Assessing goodness-of-fit for evaluation of dose-proportionality

Martin J. Wolfsegger[1,9], Alexander Bauer[1], Detlew Labes[2], Helmut Schütz[3], Richardus Vonk[4], Benjamin Lang[5], Stephan Lehr[6], Thomas F. Jaki[7], Werner Engl[1], Michael D. Hale[8]

1. Baxalta Innovations GmbH, a Takeda company, Vienna, Austria.
2. Consultant, Berlin, Germany.
3. BEBAC – Consultancy Services for Bioequivalence and Bioavailability Studies, Vienna, Austria.
4. Bayer AG, Berlin, Germany.
5. Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach, Germany.
6. Austrian Medicines and Medical Devices Agency.
7. Department of Mathematics and Statistics, Lancaster University, Lancaster, UK.
8. Shire, a Takeda company, Lexington, MA, USA.
9. Corresponding author. Donau-City-Straße 7, 1221 Vienna, Austria.
   Phone: +43 1 20100 247 2997; E-mail: martin.wolfsegger@takeda.com

Running head: Assessment of Dose-Proportionality

Key words: dose-proportionality, linear pharmacokinetics, visual predictive checks

Abstract: For the clinical development of a new drug, the determination of dose-proportionality is an essential part of the pharmacokinetic evaluations. Including goodness-of-fit evaluations of the applied statistical models in dose-proportionality considerations may be able to provide early indications of non-linear pharmacokinetics and identify sub-populations with divergent clearances. We propose the use of simulation based visual predictive checks as goodness-of-fit examinations to improve the validity of dose-proportionality conclusions for complex designs. We provide an illustrative example and include a table to facilitate review by regulatory authorities.

1. Introduction

In early clinical development, the disposition of new drugs determines the feasibility of its further development. The assessment of dose-proportionality is of great clinical importance for predicting the consequences of rational dose adjustments [1]. It is of importance to identify a lack of proportionality since a moderate change in the bioavailability (BA) could have a large impact on the efficacy and safety for narrow therapeutic index drugs. Depending on the dosing range of the drug, non-proportional properties could challenge the use of the drug in clinical practice and may lead to the need of more extensive studies, for example in the context of bioequivalence.

Essentially, a dose-proportionality assessment is performed to evaluate whether exposure increases proportionally with the dose. One can reasonably predict drug concentration for different dosing scenarios if the pharmacokinetics (PK) can be demonstrated to be dose-proportional. One of the key questions to be answered in a dose-proportionality assessment is whether clearance is constant across the intended range of doses, which allows computation of an appropriate dose amount (i.e. dose per dosing interval) and dosing interval required to maintain an average steady-state plasma concentration which provides therapeutic benefit. Another critical element of a dose-proportionality assessment is that the data from all subjects are assumed to represent a homogenous population.

The aim of the manuscript is to provide an overview of prior research on evaluation of dose-proportionality, add useful descriptions and definitions of these methodologies and illustrate how they are applied in practice. We further discuss a tool, the visual predictive check, which is often used for other pharmacokinetic applications, but which seems to have not been used for proportionality assessment, and which hasn't received much attention in the statistical literature.

Several analysis methods have been proposed to assess dose-proportionality. Examples are analysis of variance of dose-normalized PK metrics (see section 2 below), simple linear regression or a power model approach [2-5]. The power model is usually the preferred model of choice and used as a basis for decision making [4, 6]. The decision may be made by simply estimating the degree of non-proportionality or via an equivalence testing approach. Both approaches have been subject to research [e.g., 2 and 3], although mainly in the context of a parallel group design. In clinical practice, however, there is a lack of consensus on overarching standardized rules for the analysis and reporting of dose-proportionality assessments.

The concepts of linear pharmacokinetics, dose-proportionality and dose-linearity are closely related. To put these into context, we first explain the term linear pharmacokinetics and subsequently examine the concept of and consequences of dose-proportionality, before a brief description of dose-linearity. The remainder of the paper is dedicated to establishing a generic framework for the assessment of dose-proportionality. This comprises parallel group designs, but also more complex trial designs such as cross-over or alternate panel designs. We provide recommendations for the analysis and the assessment of goodness-of-fit and for reporting of results from such an investigation. Finally, we illustrate the proposed framework with a practical study example and provide the corresponding code for the statistical software R [7].

