Defining Drug Response for Stratified Medicine

Mike Lonergan[1], Stephen J Senn[2], Christine McNamee[3], Ann K Daly[4], Robert Sutton[5], Andrew Hattersley[6], Ewan Pearson[1], Munir Pirmohamed[3]

Mike Lonergan[1], Division of Cardiovascular & Diabetes Medicine, University of Dundee, DD1 9SY

Stephen J Senn[2], Competence Center in Methodology and Statistics, (CRP-Santé) 1A-B, rue Thomas Edison, L-1445 Strassen, Luxembourg

Christine McNamee[3], Wolfson Centre for Personalised Medicine, University of Liverpool, L69 3GL

Ann K Daly[4], Institute of Cellular Medicine, Newcastle University, NE2 4HH

Robert Sutton[5], NIHR Liverpool Pancreas Biomedical Research Unit, University of Liverpool, L69 3GA

Andrew Hattersley[6], Northcote House, University of Exeter, EX4 4QJ

Ewan Pearson[1], Division of Cardiovascular & Diabetes Medicine, University of Dundee, DD1 9SY

Munir Pirmohamed[3] Wolfson Centre for Personalised Medicine and MRC Centre for Drug Safety Science, University of Liverpool, L69 3GL

Corresponding author:

Professor Munir Pirmohamed
Correspondence address: Wolfson Centre for Personalised Medicine. Institute of Translational Medicine, University of Liverpool, Block A: Waterhouse Building, 1-5 Brownlow Street, Liverpool, L69 3GL
E-mail: munirp@liverpool.ac.uk
tel: +44 151 794 5549
fax: +44 151 794 5059

Keywords: (5-10): crossover trial, personalised medicine, precision medicine, stratified medicine, pharmacogenomics, statistical analysis, variability, adverse drug reactions

Teaser: (35 words): Understanding the patterns of variation in drug response, in terms of both therapeutic efficacy and safety, can help identify individuals or groups of patients requiring different treatments for similar symptoms and improve their care.

Abstract

The premise for stratified medicine is that drug efficacy, drug safety, or both, vary between groups of patients, and biomarkers can be used to facilitate more targeted prescribing with the aim of improving the benefit-risk ratio of treatment. However, many factors can contribute to the variability in response to drug treatment. Inadequate characterisation of the nature and degree of variability can lead to the identification of biomarkers that have limited utility in clinical settings. This paper discusses the complexities associated with the investigation of variability in drug efficacy and drug safety, and how consideration of these issues *a priori*, together with standardisation of phenotypes, can increase both efficiency of stratification procedures, and identification of biomarkers with the potential for clinical impact.

Introduction

Stratified medicine is the differential prescribing of medications, or treatment programmes, to groups of individuals based on attributes other than the symptoms resulting from their disease [1] (figure 1). The term is often considered synonymous with "personalised" "precision" or even "P4" medicine. Stratification could be seen as less ambitious, and more immediately realizable, than the other approaches because of its focus on identifying groups needing particular treatments rather than the direct optimisation of treatment for each individual patient [2]. Its purpose is to improve outcomes by refining drug dosages or by administering more appropriate treatments to specific groups of patients. It requires identification of appropriate subgroups of individuals, through the use of biomarkers, and suitable measures of the benefits and costs of the potential treatments for each group. The advantages of stratified medicine flow from the existence of subgroups, or strata, of patients such that individuals within the same subgroup tend to have similar responses to treatment, while those in different strata respond differently and require different treatment.

The evaluation of variability in individuals' responses to treatments, in terms of both efficacy and safety (figure 2), is central to assessing potential benefits from stratified medicine. While this requirement is obvious, fulfilling it is far more difficult than is generally appreciated. A lack of clarity about the extent and causes of variability in the effects of a particular treatment may result in efforts being focussed on investigations of inappropriate or irrelevant biomarkers. Therefore, in this position paper, we consider issues relating to three areas with respect to drug response:

- The variability of drug outcomes (in terms of both efficacy and adverse effects);
- The identification of subgroups that show differential efficacious responses; and
- The identification of subgroups most at risk of adverse drug outcomes.

