ELSEVIER

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] and Munir Pirmohamed[3]

[1] Division of Cardiovascular & Diabetes Medicine, University of Dundee, Dundee DD1 9SY, UK
[2] Competence Center in Methodology and Statistics, (CRP-Santé) 1A-B, Rue Thomas Edison, L-1445 Strassen, Luxembourg
[3] Wolfson Centre for Personalised Medicine, University of Liverpool, Liverpool L69 3GL, UK
[4] Institute of Cellular Medicine, Newcastle University, Newcastle NE2 4HH, UK
[5] NIHR Liverpool Pancreas Biomedical Research Unit, University of Liverpool, Liverpool L69 3GA, UK
[6] Northcote House, University of Exeter, Exeter EX4 4QJ, UK

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. Here, we discuss 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 the 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 programs, to groups of individuals based on attributes other than the symptoms resulting from their disease [1] (Fig. 1). The term is often considered synonymous with 'personalised', 'precision', or even 'P4' medicine. Stratification could be seen as less ambitious, and more immediately realisable, 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 the 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, whereas 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 (Fig. 2), is central to assessing potential benefits from stratified medicine. Although this requirement is obvious, fulfilling it is more difficult than is generally appreciated. A lack of clarity about the extent and causes of variability in the effects of a particular treatment can 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: (i) the variability of drug outcomes (in terms of both efficacy and adverse effects); (ii) the identification of subgroups that show differential efficacious responses; and (iii) 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.

Corresponding author: Pirmohamed, M. (munirp@liverpool.ac.uk)

