Managing Clinical Heterogeneity: An Argument for Benefit-Based Action Limits

The use of reference ranges is well established in medical practice and research. Classically, a range would be derived from the local healthy population and matched in age, gender, and other characteristics to the patients under investigation. However, recruiting suitable controls is problematic and the derivation of the range by excluding 2.5% at each end of the distribution results in 5% of the values being arbitrarily discarded. Thus, the traditional reference range is derived using statistical and not clinical principles. While these considerations are recognized by clinicians, it is often not realized that the application of whole population derived reference ranges to complex pathologies that comprise patient subgroups may be problematic. Such subgroups may be identified by phenotypes including genetic etiology, variations in exposure to a causative agent, and tumor site. In this review, we provide examples of how subgroups can be identified in diverse pathologies and how better management can be achieved using evidence-based action limits rather than reference ranges. We give examples from our clinical experience of problems arising from using the wrong reference ranges for the clinical situation. Identifying subgroups will often enable clinicians to derive specific action limits for treatment that will lead to customized management and researchers a route into the study of complex pathologies.

Keywords: laboratory results, heterogeneity, reference ranges, action limits, evidence-based medicine, lipid lowering therapy, testosterone replacement therapy

Introduction

Determining the concentration of chemicals in body fluids is commonplace in routine clinical management and research into disease mechanisms. Thus, in establishing whether an individual has a disease, the value of an analyte is often compared with those in a reference range (often mistaken for a healthy or normal range) derived from the distribution of the analyte in an assumed healthy population. Such distributions are often Gaussian (e.g., serum glucose) though asymmetrical distributions are observed (e.g., serum bilirubin). In a Gaussian distribution, 95% of the samples will be symmetrically distributed around the mean and included in the formula, mean ±2 SD (Fig. 1). Rather arbitrarily, this formula describes a reference range. Importantly, 2.5% of the samples at the high and low ends have been ignored. Thus, the classic reference range is based on two assumptions: first, the selected individuals comprise a reasonable comparison group for patients under investigation for a particular pathology, and second, 5% of these apparently healthy subjects tested have an analyte level out with the reference range.

While reference ranges are widely used and are usually helpful, their use in clinical medicine without consideration of relevant factors such as clinical presentation and medication can be problematic. For example, the reference range for cortisol (>550 nmol/l) was removed from one Hospital laboratory’s reports following the death of a patient, admitted in a semi-conscious state exhibiting a purpuric
Disease Viewed as a Homogeneous Entity

Figure 2 shows a simple pathway with causative factor(s) leading to a disease, perceived as homogeneous. It is diagnosed via clinical presentation(s) and further investigations. Often guidelines seeking simplicity adopt this format with heterogeneity being ignored. This approach leads to relatively easy diagnosis and treatment with minimal cost. However, such simplicity may compromise understanding of the disease process, optimization of treatment, and slow further research.

A good example of this approach is the United Kingdom National Institute for Health and Care Excellence (NICE) guideline for the management of type 2 diabetes (T2DM), which appears to consider this disease as a single pathology with no allowance for different causative factors. Therapeutic management is based on glycemic control, hypertension, and cardiovascular disease (CVD) risk and is not tailored to the major underlying causative factors of beta cell dysfunction and insulin resistance.

Guidelines are largely driven by evidence-based medicine, principally primary end points of randomized controlled trials (RCT) and cost-effectiveness, often measured as quality adjusted Life Years if benefit has been demonstrated. Thus, subgroups will only be considered if the RCT inclusion criteria restrict the patient cohort to one (or more) identified subgroup. This is evident in trials investigating the benefits of lipid lowering in differing clinical presentations (which can be considered subgroups). The NICE guidelines (CG181), recommending risk assessment and reduction of adult CVD including lipid modification, consider CVD (a composite of coronary heart disease (CHD) and stroke) as a single pathology with no specific causative factors. However, where therapeutic management is considered for CVD risk reduction, subgroups such as patients with T2DM and those requiring primary prevention or secondary prevention are considered separately, because of the inclusion criteria of RCT. Interestingly, separate guidelines (CG71) have been designed for familial hypercholesterolemia (FH), but no other etiology or classification such as the metabolic syndrome (MetS). Although subgroup analyses of the trials show that patients with the MetS (up to 33% of patients in the five main fibrate trials) do benefit from fibrate therapy [1], the NICE guidelines do not recommend these therapeutic agents. Thus, it is clear that most guidelines are rigidly based on the populations included in RCTs. Often, different subgroups are grouped together; this broad-brush methodology could lead to a subgroup benefiting greatly from an intervention not being identified due to dilution of the outcome results with subgroups that do not benefit.