This position paper was developed out of discussions during a workshop on Defining Drug Response for Stratified Medicine held by the UK Pharmacogenetics and Stratified Medicine Network.

Variability in drug response

In 1997 Sir Richard Sykes predicted:

> *"it will soon be possible for patients in clinical trials to undergo genetic tests to identify those individuals who will respond favourably to the drug candidate, based*

on their genotype. This will translate into smaller, more effective clinical trials with corresponding cost savings and ultimately better treatment in general practice. Individual patients will be targeted with specific treatment and personalised dosing regimens to maximise efficacy and minimise pharmacokinetic problems and other side-effects" [3].

To some extent this prediction is being realised, but variation in drug response is complex. It is important to acknowledge that a responder in a drug trial is a person who was *observed* to improve by some pre-defined standard, and not to automatically assume the drug *caused* the patient to get better. There are many sources of variation in clinical trials. These arise from:

- differences between treatments, averaged over all patients;
- differences between patients given the same treatment; and;
- differences in the effect of a single treatment given to the same patient on different occasions.

It is important to ensure that comparative clinical trials are conducted and analysed carefully to avoid the introduction of un-intended variations and the drawing of erroneous conclusions. Improving communication with statisticians before commencing trials, adopting appropriate designs for teasing out components of variation, and applying random effect methodology for improving estimates, will limit the introduction of unwanted variations into the results of clinical trials. The use of appropriate statistical models will also contain the effects of unwanted variation by representing the factors that are known, or found to correlate with, drug response.

Defining variability in drug response

"True" variability in drug response is the extent to which responses to a drug differ between individual patients. In some circumstances, these responses can vary between a positive benefit and a harmful adverse reaction. Sources of variability include:

- subject phenotype (individual characteristics) and genotype
- disease phenotype and genotype as identified by biomarkers, clinical features and outcomes
- drug formulation, route of administration, dosage, frequency and timing of administration
- subject diet and life style, including whether fed or fasting at the time of drug administration
- absorption, distribution, metabolism, excretion, bioavailability, target action, off-target action and whether the drug is administered singly, short-term or long-term

This large number of factors make it difficult to identify the true causes of observed differences in drug response and the correlations among them [4].

Uncertainty of results

Variability must be distinguished from uncertainty that a response is valid, since variability may be uncertain, but uncertainty is not necessarily due to variability. For example, during measurement of blood pressure there may be technical errors that result in values outside the range possible for true blood pressures. Usually such errors dramatically alter only a small proportion of measurements, while limitations on the accuracy of the techniques and equipment used in measuring tend to have smaller effects but act on all measurements. These two issues require different solutions [5]. Sources of variability in the results, such as that caused by limits on the accuracy of measuring of blood pressure should be considered in the design before the commencement of the experimental

study, and should be allowed for in the analysis of results. Valid, reliable measures are essential to every study; appropriate methods need to be chosen at the start of each study, and a strategy set out for discarding spurious values.

Study designs incorporating repeat measurements are likely to be substantially more accurate than designs using single measurements [6]. Repetition can be particularly useful in early phase trials, when fewer subjects are taking part. These require clear planning of study progression using predictors and sub-groups, backed up by appropriate stopping rules. Undertaking multiple studies, or sampling in multiple locations, will extend this approach providing a strict code is followed to ensure exact duplication of the method of conducting the study. Not adhering to such a code may introduce uncertainty about the true value of the results.

Simple treatment algorithms are less likely to introduce uncertainty and confusion, and so often ultimately provide more meaningful results. The uses and limitations of summary statistics have also to be fully understood. For example: mathematical transformations, such as logarithms or the use of ratios, can make the effects of drugs appear uniform and show that all subjects benefit even when the absolute benefits for those at lower risk are small. A combination of suitable transformation for analysis and back transformation for prediction can then be beneficial [7].