2. Terminology

Pharmacokinetic characteristics in early stages of drug development are typically quantified and described using a non-compartmental (NCA) approach. The corresponding estimated quantities (e.g., C_{max}, AUC_{0-t}, etc) are referred to as PK metrics instead of PK parameters since the latter are elements of a PK model which are not directly observable, and which describes concentration as a function of time and other factors.

Linear pharmacokinetics theoretically requires that all transport processes (e.g. absorption, distribution, elimination) can be described by first order kinetics, in which the instantaneous rate of change of concentration depends only on the current concentration (i.e., the rate of change for concentration is proportional to the current concentration). Therefore, transport processes which are not described by first order kinetics will lead, in general, to non-linear pharmacokinetics.

On the premise of linear pharmacokinetics:

70   • Changing dose by the factor k leads to the corresponding change of any individual concentration by the
71   same factor k at any given time point. This means that PK metrics which are simple linear functions of
72   concentration, such as area under the curve (AUC), maximum concentration (C_{max}), and trough
73   concentration (C_{trough}) change proportionally with dose

74   • PK metrics such as time to reach maximum concentration (t_{max}), apparent terminal half-life ($t_{1/2}$), systemic
75   clearance (CL), volume of distribution at steady state (Vss), mean residence time (MRT), and
76   bioavailability (F) are independent of dose, as may be surmised by examination of the units of
77   measurement

78   • Concentrations after repeated dosing can be predicted, by the principle of superposition, from
79   concentrations following a single dose (e.g., $AUC_{0-\tau}$ at steady state equals $AUC_{0-\infty}$ after a single dose
80   where τ represents the constant dosing interval).

81   The premise of linear pharmacokinetics may be shown to be unsupportable when exposure is not proportional to
82   dose. However, demonstrating dose-proportionality alone does not prove linear pharmacokinetics. For example,
83   there might be situations with linear clearance (i.e. concentration independent) although plasma or tissue binding,
84   or distribution may be non-linear. In addition, dose-proportionality or dose-linearity assessments are exposure
85   analyses that are based on three or more dose levels using exposure metrics obtained with a non-compartmental
86   approach (NCA) utilizing empirical models. These kinds of analyses are based on empirical observations rather
87   than on theoretical pharmacokinetic properties and cannot provide details on the different PK processes which
88   might impact dose-proportionality or dose-linearity.

89   ### 3.   Concept of dose-proportionality

90   Dose-proportionality over a dose range after a single administration can be assessed empirically by the power
91   model [2-4] which has the following form:

92    $$Y = \mu \times d^{\beta} \tag{1}$$

93   where Y represents the exposure metric of interest (e.g., maximum concentration C_{max}, area under the curve from
94   zero to the last time point with a quantifiable concentration $AUC_{0-tlast}$, area under the curve from zero to infinity
95    $AUC_{0-\infty}$) and d the corresponding dose administered. On the premise of dose-proportionality, the parameter of
96   interest β equals 1 (i.e., the relationship between dose and exposure is given as a straight line passing through zero
97   on the ordinate). Lack of dose-proportionality can be due to many mechanisms (e.g. limited solubility will result
98   in a less than proportional increase) but is typically due to the saturation of some components in the system (e.g.
99   under proportionality for saturable absorption and over proportionality for a saturable metabolism process [8]). A
100   comprehensive discussion regarding mechanisms leading to lack of dose-proportionality is provided in [9] and
101   will therefore not be further discussed.

102   Note that when we discuss proportionality it is necessary to specify a range over which proportionality will be
103   assessed. This is because proportionality from zero to infinity is not physically plausible, and for practical
104   applications we just need a defined range of proportionality. Assessing dose-proportionality over a dose range can
105   be considered as an equivalence hypothesis problem and can be assessed utilizing the power model based on
106   certain margins of equivalence [2,3]. In this setting, the corresponding acceptance regions for the parameter β from
107   the power model can be derived based on conventional margins of (bio)equivalence ranging from $\theta_L = 0.8$ to
108    $\theta_U = 1.25$ [2] or based on a more lenient acceptance criterion ranging from $\theta_L = 0.5$ to $\theta_U = 2.0$ arguing that the
109   conventional margins are impractically strict for large dose ranges [3]. Acceptance regions can be derived as
110   follows [2,3]:

111    $$\left(1 + \frac{ln(\theta_L)}{ln(r)}\right) < \beta < \left(1 + \frac{ln(\theta_U)}{ln(r)}\right) \tag{2}$$

112   where r corresponds to the dose ratio investigated relating acceptance regions to the dose range. It follows that the
113   acceptance region is more stringent for a large dose range than for a narrow dose range. If the calculated two-sided
114   90% confidence interval for the estimated slope β falls completely within the acceptance region, dose-

115 proportionality over a dose range can be claimed statistically at the 5% level of significance [10] which allows a
116 yes/no decision in the case that margins were specified a-priori. If the calculated two-sided 90% confidence interval
117 (CI) for the estimated slope β doesn't completely fall within the acceptance region, the highest dose $\hat{D}_{max}$ can be
118 calculated such that the two-sided 95% CI for β is included entirely within the pre-specified margin [2] by:

119 $$\hat{D}_{max} = D_{min} \times \theta_U^{\frac{1}{max(1-L,U-1)}}$$ (3)

120 where D_{min} is the lowest dose studied and L and U the lower and upper limits of the two-sided 90% confidence
121 interval for slope β, respectively. Extrapolation beyond the studied dose range is typically not recommended.

122 A definitive assessment of confirmatory dose-proportionality is generally conducted during later phases. The
123 corresponding sample size planning is discussed in [11] and can be performed by function `power.dp` of R
124 package `PowerTOST` [12].

125 For the possibility of considering a waiver for investigation of additional dose amounts in the context of
126 bioequivalence studies, the corresponding EMA guideline [13] suggests an equivalence margin for dose-
127 normalized AUCs of ±25% which would allow evaluating bioequivalence at the highest dose only where for other
128 doses a waiver is possible via a dissolution similarity exercise. Of note, the power model has a direct relationship
129 to dose-normalized AUCs because on the premise of dose-proportionality the equation Y/d=a=constant holds true.
130 Consequently using ±25% as margin for dose-normalized AUCs leads therefore to margins of $\theta_L = 0.75$ to
131 $\theta_U = 1/0.75$ for the estimated slope β. It should be noted that analysis of dose-normalized exposure metrics using
132 analysis of variance (ANOVA) may be considered an inefficient use of available data since this approach handles
133 the dose information as a categorical variable and therefore may have less statistical power compared to the power
134 model which explicitly utilizes dose information as a continuous variable [4].

135 It is commonly assumed that C_{max}, $AUC_{0-tlast}$, $AUC_{0-\infty}$ are log-normally distributed [14] with a proportional
136 increase of the variance with the size of the PK metric (i.e., constant coefficient of variation rather than constant
137 variance). For the situation at hand (i.e., assumed log-normally distributed exposure metrics with a proportional
138 error), the transform-both-sides approach [15] using the natural logarithm results in the following functional
139 relationship between dose and exposure:

140 $$ln(Y) = ln(\mu) + \beta \times ln(d) + \varepsilon$$ (4)

141 where ε represents a normally distributed residual error with zero mean and a common variance.

142 Dose-proportionality over a dose range after repeated administration at steady state can also be assessed based on
143 PK metrics $C_{max;ss}$, $C_{trough;ss}$ and $AUC_{0-\tau;ss}$ grounded on this concept. However, it should be noted that evaluation
144 of $C_{trough;ss}$ can be difficult in case of values below the lower limit of quantification (LLOQ) Also, $AUC_{0-\infty}$ is
145 typically not calculated at steady state since it would not be interpretable whenever accumulation occurs, and
146 therefore one would need to carefully consider and justify the use of this metric for evaluation of dose
147 proportionality after repeated administration.

148 As already mentioned, one of the key questions to be answered in a dose-proportionality assessment is whether
149 clearance is constant across the selected dose range. Apparent clearance is calculated as dose/AUC, so that AUC
150 can be used to examine whether apparent clearance appears to be a constant across the dose range. However, there
151 can be situations where AUCs may appear to be dose proportional but individual concentration time curves are
152 not. This is the motivation for evaluating dose-proportionality also on other exposure metrics such as C_{max} or
153 C_{trough}. C_{max} is important regarding the potential for safety concerns which may arise if one might experience of
154 unexpectedly high peak exposure. Likewise, C_{trough} may be a critical consideration for efficacy, such as FVIII for
155 hemophilia patients, where maintenance of a suitably high C_{trough} is important.