The above argument demonstrates the importance of subgroups based on etiology and/or clinical presentations being identified. This leads to a more complex network of phenotypes grouped together as a single pathology. Figure 3 demonstrates potential aspects of this complexity. It is clear that over time, the guidelines dealing with lipid lowering have partially moved toward Fig. 3 with FH (etiology) and primary/secondary CVD (clinical presentation) prevention having separate management strategies.

Identification of subgroups can be based on relatively simple presentational phenotypes as well as more complex genetic susceptibility markers. For example, in patients with basal cell carcinoma (BCC), we identified different tumor presentation phenotypes and some of their associated susceptibility and modifying factors [2–5]. Thus, Fig. 4 shows putative subgroups we identified in patients with at least one BCC. While many patients develop only one tumor, others were found to have “clusters” of multiple lesions over similar follow-up periods. Tumor size is a further aspect of heterogeneity. We speculated that, and obtained evidence for the concept that, each subgroup may describe differing disease causation pathways, host susceptibility factors, environmental exposures, prognosis, and possibly treatment options.

Identification of these subgroups enabled us to present a hypothetical model incorporating susceptibility factors and a variable disease risk threshold for BCC [3]. Figure 5 demonstrates how age-related changes in immune surveillance may influence the numbers of BCC: single tumors or clusters (presentation phenotypes observed). Reduced immune surveillance leads to crossing the BCC threshold and is associated with the release of pre-existing microtumors and a visible, initial BCC. Recovery of immune surveillance can also influence phenotype and outcome by immune surveillance rising above the threshold. Continuous, and hence, more severe loss of surveillance with the patient remaining below the threshold leads to a higher rate of BCC development often seen as clusters. The threshold level is not considered as absolute, but a variable level is specific for each individual and is influenced by other risk factors. This generic model could be applied to other mechanisms and presentation subgroups.

The above model can be used to consider the relationship between clinical heterogeneity and risk factors in any chronic pathology. Thus, Horton et al. in 2009 published a similar scheme showing subgroups, risk factors, and disease thresholds for CHD [6]. Subgroups recognized were FH (both homozygous and heterozygous) and “loss of function” PCSK9 states. The improved prognosis associated with loss of function PCSK9 mutation subgroup has given rise to PCSK9 inhibitors and therapeutic agents in which early trials show encouraging results.

Action Limits Replacing Reference Ranges in Patient Management

Since there will sometimes be profoundly different mechanisms determining causation and prognosis in subgroups, the use of a universal reference range for an analyte of interest may be inappropriate. The use of subgroup specific action limits may be a better approach. Evidence based on RCT’s has led to action limits...
replacing reference ranges in guidelines, regardless of whether a pathology is considered homogeneous (Fig. 2) or heterogeneous (Fig. 3). This move driven by clinical research shows the need for researchers to allow for heterogeneity in the design of trials.

We now consider two clinical areas at differing stages of acceptance of use of action limits. First, how low density lipoprotein cholesterol (LDL-c) action limits replaced reference ranges in lipid lowering, and second, male hypogonadism (HG) where the evidence base is less mature and reference ranges are still quoted and used when it may be more advisable to move to action limits.

Low Density Lipoprotein-Cholesterol and Cardiovascular Disease

Atherosclerotic obstruction of arteries by plaque formation leading to CVD such as myocardial infarction and stroke is one of the commonest causes of mortality in developed countries. Though the rate has been falling, CVD is still a leading cause of death in the United Kingdom [7] (Office for National Statistics, Deaths registered in England and Wales 2015). Various CVD risk factors have been identified. In 1948, the Framingham Heart Study followed 5209 adults aged 30–62 years to identify contributing factors associated with CVD [8]. Every 2 years, participants provided a detailed medical history and underwent a physical examination and laboratory testing. Since the initial phase, the study has expanded with two further generations being recruited. Age, male gender, elevated blood pressure, diabetes, smoking, total cholesterol (TC), and high density lipoprotein cholesterol ratio are associated with CHD, and accordingly, these factors were built into a primary prevention algorithm to estimate CVD risk in patients without established CVD. Association between lipids and CVD does not imply causation and it required intervention trials where lipid lowering (with more than one agent working via different pathways) led to a reduction in CVD risk before a causative role was established with a high degree of certainty.