Clinical Trials

Different sources of variability tend to predominate in each stage of clinical trials. In preclinical studies, the use of a different species in the trial can result in variability in drug response being recorded that may not be relevant to humans. During first-in-human, early phase trials, late phase trials and observational studies, the emphasis is initially on dose identification, and tolerability, and this is followed by an increasing emphasis on efficacy as the drug progresses towards licensing approval. However, during these phases, there is much less emphasis on the variability of response. Early phase studies typically recruit small numbers of participants and can identify important differences in pharmacokinetics, whereas late phase studies recruit larger numbers and are more likely to identify off-target (usually less frequent) adverse effects. The range of drug responses makes the identification and understanding of variability in drug response a critical component at each stage of drug discovery, development, and medical practice. It is also important to note that very few trials publish explicit information on the sources and patterns of variation in their data, and this is something that should be mandated by journal editors.

Variability in drug response may also confound study outcomes. Where possible, study design should recognise and address this. Adopting a specific trial design may make it more likely to identify a particular type of variability in drug response [8] (see box 1). Given this, the ideal studies for investigating subgroups or individuals who respond better to one drug than another are crossover studies, or where possible, repeated crossover studies [7]. Such within patient studies are only suitable for relatively stable diseases treated with drugs whose effects are reversible. It is important to note that this is most relevant to efficacy, where other trial designs are commonly utilised. Randomised controlled trials are rarely used in drug safety studies; the majority of the evidence for drug-biomarker associations is based on observational study designs [9]. This is of course to be expected, particularly when the adverse event is rare and it would not be feasible both clinically and economically to mount a RCT. Nevertheless, with some phenotypes, it has been

possible to undertake RCTs to evaluate the evidence between a drug and its biomarker, for example, the use of *HLA-B*57:01* genotyping in preventing abacavir hypersensitivity [10].

Efficacy and drug response

If efficacy is the beneficial response from treatment, then trials and retrospective studies provide estimates of the efficacy of specific doses of drugs for particular sets of individuals. The use of their results in traditional, "non-stratified", medicine requires two sets of decisions:

- What treatment would be most appropriate for individuals similar to those that were studied?
- What wider group of patients can be considered sufficiently similar to the study population to be treated in the same way as them?

Those decisions remain important in stratified medicine, though this approach also requires decisions to be made about how best to subdivide the population into groups requiring different treatments. Both the generic challenges in defining efficacy and of identifying strata with different drug efficacies are considered in this section.

Defining drug efficacy

Hard endpoints, such as all-cause mortality or rates of cardiovascular death, are the simplest outcomes of drug action to interpret. However very large studies with lengthy follow-up can be necessary to produce clear results from such outcomes. Surrogate endpoints are therefore often used to evaluate the efficacy of drugs. Many of these surrogates, such as Blood Pressure, Glycosylated Haemoglobin and LDL-Cholesterol, are measured on continuous scales. The use of such surrogate endpoints introduces additional challenges, including the definition of appropriate response measures and how to account for the pre-treatment situation.

Drug responses are rarely binary, so the terms 'responder' and 'non-responder' can be misleading, and are best avoided. The terms 'responder' and 'non-responder' greatly over-simplify the issues. A responder is a patient *observed* to have had a favourable outcome, typically by some arbitrary standard, but a responder is not a patient who was necessarily *caused* to get better by a specific drug treatment. Subsequence is not consequence; careful examination of results is required to actually determine which group of patients have received an active dose of a drug that has provided them with a beneficial outcome.

It is also conventional to relate clinical outcomes to prescribed doses. However, this is problematic in two regards: first, the dose does not necessarily equate to exposure as there is often marked inter-individual variability in exposure which may be the result of genetic or disease factors, or because of co-prescribing of interacting drugs [11]. Second, adherence represents a major issue in many disease areas [12], even in trials, and is rarely accounted for in clinical studies. The labelling of non-responders without taking into account non-adherence can lead to the development of biomarkers that perform poorly in clinical practice. Further attention to the recording of adherence, and the development of novel methods to assess adherence, would increase the information available from many trials. While analysis on the basis of intent-to-treat will remain important, this additional information may help identify situations where limited efficacy and patterns of adverse

effects in individual patients or groups of patients are related more to poor adherence than the drug *per se*.