156 Lastly, as alluded to above, most compounds can be expected to exhibit a non-linear relationship between dose
157 and exposure when administered over a sufficiently large range (i.e., at extreme doses). Therefore, one should only
158 assess dose-proportionality for the clinically relevant or defined dose range, although a broader dose range is
159 typically investigated in early pharmacological studies since a final recommended dose range is not yet known or

160 decided upon. Another point to consider is that drugs may be evaluated at a large dose range during clinical
161 development, but much smaller dose ranges may apply in various situations, such as for different indications. For
162 instance, intravenous immunoglobulins are used at doses between 0.4 to 2 g/kg body mass [16], but narrower dose
163 ranges are used for replacement therapy (0.4–0.8 g/kg) and immunomodulating therapies (1–2 g/kg). More
164 generally the recommended dose range may vary for different countries, by age, by body weight, by hepatic or
165 renal impairment, for concomitant medications, or by other factors which may impact the dose-exposure
166 relationship.

167 In contrast to dose-proportionality, dose-linearity after a single dose study implies that the relationship between
168 dose and exposure (e.g., AUC) is given as a straight line starting on the ordinate at any value greater or smaller
169 than zero. This concept is therefore applicable for a) endogenous compounds implying starting on the ordinate at
170 a value greater than zero or b) compounds that are invariably lost upon administration at low dose levels (e.g., a
171 low dose of a drug which is given far in excess of a receptor) starting on the ordinate on a value lower than zero.
172 Consequently, an alternate non-linear modeling approach, such as model 4 in [17], may need to be considered as
173 a basis for a description of the underlying relationship between dose and exposure.

4. Assessing Goodness-of-fit

175 Prior to making any conclusions regarding dose-proportionality, the goodness-of-fit of the model must be assessed
176 to evaluate the model performance which is usually done using graphical evaluation. These assessments should
177 include residual diagnostic plots (i.e., plotting residuals versus predicted). A comprehensive model evaluation
178 should also include plotting the residuals (y-axis) against potential covariates (x-axis) which may represent a sub-
179 population with different drug clearance profiles (e.g. race, ethnicity, sex, age, disease status, phenotypes, etc). If
180 these figures show any unacceptable trends (i.e. not symmetrically scattered residuals around zero), inclusion of
181 covariates in the model or sub-group analyses should be considered.

182 An important goodness-of-fit plot is a scatter plot of the observed dose-exposure values superimposed with the
183 model-predicted dose-exposure relationship, along with corresponding confidence and prediction intervals [2] on
184 linear-linear and log-log scales. However, the latter goodness-of-fit plot brings along some difficulties when more
185 complex designs are used. This section provides a solution to the difficulties that goodness-of-fit plots exhibit for
186 more complex designs.

187 As mentioned above, the power model can be straightforwardly applied to exposure metrics estimated from a
188 parallel group design using the following simple linear regression model:

189 $$ln(Y_i) = ln(\mu) + \beta \times ln(d_i) + \varepsilon_i \qquad (7)$$

190 where Y_i represents the exposure metric of interest and d_i the corresponding dose for subject i. Model parameter μ
191 represents the intercept, i.e. the average $ln(Y)$ value for $ln(d) = 0$, and β the slope, i.e. the increase of $ln(Y)$ for an
192 increase of 1 unit in $ln(d)$. The residual error ε is a normally distributed random variable with mean zero and
193 common variance. A plot of the observed dose-exposure values superimposed with the model predicted dose-
194 exposure relationship, along with corresponding confidence and prediction intervals, can be derived easily.

195 However, dose-proportionality is frequently assessed using a more complex design, such as higher order cross-
196 over designs, which in turn requires the application of more complex models, with additional parameters taking
197 the design features into account. We generally recommend a balanced design which requires that 1) each dose
198 occurs only once with each subject, 2) each dose occurs the same number of times in each period and 3) the number
199 of subjects who receive dose i in some period followed by dose j in the next period is the same for all $i \neq j$.