The association between cholesterol and atherosclerosis led to the coronary primary prevention trial which demonstrated reduction of CVD following cholesterol reduction with cholestyramine [9]. This gave rise to the LDL hypothesis; LDL-c being a causative factor of atherosclerosis. Since the Scandinavian Simvastatin Survival Study (4S) in 1994, statins provide the mainstay of lipid lowering treatments [10]. Statins block hepatic cholesterol synthesis leading to upregulation of LDL receptors and increase in LDL particle uptake, thereby lowering serum LDL-c and TC. The evidence from 4S prompted us to move away from TC reference ranges and adopt an action limit of below 5.2 mmol/l for the coronary cholesterol clinic treating secondary prevention patients in the Manor Hospital, Walsall [11]. The significant CVD risk reduction observed in 4S was replicated in many other intervention trial studies [12–18].

Following these intervention trials, the Joint British Societies (JBS) and European Atherosclerosis Society recommended secondary prevention treatment targets of TC < 5 mmol/l and/or LDL-c < 3 mmol/l in 1998 [12,18]. More recent studies clearly demonstrate that “lower is better” regards TC and LDL-c levels [14,17,19]. The LDL hypothesis was given a boost by the SHARP [20] and IMPROVE-IT [21] studies where addition of ezetimibe, a LDL-c lowering agent with a mechanism of action different from statins, was demonstrated to reduce CVD risk as predicted by the LDL-c reduction seen in various statin trials. These intervention trials formed the basis of the JBS 2 guidelines [22]. They recommended treating with new optimal targets of TC < 4 mmol/l and/or LDL-c < 2 mmol/l with the previous targets, TC < 5 mmol/l.
and LDL-c <3 mmol/l, now viewed as minimal targets which were now adopted as lipid targets in the quality and outcomes framework, a primary care incentive scheme in the United Kingdom [23]. The revamped JBS 3 guidelines published in 2014 suggest that non-high density lipoprotein–cholesterol target of <2.5 mmol/l, considered equivalent to a LDL-c <1.8 mmol/l in secondary prevention and lifetime risk reduction as opposed to absolute risk reduction in primary prevention should also be considered [24]. Simplified versions of the Framingham risk algorithms were created for routine CVD risk estimation such as the Sheffield Tables in the UK [25]. Currently, the NICE guidelines (CG 181) advise primary care to use the QRISK2 risk calculator instead of the Framingham risk score. QRISK is a prediction system, similar

Fig. 4 Susceptibility, environmental and modifying factors leading to various BCC subgroups (1–n) defined by presentation, each with specific causative mechanisms (1–n)

Fig. 5 Scheme depicting how immune surveillance could determine BCC presentation phenotypes
to the Framingham risk score, which uses a combination of traditional risk factors (age, blood pressure, and cholesterol) and other factors (including family history, body mass index, and ethnicity). It has the advantage of being derived from data obtained from patients in the UK.

These algorithms estimating risk are not ideal in high risk individuals such as those with established CVD (secondary prevention) and FH. This indicates the presence of functional subgroups determined by risk and not etiology; greater benefit may be seen with lipid lowering treatment in patients at higher risk. Thus, action limits for lipid lowering treatment can differ. Heterozygous FH is the commonest single gene disease in the UK (1:500) [26]. If left untreated, around 50% of men and 30% of women with FH will develop CVD by the age of 50 and 60 years, respectively [26]. Most CVD risk factors are modifiable and treatable. These include high blood pressure, lipids, obesity, diabetes, physical activity levels, and tobacco usage. Regarding primary prevention, it is recommended that patients have their probability of CVD estimated via a risk calculator (QRISK 2) and statin treatment offered based on the underlying 10-year probability (NICE CG 181).4 The 10-year CVD probability treatment threshold has decreased from 30% [27] to 20% [22] and now to 10% [24]. Once again, population-based reference ranges have been superseded by more complex indices. Different approaches toward primary and secondary prevention are due to the inherent risk of the population being considered. Functional subgroups demonstrating varying therapeutic benefits have also been seen in CVD prevention.