Response definition determines the "good responders": If a quantity, such as LDL-Cholesterol, is recorded before and after treatment, with results A and B respectively, the values have been used to assess drug response using different calculations (for example, the post-treatment value B; the absolute reduction A-B; and the relative reduction (A-B)/A). However, these different parameters actually measure the same value with different degrees of precision [13]. Another approach is to define success to the satisfaction of some criterion after treatment, and ask whether that treatment target was achieved. For example: a measure of success for a lipid lowering agent could be whether the LDL-Cholesterol level fell below 2 mmol/L. These measures are likely to identify different individuals as having responded best to treatment, and can identify different covariates as being associated with good response to treatment. In many publications, the drug response phenotype seems to have been selected to present a particular picture of the efficacy of the drug in question. Care therefore needs to be taken in the interpretation of results based on these types of estimates of treatment response.

The impact of the baseline on subsequent response: A high baseline pre-treatment value can be associated with a greater reduction upon treatment, but a lower likelihood of achieving a treatment target. Incorporating the pre-treatment baseline measure as an interaction term can be useful in these cases, and may provide insight into the causes and mechanisms of disease and treatment. It can also help shift the focus from identifying groups responsive to particular treatments to, often more clinically meaningful, questions about which treatment is best for people with particular characteristics. However, uncertainty in baseline estimates also complicates the analysis. Estimates of absolute reduction and treat-to-target approaches, in particular, are prone to regression to the mean: individuals whose initial results were high due to chance are likely to produce lower results on retesting, even without any treatment. It is also well known that simple regression techniques underestimate the sizes of effects, and produce misleading estimates of uncertainty and statistical significance, when the explanatory covariates they incorporate are imprecisely measured [14]. Some research groups therefore prefer not to adjust for the pre-treatment measure; for example, this has been used to assess statin efficacy where genome-wide pharmacogenetic variants that determine differences between the baseline and treatment measure have been identified [15].

The effects of time on measures of response: Estimates of individuals' responses to treatments can also be sensitive to the timing of the on-treatment measurement. Both the time the drug takes to act and the rate of progression of the underlying disease can affect the results of studies. For example, in diabetes, the HbA1c reduction achieved can increase over the first 3-6 months of treatment, and this is usually followed by a gradual deterioration thereafter. While that pattern could reflect gradual decrease in efficacy of the drug, it seems more likely to reflect progression of the underlying disease. Studies of treatments of diabetes, and similar disorders, therefore need to choose, and justify, the time from starting on drugs, at which they measure response and efficacy.

The use of efficacy in stratification

As well as the issues that concern the definition of drug efficacy *per se*, there are further ones that affect the identification of strata, the groups of individuals who respond to treatment in distinct

ways. Most of the issues around the definition of efficacy are exactly the same for stratified medicine as for any other decisions about treatment. Stratification can almost be considered as an add-on, taking the clinically relevant definition of efficacy and applying it to the splitting of the population into strata with different treatment responses. However, the choice of the numbers of strata, and the positioning of their boundaries, raise additional issues that need to be addressed.

***Defining stratum boundaries*:** Strata are categories of patients identified for different treatments, of symptoms they share, on the basis of characteristics other than the severity of their disease symptoms. The simplest way to define strata is by using easily observed properties of individuals: for example treating women differently from men, or separating out those with a particular genetic variant. A set of such categorical distinctions can easily be drawn up, based on some combination of subjective judgements of plausibility and the limitations imposed by data availability.

Size of effect as well as significance is important for clinical implementation: Various techniques, ranging from the examination of differences in mean responses to the fitting of sophisticated statistical models, can be used to decide which distinctions are worth respecting. These decisions only partly depend on the statistical significance of estimates of differences: complexity is a cost in itself, partly because it is likely to increases the frequency of mistakes. Just because a significant interaction has been noted between the treatment and one or more covariates does not imply that using these covariates to determine treatment is necessarily a good idea. The variation in effect of treatment may not be enough to justify the extra complication and cost in using such covariates to guide treatments.

***Location of boundaries between strata*:** Continuous variables can be split to give categories that can be treated in the way described above. This can produce benefits of simplicity, convenience and consistency. For example, splitting BMI at 30 and considering "healthy weight" and "obese" people separately might naturally fit into many practitioners' views of the population. However, such intuitive boundaries might not be appropriate: it could be that the greatest change in response to a medication occurs at around a BMI of 28, or 34. In such circumstances the convenience of using a BMI of 30 as a boundary needs to be weighed against the risk of reducing the quality of medical treatment received by substantial numbers of people.