200 Alternate panel designs (e.g., single sequence k period design) do not meet the properties of a balanced design and
201 require additional unverifiable premises such as of no period effect and is typically evaluated using a linear mixed
202 effects model with log-transformed dose as a fixed effect and subject as a random effect. However, Williams'
203 designs (i.e., special cases of orthogonal Latin squares design) meet these properties of a balanced design and the
204 following equation shows how to model the dose-exposure relationship for a higher order cross-over design [18]:

205 $$ln(Y_{ijk}) = ln(\mu) + \beta \times ln(d_{ijk}) + p_j + s_k + \gamma_{ik} + \varepsilon_{ijk} \qquad (8)$$

206   where μ represents the intercept parameterized as the average $\ln(Y)$ value for $\ln(d) = 0$. Model parameter p_j
207   represents the fixed effect for period j and s_k the fixed effect for sequence k. The effect for subject nested in
208   sequence γ for subject i in sequence k can be modeled as fixed effect or as random (intercept) effect representing
209   a normally distributed random variable with mean zero and common variance. Of note, the model in (8) fits a
210   common slope to all subjects where the model can be easily expanded by modeling a normally distributed random
211   slope θ_{ik} with mean zero and common variance to get individual slopes for each subject.

212   The estimated dose-exposure relationship now depends on additional fixed effects for period and sequence (and
213   subject) preventing an overall assessment of goodness-of-fit by plotting observed dose-exposure values
214   superimposed with the estimated dose-exposure relationship and corresponding confidence and prediction
215   intervals.

216   One possibility to overcome this issue is to use simulation based visual predictive checks (VPCs) to assess
217   goodness-of-fit of the model. The general concept of VPCs is to assess graphically whether simulations from a
218   model can reproduce the central trend as well as the variability in the observed data where Duffull *et al.* [19] may
219   be the earliest recognizable use of VPCs with their figures 4 and 5. In a first step, many replicates of the original
220   dataset are simulated from the model. Percentiles of interest for each simulated dataset are calculated and used to
221   generate non-parametric confidence intervals for the predicted percentiles which are then superimposed with the
222   observed percentiles.

223   In situations with pronounced period and/or sequence effects, it may be helpful to perform this graphical
224   assessment based on a period and sequence-corrected exposure metric Y^{cor} and corresponding simulations to get a
225   better understanding of the underlying inter-subject variability:

226 $$Y_{ijk}^{cor} = exp\big(ln(Y_{ijk}) - p_j - s_k\big) \tag{9}$$

227   Typically, such studies are based on fixed dose groups and, therefore, observed and simulated percentiles can be
228   straightforwardly calculated per dose group. If fixed doses are administered but body mass adjusted doses are of
229   interest, it is possible that dose groups overlap after adjustment, and data needs to be grouped together by binning.
230   If predictions within a bin differ largely due to different values of other independent variables, prediction corrected
231   VPCs (pcVPC) should be employed [20]. Of note, VPCs are often used for complex models, such as population
232   pharmacokinetic-pharmacodynamic (Pop PK/PD) modelling. We apply the model to a simpler log-transformed
233   power model to illustrate its use in dose-proportionality assessments in the next section.
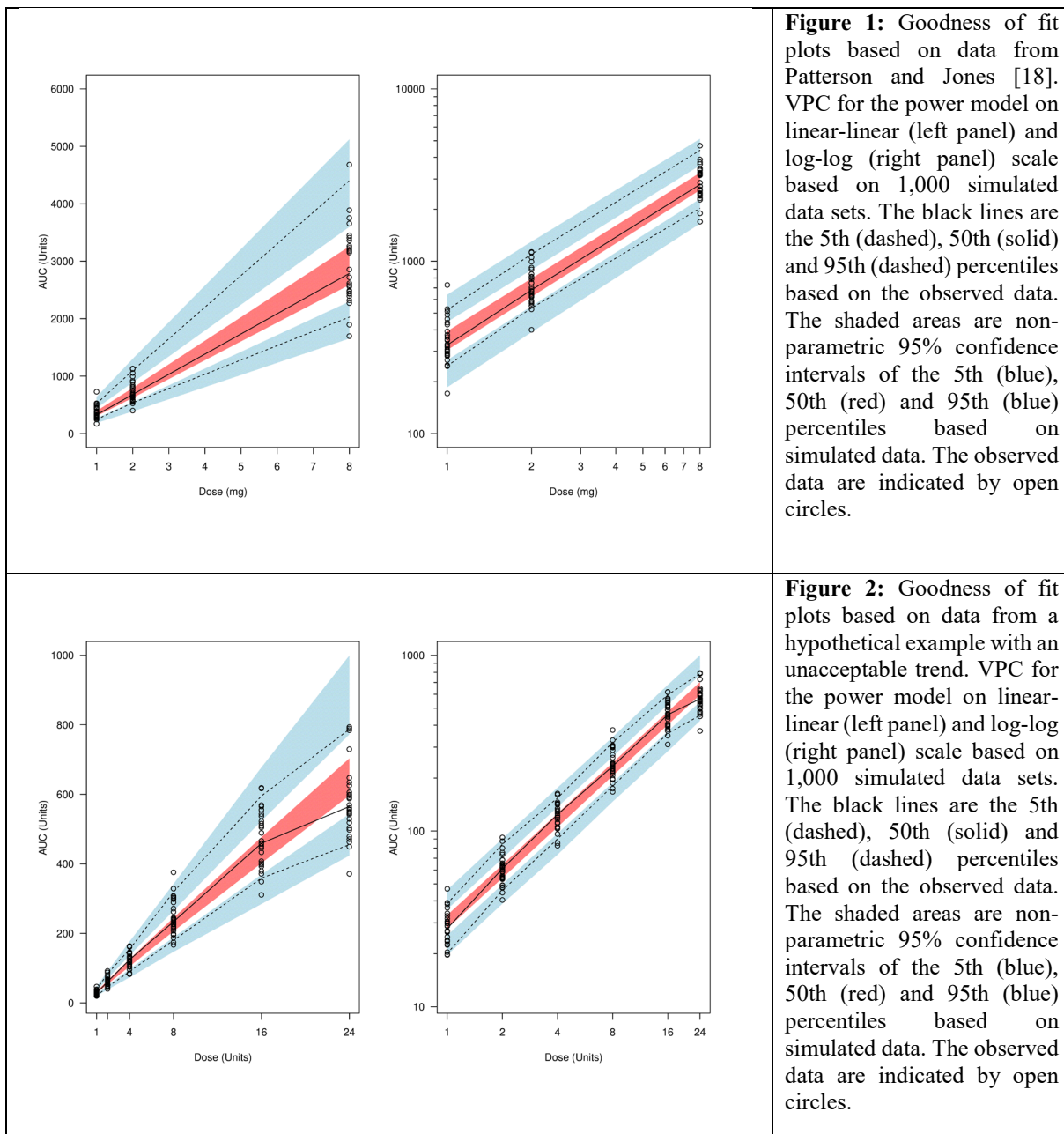
234   **5. Illustrative Example**

235   The following illustrative example was taken from Patterson and Jones [18]: A randomized cross-over study in 28
236   normal healthy volunteers was performed to assess dose-proportionality (and the effect of food) using a Williams'
237   design (4 treatments, 4 sequences, 4 periods). Fixed doses of 1, 2, and 8 mg were administered after an overnight
238   fast, with administration of each dose separated by a washout period. The fourth treatment was a fixed dose of 8
239   mg with a meal instead of the overnight fast. Data of the latter regimen was not used for the assessment of dose-
240   proportionality.

241   A linear mixed effects model using R function `lmer` of R package `lme4` [21, 22] was fitted with fixed effects for
242   period and sequence and a random (intercept) effect for subject nested in sequence. Corresponding VPCs for AUC
243   on linear-linear and log-log scale based on 1,000 simulated data sets are shown in Figure 1. The corresponding R
244   code for different goodness-of-fit plots is given as online supplement.

245   Figure 1 shows no unacceptable trends since observed percentiles of interest were well within simulated non-
246   parametric 95% confidence intervals. Using the critical margins of 0.8 and 1.25 [2], dose-proportionality for the
247   dose range investigated (1 to 8 mg) is judged to be satisfactory at the 5% level of statistical significance since the
248   calculated two-sided 90% confidence interval for the estimated slope β falls completely within the critical region
249   (Table 1).

250   An additional alternate numerical model evaluation could be based on the power-model predicted geometric means
251   per dose level as well as observed geometric means and their corresponding ratios.

252   An illustrative example with an unacceptable trend based on simulated data is shown in Figure 2. The power model
253   fails to reproduce the central trend of the data. The corresponding VPC shows the discrepancy of the model-based
254   non-parametric confidence intervals with the percentiles of the observed data for the highest dose investigated
255   indicating an approximately 20% smaller AUC than predicted by the model.