There appears to be considerable residual CVD risk following statin treatment especially in cases with the atherogenic lipoprotein phenotype, the dyslipidemia characterizing the MetS [28]. Thus, in patients with MetS, it may be unwise to consider LDL-c in the same way as in other patients. Thus, a different approach to hyperlipidemia may yield benefits in patients with the MetS. The MetS in our opinion is poorly classified with defining variables dichotomized and many associated factors omitted [29]. Thus, we would consider that heterogeneity exists even within the MetS, which itself must be distinguished from other dyslipidemic patterns. Meta-analysis has suggested that fibric acid derivatives should only be used in patients demonstrating the atherogenic lipoprotein phenotype [1,30]. Thus, optimal management in dyslipidemia requires recognition of subgroups based on etiology, evidence, and underlying risk levels and customizing the treatment. This leads to a more sophisticated disease model with a requisite for greater knowledge and skill on the part of health care professionals and researchers.

Testosterone, Late Onset Hypogonadism and Testosterone Replacement Therapy

Late onset HG in men, as opposed to primary (testicular pathology) and secondary (pituitary/hypothalamic pathology), is defined as a combination of sexual symptoms and serum total testosterone levels <12 nmol/l [31]. Age-related late onset HG is seen in 6–12% of men and affects metabolic parameters associated with ill health and morbidity/mortality risk in adult men [32].

The European Male Ageing study of 2599 men aged 40–79 years (T2DM: 7%, follow up 4 years) showed hypogonadal symptoms and total testosterone levels <8 nmol/l, independently and in combination predicted overall and CVD mortality [33]. However, causality of the association could only be suggested if testosterone replacement therapy (TRT) led to risk reduction. Shores et al. investigated the effects of TRT in 1031 males aged over 40 years with total testosterone levels ≤ 8.7 nmol/l [34]. TRT in 398 of these men halved mortality (treated men: 10.3%, mean follow-up: 42.8 months; untreated men: 20.7%, mean follow-up: 38 months). When the patient cohort was stratified, the mortality reduction was significant in those with T2DM (HR:0.44, CI:0.23-0.84) but not in non-T2DM men (HR:0.72, CI:0.46-1.13). Both the association between late onset HG and mortality and mortality reduction following TRT in men with T2DM were confirmed by Muraleedaran et al. in 581 men with T2DM over 6 years follow-up [35]. The patient cohort was stratified by total testosterone levels of 10.4 nmol/l; 343 patients were classified as normal and 238 patients classified as having a low testosterone. In the low testosterone group, 174 untreated patients demonstrated significantly higher mortality compared to men on TRT after adjustment for covariates. We obtained similar results in 857 men with T2DM and late onset HG, using a total testosterone cut-off of 12.0 nmol/l or a free testosterone of 0.25 nmol/l [36,37].

While these studies do not allow definition of precise, cut-off values for testosterone that are associated with increased mortality in either diabetic or nondiabetic subjects, they do raise issues regarding what level should be considered as normal or healthy. Thus, laboratory testosterone measurement is recommended in men with HG and obesity [32]. There is a wide variation in reference ranges for total testosterone. The reference range of the Mayo Medical Laboratories for total testosterone is 8.3–33.0 nmol/l (240–950 ng/ml) in adult men and is not in line with the increasing benefits seen in some of the above cited longitudinal studies [34–37]. These studies suggest total testosterone thresholds should be < 8.7 nmol/l (250.9 ng/dl) in nondiabetic [34] and <10.4 nmol/l (300.0 ng/dl) [35] or even 12 nmol/l [36,37] in men with T2DM, when considering TRT. An understanding of the meaning of reference ranges and current evidence will enable a change toward action limits for testosterone as seen in lipidology [38].

Is It Time to Move Away From Population-Based Reference Ranges?

