EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

26 July 2018
EMA/810713/2017
Human Medicines Research and Development Support

Question and answer on the adequacy of the Mahalanobis distance to assess the comparability of drug dissolution profiles

Draft agreed by Biostatistics Working Party	June 2018
Adopted by CHMP	26 July 2018

Keywords	Bioequivalence, dissolution profiles, f2, Mahalanobis distance, biowaiver

The aim of this question-and-answer document is to provide clarification about the suitability of the Mahalanobis distance as a tool to assess the comparability of drug dissolution profiles and to a larger extent to emphasise the importance of confidence intervals to quantify the uncertainty around the point estimate of the chosen metric (e.g. the f2 factor or the Mahalanobis distance).

30 Churchill Place ● Canary Wharf ● London E14 5EU ● United Kingdom
Telephone +44 (0)20 3660 6000 **Facsimile** +44 (0)20 3660 5555
Send a question via our website www.ema.europa.eu/contact

Question:

Is the Mahalanobis Distance (MD) an adequate measure for use in the assessment of dissolution similarity, in particular in cases where the f2 statistic is not suitable?

Can interval estimation be used to inform decision making for the similarity of dissolution profiles based on an inferential statistical approach (with MD or other statistical measures)?

Answer:

As background to the comments provided below, it is considered important to note that in cases when f2 is considered suitable, i.e. can be used as outlined in Appendix 1 of the CHMP guideline on the investigation of bioequivalence [CPMP/EWP/QWP/1401/98 Rev. 1/ Corr **], guideline-compliant evaluation of dissolution similarity does not involve confidence interval estimation to decide upon similarity. The recommended decision criterion is based only upon the derived numerical value for f2 (point estimate ≥ 50)[1]. This means that the uncertainty related to the f2 sampling distribution is not accounted for. Against this background, the question concerning an adequate alternative statistical decision criterion, for cases where f2 should not be used, is difficult to answer. Some of the recently suggested alternative statistical approaches to measure the distance between two dissolution profiles involve an inferential element, i.e. the estimation of a confidence interval or region. Since f2 employed on its own does not have any inferential element, comparing these potential alternatives to the standard f2 criterion is hence not straightforward.

Regarding the MD, it is a dissimilarity measure between two random vectors x and y of the same length, which takes into account the correlations in the data set. MD is the multi-dimensional generalisation of the idea of expressing the distance between two points using standard deviation as the unit of measurement. This standardisation means that MD is dependent on variance and covariance estimates. In dissolution data sets, covariates generally correspond to dissolution percentages collected for different time-points. Under some assumptions, the MD becomes smaller, indicating similar dissolution profiles, with increasing variability observed in the data. This property makes its use undesirable for deciding upon similarity in dissolution, in particular with regard to the additional criterion that similarity limits should not be greater than a 10% difference at any time point is satisfied [CPMP/EWP/QWP/1401/98 Rev. 1/ Corr **]; depending on the variability observed it is quite possible to have an observed difference of over 10% at some time point, yet MD-based criteria could declare the difference to be unimportant.

Based on these considerations, the MD metric cannot be supported as a preferred methodological approach to decide upon similar dissolution, even in situations where the f2 statistic should not be used in the way outlined in the CHMP bioequivalence guideline.

[1] Another drawback of the f2 is that neither the shape of the dissolution profiles nor the time correlation is taken into account. Permuting the order of the time points would give the same f2 estimate, provided the time points in the test group are always compared to the same time points in the reference group. This means the f2 metric is not using all potentially relevant information available. Implications of this property on the sensitivity/specificity of the f2 decision criterion remain uncertain.

Any approach based upon confidence intervals for f2 would, however, be considered appropriate whether the validity criteria outlined in CHMP guidance are met or not [CPMP/EWP/QWP/1401/98 Rev. 1/ Corr **]. Similarity could then be declared if the confidence interval for f2 were entirely above 50. However, regardless of whether the conditions to adequately apply f2 in a dissolution experiment are fulfilled or not, the properties of the f2 sampling distribution do not allow the derivation of exact confidence intervals to adequately quantify the uncertainty of the f2 estimate. To address this, bootstrap methodology could be used to derive confidence intervals for f2 based on quantiles of re-sampling distributions, and this approach could actually be considered the preferred method over f2 and MD.