256

Figure 1: Goodness of fit plots based on data from Patterson and Jones [18]. VPC for the power model on linear-linear (left panel) and log-log (right panel) scale based on 1,000 simulated data sets. The black lines are the 5th (dashed), 50th (solid) and 95th (dashed) percentiles based on the observed data. The shaded areas are non-parametric 95% confidence intervals of the 5th (blue), 50th (red) and 95th (blue) percentiles based on simulated data. The observed data are indicated by open circles.

Figure 2: Goodness of fit plots based on data from a hypothetical example with an unacceptable trend. VPC for the power model on linear-linear (left panel) and log-log (right panel) scale based on 1,000 simulated data sets. The black lines are the 5th (dashed), 50th (solid) and 95th (dashed) percentiles based on the observed data. The shaded areas are non-parametric 95% confidence intervals of the 5th (blue), 50th (red) and 95th (blue) percentiles based on simulated data. The observed data are indicated by open circles.

257

6.   Reporting of Results

259   In the event that the goodness-of-fit plots shows unacceptable trends, any conclusions based on the power model
260   are questionable and alternative models to describe the dose-exposure relationship can be utilized. In such cases
261   we suggest investigating covariates to be included in the power model where influential covariates and their
262   relationship to the exposure metric can be identified based on residual plots showing the relationship between
263   residuals (y-axis) and covariates (x-axis).

264 However, in case that the fitted basic power model does not show unacceptable trends and to facilitate
265 interpretation of the dose-proportionality assessment in terms of clinical relevance, we recommend estimating the
266 increase in the exposure metric per doubling of the dose [4]:

267 $$\text{Increase per doubling of dose} = exp(ln(2) \times \beta) \tag{10}$$

268 In addition to the factor for increase in PK metric, the size of deviation from dose-proportionality can be readily
269 calculated from the power model and using the dose ratio (r) via [2,9]:

270 $$\text{Deviation from dose-proportionality} = r^{\beta-1} \tag{11}$$

271 As already mentioned, many compounds are expected to exhibit a non-linear relationship between dose and
272 exposure when administered at extreme doses, which encourages a formal assessment of dose-proportionality for
273 the a-priori specified anticipated clinical dose range in addition to the complete dose range investigated. We found
274 a table like those presented in [2] useful for reporting the corresponding results.

275 For the purpose of an exploratory assessment of dose-proportionality, acceptance regions for the slope based on
276 [2] and/or [3] and/or [13] could be added as an additional column or footnote to the table to facilitate review by
277 regulatory authorities. Table I shows results for the illustrative example where the deviation from dose-
278 proportionality is 4% (calculated as described in equation (11) as $8^{1.02-1} = 1.04$) regarding extent of absorption (i.e.
279 the AUC is about 4% higher at an 8-fold increased dose than expected under dose-proportionality).

Table I: Exploratory assessment of dose-proportionality based on the power model

Dose range studied	PK metric[*]	Model predicted geometric mean values[*] for dose range studied	Estimated slope (90% CI)	Increase in PK metric per doubling of dose (90% CI)
1 to 8 (mg)	AUC	346 – 2,901	1.02 (1.00 – 1.04)	2.03 (2.00 – 2.06)
	C_{max}	75 – 588	0.988 (0.958 – 1.02)	1.98 (1.94 – 2.03)

Acceptance region for the estimated slope as described in equation 2 according to Smith et al. [2]: 0.893 to 1.107 (±20%); EMA guideline [13]: 0.862 to 1.138 (±25%); Hummel et al. [3]: 0.667 to 1.333 (±50%). [*] Units of AUC and C_{max} not available for the example in Patterson and Jones [18]

7. Discussion and Conclusion

281 Assessment of dose-proportionality over a dose range is an essential part of characterizing the PK properties of a
282 drug during the development process. The dose-proportionality assessment is typically an exposure analysis of
283 three or more dose levels based on exposure metrics obtained with the non-compartmental approach (NCA)
284 utilizing an empirical model such as the power model. For this reason, a dose-proportionality assessment is based
285 on empirical observations rather than on theoretical pharmacokinetic properties of the molecule at hand.
286 Essentially, this kind of exposure analysis is aimed to evaluate whether exposure increases proportionally with
287 dose where one of the key questions to be answered in a dose-proportionately assessment is whether clearance is
288 "apparently" constant across an appropriate dose range. Dose-proportionality assessment can therefore be seen as
289 a basic tool to help decide whether and which more advanced investigations are called for. This could be to use
290 more sophisticated modeling approaches to better understand and characterize the compound or to collect
291 additional data.