These two examples show areas with mature data (LDL-c) and accumulating data (testosterone) that allow derivation of action limits. Lipid lowering treatment is based on either action limits or overall CVD risk assessment. To obtain optimum benefit from therapy, clinicians need an understanding of disease mechanisms, up to date evidence from trials and other data such as meta-analyses as well as the drug pharmacology. Importantly in both clinical practice and research, CVD and most other diseases should be viewed as an end clinical state reached via various pathways with different risk factors. Thus, the population-based LDL-c reference range has been ignored by clinicians and researchers, although they are found in laboratory reports. Unlike in the case of testosterone, there are sufficient guidelines over the past 20 years for clinicians not to be influenced by the LDL-c reference range. This is unfortunately not the case with testosterone where evidence of benefit following TRT has not been widely disseminated. The lesson from lipid lowering is that considerable time is required to erode skepticism.

It is likely that heterogeneity exists within late onset HG. This can be, at least initially, identified by stratifying disease presentation patterns and causative factors. Following this, action limits for treatment will have to be periodically refined based on evidence.

Case History

The potential problem resulting from unconsidered use of a laboratory reference range is exemplified by the case of a middle-aged male who led an active and healthy lifestyle (unpublished). He gave a 12-month history of worsening fatigue, frequent hot flushes, loss of libido, erectile dysfunction, depression, and loss of mental focus, which are common symptoms of HG. He requested to measure the testosterone levels and his morning total testosterone one level was found to be 9.9 nmol/l. His request for TRT was refused as the total testosterone level was within the local laboratory reference range of 4–32 nmol/l. Desperation led to the patient sourcing testosterone enanthate and self-injecting every 10 days using a dose higher than those conventionally used for testosterone undecanoate. This treatment resulted in markedly improved

http://www.mayomedicallaboratories.com/test-catalog/clinical-rand+interpretive/83686
HG symptoms. Since then, he has contacted a specialist and is now receiving testosterone undecanoate administered under professional care. The challenge might be the achievement of long-term maintenance of testosterone levels above 15 nmol/l as recommended by expert guidelines [38], if primary care follow-up is dependent on once more laboratory reference ranges. This real life case demonstrates the danger of using reference ranges without consideration of clinical presentation(s) though we recognize that a single anecdotal example cannot determine clinical policy. However, there is already ample evidence showing that a total testosterone level below 15 nmol/l can be associated with symptoms [39] and that TRT in men with HG and total testosterone <12 nmol/l leads to clinical improvement of symptoms [40,41].

Conclusion

Clinical heterogeneity is a facet of the developing interest in personalized medicine. In our experience, research scientists (outside clinical biochemists) have a limited understanding of how a reference range is derived and used and of clinical heterogeneity. However, perhaps surprisingly, many medical undergraduate courses include little formal training on how to use chemical pathology services and reference ranges that are often viewed by newly qualified doctors as absolute markers of health or disease. Experienced clinicians of course recognize disease heterogeneity in their patients. Interaction between such clinicians and clinical scientists and their laboratories should improve the exchange of scientific and medical principles and use of research data. The benefits of this approach can be seen in various NICE guidelines for primary and secondary prevention of CVD as well as familial hypercholesterolemia. In our view, the MetS and the pathologies associated with it including late onset HG are clinical conditions that would benefit from this approach.

In conclusion, while we believe the identification and study of patient subgroups is important, we recognize that choosing the characteristic(s) used to classify subgroups is critical. Clearly, overenthusiastic use of multiple characteristics will result in relatively large numbers of subgroups comprising relatively few subjects making statistical analysis problematic. This problem will be compounded by using characteristics that are difficult to quantify precisely. Thus, the use of genetic factors (e.g., allelic variants) may allow more robust analysis than phenotypes such as skin type. This issue is of particular importance in the design of RCTs; inclusion criteria for the trial cohort should be based on a particular subgroup with a relevant outcome as primary outcome.

Issues of assay accuracy and precision also influence the reliability of clinical action limits. Thus, selecting an action limit for an analyte should reflect the coefficient of variation of the assay used to determine its concentration. Further, as different analytical methodologies may be used by laboratories, quality assurance schemes should be used to ensure that particular action limits are standardized.

Summary

It is evident that clinical heterogeneity exists in all pathologies and it seems reasonable to speculate that this may reflect different causal mechanisms and prognoses. We believe it would be helpful if intervention trials paid more attention to identifying possible subgroups on the basis of etiology or clinical presentation thereby allowing a complete understanding of the benefits of therapy (including customized therapy), and better options for researchers to unravel causative and prognostic mechanisms.

References