While some characteristics, such as BMI, that affect people's responses to medication are measured on a continuous scale, treatment decisions tend to be less flexible. Usually the choice is between a limited set of medications, each of which is available in a small number of different doses. Stratification on BMI therefore requires choosing thresholds to separate people requiring different treatments. The uncertainty in estimates of responses to treatment is important to the evaluation of both particular stratification schemes and the sensitivity of their benefits to the exact choice of stratum boundaries. For example, if one drug was better than another at low BMI but the two were equally effective for people with BMIs greater than 28, then whether to stratify at all would depend on the size of the difference between treatment responses at very low and high BMIs, but whether to move the division to 30, the conventional definition of obesity, would depend on the estimated size of the differences in response to the two treatments at BMIs between 28 and 30.

Other requirements for stratification: The identification of strata of patients who respond particularly well or poorly to a drug is an important step, but does not immediately justify a stratified approach. For example, where a drug is not effective there also needs to be a better alternative treatment strategy, which might even be to do nothing, available for that stratum of patients. That would not apply if there were simply some people with serious conditions who responded poorly to all available treatments, so stratification requires intra-individual differential responses to drugs, a situation that in general is assumed rather than established. Parallel design RCTs can show that patients with one particular phenotype tend to respond poorly to one drug, and well to another, but crossover trials, where individuals are exposed to a sequence of treatments, are necessary to show that particular individuals respond better to one of the treatments. These also tend to be more efficient and powerful than parallel trials, though there are situations, particularly with rapidly resolving, progressive and intermittently recurring conditions, where crossover designs may be difficult or impossible to implement. Observational data can demonstrate that individuals have responded better to one of a set of treatments, but their interpretation can be complicated by their lack of randomisation and difficulties in assessing the comparability of the conditions under which each treatment was used.

Choice of biomarkers for stratification: There are many potential biomarkers that could be used in the stratification of treatment. These range from transient biochemical changes, through longer lasting social, physical, and environmental conditions, to permanent, typically genetic, characteristics of individuals. Different biomarkers will be suitable for different stratification processes. The choice of biomarkers will depend on their relevance to the particular disease, the method and reproducibility by which they have been identified and the stratification decision. Currently a major focus of stratified medicine is on -omic biomarkers (e.g. proteomic, metabolomics, genomic), but it is also important to consider simple phenotypic biomarkers and traits such as age, duration of disease, BMI, and ethnicity. The influence of these simple characteristics on drug efficacy has been studied in very few diseases. Transparent reporting of any prediction models developed is also important [16].

Stratification to reduce adverse reactions to drugs

Medicines are not always effective in improving health: some patients have adverse reactions to drugs. These cause 6.5% of admissions into hospital [17], including 2.5% emergency admissions [18]. 14.7% of inpatients develop an ADR in hospital [19], and 25% of primary care consultations involve patients seeking help for an adverse drug event [20]

Adverse drug reactions may be classified as on target (predictable from the known primary or secondary pharmacology of the drug, and with a clear dose-dependence relationship within the individual) or off target (not predictable from a knowledge of the basic pharmacology of the drug and can exhibit marked inter-individual susceptibility with complex dose-dependency) [21]. Both on-target and off-target adverse drug reactions are important contributors to mortality and morbidity in clinical practice in any healthcare environment.

Many factors predispose patients to adverse drug reactions; these may be environmental, clinical, or genetically based. Despite some success such as the identification of HLA-B*57:01 as the main genetic risk factor for abacavir hypersensitivity [10], we still do not know the full extent of the

genetic contribution to most adverse drug reactions [21]. To some extent, those predictors of ADRs which have already been identified can be regarded as "low hanging fruit" for stratification. This suggests that the predictors that remain to be identified may be more complex and require major efforts for identification. As with efficacy issues, it is important to remember that predisposition to ADRs may be due to genetic as well as non-genetic factors, and a holistic approach that simultaneously evaluates both is important.