http://dx.doi.org/10.1016/j.drudis.2016.10.016

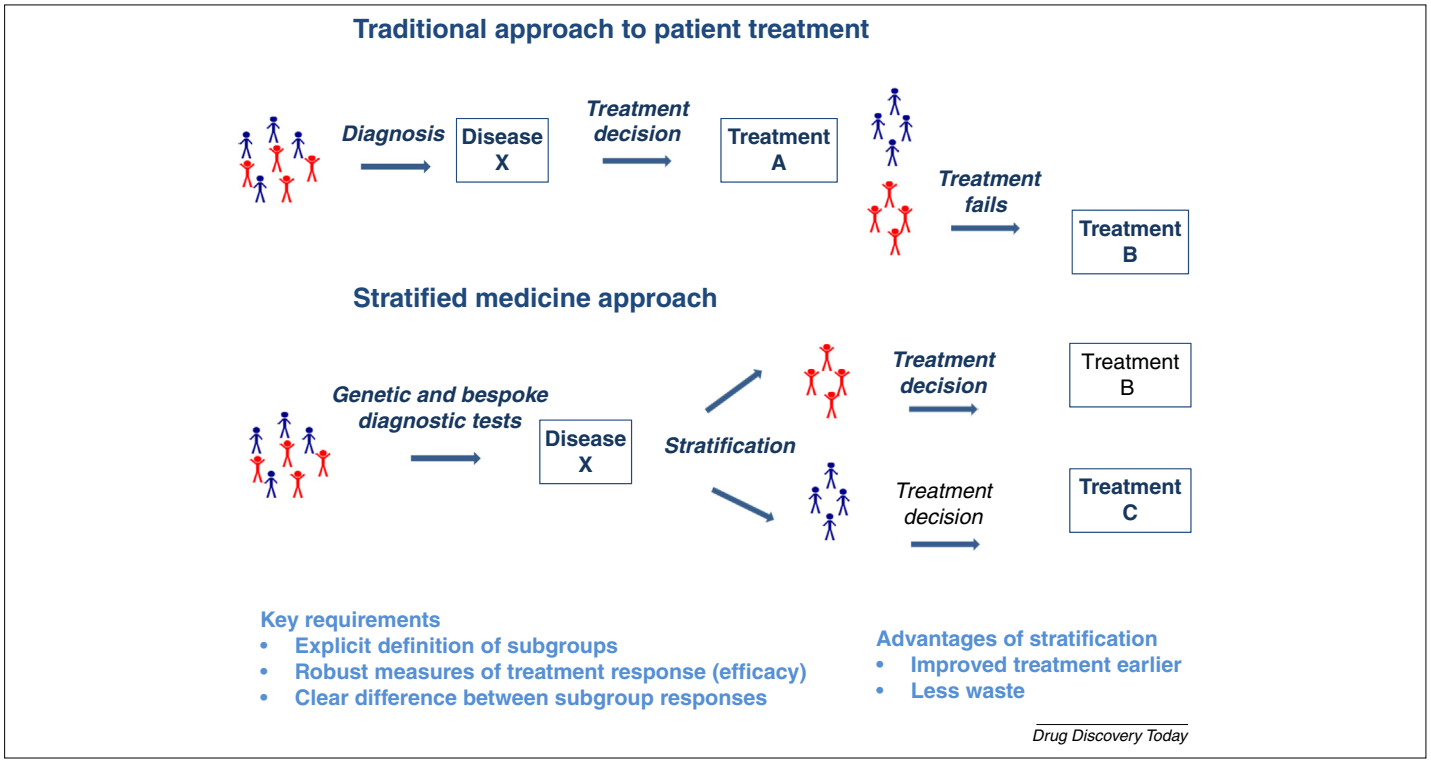

FIGURE 1

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.

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 predefined standard, and not to automatically assume that the drug caused the patient to get better. There are many sources of variation in clinical trials. These arise from: (i) differences between treatments, averaged over all patients; (ii) differences between patients given the same treatment; and (iii) 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 unintended 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.

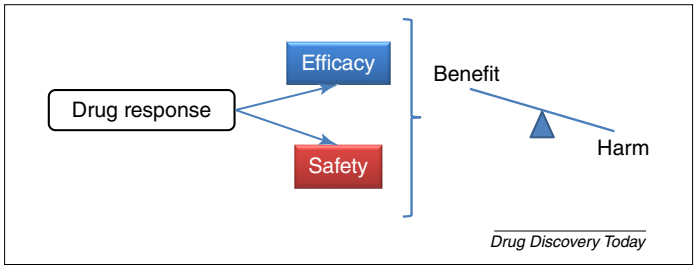

FIGURE 2

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

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: (i) subject phenotype (individual characteristics) and genotype; (ii) disease phenotype and genotype, as identified by biomarkers, clinical features, and outcomes; (iii) drug formulation, route of administration, dosage, frequency, and timing of administration; (iv) subject diet and life style, including whether fed or fasting at the time of drug administration; and (v) 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 makes 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, because variability can be uncertain, but uncertainty is not necessarily the result of variability. For example, during measurement of blood pressure, there might 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, whereas 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 during early-phase trials, when fewer subjects are taking part. These require clear planning of study progression using predictors and subgroups, 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 can 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 during each stage of a clinical trial. In preclinical studies, the use of a different species in the trial can result in variability in drug response being recorded that might 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 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 crucial component at each stage of drug discovery, development, and medical practice. It is also important to note that 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 can also confound study outcomes. Where possible, study design should recognise and address this. Adopting a specific trial design might make it more likely to identify a particular type of variability in drug response [8] (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 (RCT) 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: (i) what treatment would be most appropriate for individuals similar to those that were studied?; and (ii) 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, although this approach also requires decisions to be made about how best to subdivide the population into groups requiring dif-

BOX 1

Trial designs and identification of variability in drug response

Parallel group designs (where each patient is randomised to receive a single treatment) can identify differences between treatments, but are less likely to identify variability in drug response arising from differences between patients (e.g. severity of disease), from interaction between individual patients and their disease, or variability within a patient on different occasions.

Crossover designs (where each patient receives each treatment in one period only) can identify variability in drug responses arising from differences between treatments as well as from differences between patients.

Repeated crossover designs (where each patient receives each treatment at least twice), go further and, importantly, can identify variability arising between treatments and from interaction between individual patients and their treatment; however, identification of variability within a patient on different occasions requires repeated measurements of that patient.

ferent 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. Therefore, surrogate endpoints are often used to evaluate the efficacy of drugs. Many of these surrogates, such as blood pressure, glycosylated haemoglobin and low-density lipoprotein (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 pretreatment situation.