Sources of variability in drug safety studies

Many of the issues discussed above with regard to drug efficacy studies are also important for drug safety. In addition, some other specific areas need to be considered for drug safety.

Definitions: There are many different definitions of adverse drug reactions [22]. For example, FDA considers lack of efficacy to be a safety issue though this view is not universal. More recently, the EU pharmacovigilance legislation has widened the definitions of adverse drug reactions to include drug misuse, drug abuse, and medication errors [23]. Thus it is important to be clear what is being covered within a particular study.

Causal relationship: An additional issue to consider is that an event occurring during exposure to a drug may not necessarily be causally related to the drug. It is therefore important to undertake causality assessment. Over 30 different tools have been developed, and none of them are perfect. For instance, the Liverpool causality assessment tool [24] was developed as a more user friendly alternative to the older Naranjo tool [25]. However, further work that undertakes comparative assessments of different causality tools is required. For any ADR study, it is important to highlight which causality assessment tool was used, who undertook that assessment, and how consensus was reached on difficult cases where causality was difficult to assign.

Phenotype standardisation: An important limitation to progress is that the quality of clinical phenotyping is poor, with considerable heterogeneity within the same overall adverse drug reaction class. Imprecise phenotyping, and mixing of phenotypes, can lead to both false-positive and false-negative associations, and it is therefore important to ensure that standardised phenotypes are used. Some progress has already been made by the phenotype standardisation project. [26] This has concentrated on selected idiosyncratic reactions involving liver [27], skin [28], torsades de pointes [29], muscle [30], renal injury [31] and angioedema [32]. If such standardisation is not available for a particular phenotype, then it is important to pre-specify the phenotype being studied.

The issue of obtaining an accurate phenotype is relevant to adverse reactions during both clinical trials and more generally during routine health care. Adverse reactions recorded during routine clinical practice may not be adequately detailed, making it difficult to identify and recruit patients with the inclusion criteria to studies; a particular problem with retrospective case-control designs. Phenotype can involve very simple and clear cut parameters (for example date of death) or very complex issues (such as cause of death) but must be clearly defined in order to accurately determine what data is required to establish the true phenotype of the reaction. This information is of course often simpler to obtain and more accurate in prospective studies, though these studies can be costly and take a long time to complete.

Increasing emphasis is being placed on the role of electronic health records in identifying and recruiting patients for stratified medicine studies [1]. Where the adverse event is an easily quantifiable parameter (for example, change in creatine kinase levels after the start of statin treatment [33]), electronic records may enable large-scale recruitment to studies. However, when the phenotype is more complex, and more deep phenotyping procedures are needed to correctly categorise patients, this information may not necessarily exist within the electronic health record. Adverse reactions are often described in the simplest of terms (for example "skin rash"). In the long-term, better standards for the capture of clinical data are needed.

Finally, although the underlying aim of the phenotype standardisation project is to have specific sets of criteria for defining the ADR, issues may still remain which may render both the phenotype standardisation and causality imprecise. For instance, with idiosyncratic reactions involving the liver, a specific issue to be considered is whether the definition of toxicity is based directly on changes in liver function test results or whether Hy's law (drug-induced hepatocellular injury associated with at least 3-fold elevation of transaminases above the upper limit of normal (ULN) and a concomitant elevation in bilirubin by at least 2-fold above the ULN) is applied [27]. Some studies, for example for lumiracoxib-induced liver injury, have shown a correlation between the severity of the liver derangement and strength of the association with the genetic biomarker [34]. It is also important to consider that for some ADRs, the symptoms and signs abate even though the drug is continued. For example, with liver injury, with certain drugs (for example isoniazid), after an initial rise in liver enzymes, adaptation usually occurs and the liver enzymes revert to normal without drug discontinuation. The phenomenon of adaptation [35] is usually not considered in genetic studies of drug-induced liver injury. Indeed, given that most guidelines suggest that a drug should be discontinued when the ALT>5xULN, it would appear unethical to continue the drug beyond that point.

In summary, therefore, all studies on drug safety should be required to show:

- the definition used for the adverse event
- how causality assessment was undertaken
- how the phenotype was defined
- for quantitative variables, the cut-off values used, along with the justification for these values

Summary

Many areas such as cancer, inflammatory and rare diseases have made great progress in stratifying patient's treatment to deliver "the right drug, at the right time, at the right dose". Grouping patients into strata based on their genetic profile, or molecular basis of their disease, then treating them based on their strata (rather than the symptoms they present) has already improved healthcare. Not only do patients recover more quickly but there is a potential to improve drug safety by drastically reducing adverse drug reactions, with the knock on effect of significantly reducing costs for the health service. However, in order for stratified approaches to succeed, it is important to consider for a particular intervention and disease area how variability is actually defined, define processes in a standardised manner from the beginning, and ensure that the study design is optimal to answer the

hypothesis. Considerations of these issues upfront will increase both the internal and external validity of stratification procedures, and enable better translation into clinical practice.

Acknowledgments

The article is based on a workshop held by the UK Pharmacogenetics and Stratified Medicine Network [http://www.uk-pgx-stratmed.co.uk/], which was funded by the Knowledge Transfer Network and Innovate UK at the time of the workshop. MP is an NIHR Senior Investigator and acknowledges the support of the MRC Centre for Drug Safety Science and Wolfson Centre for Personalised Medicine, both in Liverpool. RS is also an NIHR Senior Investigator, and acknowledges the support of the NIHR Liverpool Pancreas Biomedical Research Unit. SJS gratefully acknowledges funding through grant 602552 for the IDEAL project under the European Union FP7 programme. We acknowledge all those who participated in the workshop (Alan Houston, Alberto Martinez, Andrew Read, Angus Jones, Beverley Shields, Carla Deakin, Catherine Angwin, Cecil Kullu, Clive Bowman, Gareth Veal, Helen Gallagher, John Warren, Kaixin Zhou, Lauren Rodgers, Louise Brown, Mark Bartlett, Matt Rowson, Matthew Hallsworth, Michael Barnes, Neil Hawkins, Peter Dewland, Richard Owen, Robin Choudhury, Rothe Hansjoerg).

References

1 The Academy of Medical Sciences. (2013) *Realising the potential of stratified medicine* The Academy of Medical Sciences

2 The Academy of Medical Sciences (2015) *Stratified, personalised or P4 medicine: a new direction for placing the patient at the centre of healthcare and health education* Academy of Medical Sciences

3 Sykes, R. (1997) *The Pharmaceutical Industry in the New Millenium: Capturing the Scientific Promise*, Centre for Medicines Research

4 Senn, S. (2004) Individual response to treatment: is it a valid assumption? *BMJ* 329 (7472), 966-968

5 Giorgini, P. et al. (2014) A review of blood pressure measurement protocols among hypertension trials: implications for "evidence-based" clinical practice. *J Am Soc Hypertens* 8 (9), 670-676

6 Vickers, A.J. (2003) How many repeated measures in repeated measures designs? Statistical issues for comparative trials. *BMC Medical Research Methodology* 3 (1), 1-9

7 Senn, S. (2016) Mastering variation: variance components and personalised medicine. *Stat Med* 35 (7), 966-977

8 Senn, S. (2001) Individual Therapy: New Dawn or False Dawn? *Drug Information Journal* 35 (4), 1479-1494

9 Mitchell, A.A. and Lesko, S.M. (1995) When a randomised controlled trial is needed to assess drug safety. The case of paediatric ibuprofen. *Drug Saf* 13 (1), 15-24

10 Mallal, S. et al. (2008) HLA-B*5701 screening for hypersensitivity to abacavir. *N Engl J Med* 358 (6), 568-579

11 Mathijssen, R.H.J. et al. (2014) Determining the optimal dose in the development of anticancer agents. *Nat Rev Clin Oncol* 11 (5), 272-281

12 Capoccia, K. et al. (2016) Medication Adherence With Diabetes Medication: A Systematic Review of the Literature. *Diabetes Educ* 42 (1), 34-71