Drug responses are rarely binary, so the terms 'responder' and 'nonresponder' can be misleading, and are best avoided. The terms 'responder' and 'nonresponder' also greatly oversimplify 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 has 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 because there is often marked interindividual variability in exposure that could be the result of genetic or disease factors, or because of coprescribing 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 nonresponders without taking into account nonadherence 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 it, would increase the information available from many trials. Although analysis on the basis of intent-to-treat will remain important, this additional information could 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 to 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 will have been used to assess drug response using different calculations [e.g. the post-treatment value B; the absolute reduction $A - B$; and the relative reduction $(A - B)/A$]. However, these different parameters in fact 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 appears to have been selected to present a particular picture of the efficacy of the drug in question. Therefore, care needs to be taken in the interpretation of results based on these types of estimate of treatment response.

The impact of the baseline on subsequent response

A high baseline pretreatment value can be associated with a greater reduction upon treatment, but a lower likelihood of achieving a treatment target. Incorporating the pretreatment baseline measure as an interaction term can be useful in these cases, and can 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 patients 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 because of 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]. Therefore, some research groups prefer not to adjust for the pretreatment 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 haemoglobin (Hb)A1c reduction achieved can increase over the first 3–6 months of treatment, and this is usually followed by a gradual deterioration thereafter. Although that pattern could reflect a gradual decrease in efficacy of the drug, it seems more likely to reflect progression of the underlying disease. Therefore, studies of treatments of diabetes and similar disorders need to choose and justify the time from starting a drug 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 and of symptoms they share, on the basis of characteristics other than on 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 increase 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 might 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 body mass index (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.

Whereas some characteristics, such as BMI, that affect patients' 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. Therefore, stratification on BMI requires choosing thresholds to separate patients 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 patients 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 patients with serious conditions who responded poorly to all available treatments; thus, stratification requires intraindividual 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 are parallel trials, although there are situations, particularly with rapidly resolving, progressive, and intermittently recurring conditions, where crossover designs might 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, metabolomic, and 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 adverse drug reaction (ADR) in hospital [19], and 25% of primary care consultations involve patients seeking help for an ADR [20].

ADRs can be classified as on target (predictable from the known primary or secondary pharmacology of the drug, and with a clear dose-dependence relation within the individual) or off target (not predictable from a knowledge of the basic pharmacology of the drug and can exhibit marked interindividual 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 can 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 ADRs [21]. To some extent, those predictors of ADRs that have already been identified can be regarded as 'low-hanging fruit' for stratification. This suggests that the predictors that remain to be identified are more complex and require major efforts for identification. As with efficacy issues, it is important to remember that predisposition to ADRs can be the result of genetic as well as nongenetic 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, other specific areas also need to be considered for drug safety.

Definitions of adverse drug reactions

There are many different definitions of ADRs [22]. For example, the FDA considers lack of efficacy to be a safety issue, although this view is not universal. More recently, the European Union (EU) pharmacovigilance legislation widened the definitions of ADRs 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 between drug and adverse event

An additional issue to consider is that an adverse event occurring during exposure to a drug might not necessarily be causally related to the drug. Therefore, it is 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 of adverse drug reactions

An important limitation to progress is that the quality of clinical phenotyping is poor, with considerable heterogeneity within the same overall ADR class. Imprecise phenotyping, and mixing of phenotypes, can lead to both false-positive and false-negative associations, and, therefore, it is 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 prespecify the phenotype being studied.

The issue of obtaining an accurate phenotype is relevant to ADRs both during clinical trials and more generally during routine healthcare. ADRs recorded during routine clinical practice might not be adequately detailed, making it difficult to identify and recruit patients with the inclusion criteria to studies; this is a particular problem with retrospective case–control designs. Phenotype can involve simple and clear-cut parameters (e.g. date of death) or complex issues (e.g. cause of death) but must be clearly defined to accurately determine what data are required to establish the true phenotype of the reaction. This information is of course often simpler to obtain and more accurate in prospective studies, although 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 ADR is an easily quantifiable parameter (e.g. change in creatine kinase levels after the start of statin treatment [33]), electronic records can enable large-scale recruitment to studies. However, when the phenotype is more complex, and deeper phenotyping procedures are needed to correctly categorise patients, this information might not necessarily exist within the electronic health record. Adverse reactions are often described in the simplest of terms (e.g. '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 might still remain that could 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 [i.e. drug-induced hepatocellular injury associated with at least threefold elevation of transaminases above the upper limit of normal (ULN) and a concomitant elevation in bilirubin by at least twofold 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 (e.g. isoniazid), after an initial increase 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 alanine transaminase (ALT) $>5 \times$ ULN, it would appear unethical to continue the drug beyond that point.