13 Senn, S. (2007) *Statistical Issues in Drug Development*, Wiley

14 Yanez, N.D., 3rd et al. (2002) A regression model for longitudinal change in the presence of measurement error. *Ann Epidemiol* 12 (1), 34-38

15 Barber, M.J. et al. (2010) Genome-wide association of lipid-lowering response to statins in combined study populations. *PLoS One* 5 (3), e9763

16 Collins, G.S. et al. (2015) Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD StatementThe TRIPOD Statement. *Annals of Internal Medicine* 162 (1), 55-63

17 Pirmohamed, M. et al. (2004) Adverse drug reactions as cause of admission to hospital: prospective analysis of 18 820 patients. *BMJ* 329 (7456), 15-19

18 Budnitz, D.S. et al. (2006) National surveillance of emergency department visits for outpatient adverse drug events. *JAMA* 296 (15), 1858-1866

19 Davies, E.C. et al. (2009) Adverse drug reactions in hospital in-patients: a prospective analysis of 3695 patient-episodes. *PLoS One* 4 (2), e4439

20 Gandhi, T.K. et al. (2003) Adverse drug events in ambulatory care. *N Engl J Med* 348 (16), 1556-1564

21 Pirmohamed, M. (2014) Personalized pharmacogenomics: predicting efficacy and adverse drug reactions. *Annu Rev Genomics Hum Genet* 15, 349-370

22 Edwards, I.R. and Aronson, J.K. (2000) Adverse drug reactions: definitions, diagnosis, and management. *Lancet* 356 (9237), 1255-1259.

23 Borg, J.J. et al. (2015) European Union pharmacovigilance capabilities: potential for the new legislation. *Ther Adv Drug Saf* 6 (4), 120-140

24 Gallagher, R.M. et al. (2011) Development and inter-rater reliability of the Liverpool adverse drug reaction causality assessment tool. *PLoS One* 6 (12), e28096

25 Naranjo, C.A. et al. (1981) A method for estimating the probability of adverse drug reactions. *Clin Pharmacol Ther* 30 (2), 239-245.

26 Pirmohamed, M. et al. (2011) The phenotype standardization project: improving pharmacogenetic studies of serious adverse drug reactions. *Clin Pharmacol Ther* 89 (6), 784-785

27 Aithal, G.P. et al. (2011) Case definition and phenotype standardization in drug-induced liver injury. *Clin Pharmacol Ther* 89 (6), 806-815

28 Pirmohamed, M. et al. (2011) Phenotype standardization for immune-mediated drug-induced skin injury. *Clin Pharmacol Ther* 89 (6), 896-901

29 Behr, E.R. et al. (2013) The International Serious Adverse Events Consortium (iSAEC) phenotype standardization project for drug-induced torsades de pointes. *Eur Heart J* 34 (26), 1958-1963

30 Alfirevic, A. et al. (2014) Phenotype Standardization for Statin-Induced Myotoxicity. *Clin Pharmacol Ther* 96 (4), 470-476

31 Mehta, R.L. et al. (2015) Phenotype standardization for drug-induced kidney disease. *Kidney Int* 88 (2), 226-234

32 Wadelius, M. et al. (2014) Phenotype Standardization of Angioedema in the Head and Neck Region Caused by Agents Acting on the Angiotensin System. *Clin Pharmacol Ther* 96 (4), 477-481

33 Carr, D.F. et al. (2013) SLCO1B1 Genetic Variant Associated With Statin-Induced Myopathy: A Proof-of-Concept Study Using the Clinical Practice Research Datalink. *Clin Pharmacol Ther* 94 (6), 695-701

34 Singer, J.B. et al. (2010) A genome-wide study identifies HLA alleles associated with lumiracoxib-related liver injury. *Nat Genet* 42 (8), 711-714

35 Dara, L. et al. (2016) Mechanisms of Adaptation and Progression in Idiosyncratic Drug Induced Liver Injury, Clinical Implications. *Liver Int* 36, 158-165.

Figure legends

Figure 1: Schematic of the differences between traditional and stratified approaches to patient treatment. Treatment failure can result from either a lack of response or unacceptable adverse-effects.

Figure 2: Drug response includes both efficacious and adverse responses; together these determine the benefit-harm ratio of drugs.