In summary, therefore, all studies on drug safety should be required to show: (i) the definition used for the ADR; (ii) how causality assessment was undertaken; (iii) how the phenotype was defined; and (iv) for quantitative variables, the cut-off values used, along with the justification for these values.

Concluding remarks

Many areas, such as cancer, and inflammatory and rare diseases, have made great progress in stratifying patients' 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 also the potential to improve drug safety by drastically reducing ADRs, with the knock-on effect of significantly reducing

costs for the health service. However, for stratified approaches to succeed, it is important to consider for a particular intervention and disease area how variability is 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.

Acknowledgements

The article is based on a workshop held by the UK Pharmacogenetics and Stratified Medicine Network (www.uk-pgx-stratmed.co.uk/), which was funded by the Knowledge Transfer Network and Innovate UK at the time of the workshop. M.P. 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. R.S. is also an NIHR Senior Investigator, and acknowledges the support of the NIHR Liverpool Pancreas Biomedical Research Unit. S.J.S. 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, and 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 Millennium: Capturing the Scientific Promise*. Centre for Medicines Research
4 Senn, S. (2004) Individual response to treatment: is it a valid assumption? *BMJ* 329, 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, 670–676
6 Vickers, A.J. (2003) How many repeated measures in repeated measures designs? Statistical issues for comparative trials. *BMC Med. Res. Methodol.* 3, 1–9
7 Senn, S. (2016) Mastering variation: variance components and personalised medicine. *Stat. Med.* 35, 966–977
8 Senn, S. (2001) Individual therapy: new dawn or false dawn? *Drug Inf. J.* 35, 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, 15–24
10 Mallal, S. *et al.* (2008) HLA-B*5701 screening for hypersensitivity to abacavir. *N. Engl. J. Med.* 358, 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, 272–281
12 Capoccia, K. *et al.* (2016) Medication adherence with diabetes medication: a systematic review of the literature. *Diabetes Educ.* 42, 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, 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, e9763
16 Collins, G.S. *et al.* (2015) Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD Statement. *Ann. Intern. Med.* 162, 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, 15–19
18 Budnitz, D.S. *et al.* (2006) National surveillance of emergency department visits for outpatient adverse drug events. *JAMA* 296, 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, e4439
20 Gandhi, T.K. *et al.* (2003) Adverse drug events in ambulatory care. *N. Engl. J. Med.* 348, 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, 1255–1259
23 Borg, J.J. *et al.* (2015) European Union pharmacovigilance capabilities: potential for the new legislation. *Ther. Adv. Drug Saf.* 6, 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, e28096
25 Naranjo, C.A. *et al.* (1981) A method for estimating the probability of adverse drug reactions. *Clin. Pharmacol. Ther.* 30, 239–245
26 Pirmohamed, M. *et al.* (2011) The phenotype standardization project: improving pharmacogenetic studies of serious adverse drug reactions. *Clin. Pharmacol. Ther.* 89, 784–785
27 Aithal, G.P. *et al.* (2011) Case definition and phenotype standardization in drug-induced liver injury. *Clin. Pharmacol. Ther.* 89, 806–815
28 Pirmohamed, M. *et al.* (2011) Phenotype standardization for immune-mediated drug-induced skin injury. *Clin. Pharmacol. Ther.* 89, 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, 1958–1963
30 Alfirevic, A. *et al.* (2014) Phenotype standardization for statin-induced myotoxicity. *Clin. Pharmacol. Ther.* 96, 470–476
31 Mehta, R.L. *et al.* (2015) Phenotype standardization for drug-induced kidney disease. *Kidney Int.* 88, 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, 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, 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, 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