292 In addition, different equivalence margins have been proposed which may limit the ability to satisfy different
293 regulatory agencies of the appropriateness of the chosen values to support an "apparently" constant clearance
294 across the recommended dose range; a globally acting sponsor company could easily end up with different
295 decisions depending on the jurisdiction. The resulting uncertainty clearly has major potential implications on the
296 internal decision- making process and induces unnecessary uncertainty in the "appropriate" margin. For this
297 reason, choosing the strictest margins should therefore be considered in a confirmatory framework.

However, we believe that, particularly in early drug development phases when the range of clinically relevant doses is not yet completely clear, the goodness-of-fit plots as well as the confidence intervals are more informative and important than the comparison to pre-specified margins of equivalence. This is because they provide information on 1) the adequacy of the model which serves as the basis for any further considerations, 2) the impact of the uncertainty of the estimates and 3) the possible size of exposure deviation from dose-proportionality for the dose range studied.

An adequately powered and carefully conducted dose-proportionality assessment, including model evaluation based on goodness-of-fit plots, might be able to provide insights into possible non-linear pharmacokinetics and/or sub-populations with divergent clearances. This is difficult to achieve in complex designs and this article proposes simulation based VPCs to address this unmet need.

8. Declarations

9. References

325

326 1. Calvo E, Zafar H, Goetz A, Bonate P, De Bono J, Patnaik A, Fourezesh B, Tolcher A, Rowinsky E,
327   Takimoto C. Analysis of dose proportionality testing methods in phase I clinical trials of anticancer
328   agents. Cancer Res. 2005;65(9 Supplement):973–74.

329 2. Smith BP, Vandenhende FR, DeSante KA, Farid NA, Welch PA, Callaghan JT, Forgue ST. Confidence
330   Interval Criteria for Assessment of Dose Proportionality. Pharm Res. 2000;17(10):1278–83. doi:
331   10.1023/A:1026451721686.

332 3. Hummel J, McKendrick S, Brindley C, French R. Exploratory assessment of dose proportionality:
333   Review of current approaches and proposal for a practical criterion. Pharmaceut. Statist. 2009;8(1):38–
334   49. doi: 10.1002/pst.326.

335 4. Gough K, Hutchison M, Keene O, Byrom B, Ellis S, Lacey L, McKellar J. Assessment of Dose
336   Proportionality: Report from the Statisticians in the Pharmaceutical Industry / Pharmacokinetics UK Joint
337   Working Party. Drug Inf J. 1995;29(3):1039–48. doi: 10.1177/009286159502900324.

338 5. Senn S. Statistical Issues in Drug Development. 2nd ed. Chichester: John Wiley & Sons; 2007. p. 345-
339   347.

340 6. Ezzet F, Spiegelhalter D. Pharmacokinetic dose proportionality: practical issues on design, sample size
341   and analysis. In: 2nd International Meeting on Statistical Methods in Biopharmacy. Société Francaise de
342   Statistique, Paris, 1993.

343 7. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical
344   Computing, Vienna, Austria. 2019. URL https://www.R-project.org/.

345 8. Sierakowki B. Study designs tailor-made for different pharmacokinetic trials. In: Cawello W. Parameters
346   for Compartment–free Pharmacokinetics. Standardisation of Study Design, Data Analysis and Report.
347   Aachen: Shaker Verlag; 2003. p. 127–144.

348 9. Eisenblaetter T, Teichert L. Dose Linearity and Proportionality. In: Vogel HG, Maas J, Gebauer A,
349   editors. Drug Discovery and Evaluation: Methods in Clinical Pharmacology. Berlin, Heidelberg:
350   Springer; 2011. p. 35-40.

351 10. Schuirmann DJ. A Comparison of the Two One–Sided Tests Procedure and the Power Approach for
352   Assessing the Equivalence of Average Bioavailability. J Pharmacokin Biopharm. 1987;15(6):657–680.
353   doi:10.1007/BF01068419.

354 11. Sethuraman VS, Leonov S, Squassante L, Mitchell TR, Hale MD. Sample size calculation for the Power
355   Model for dose proportionality studies. Pharmaceut. Statist. 2007; 6:35–41. doi: 10.1002/pst.241.

356 12. Labes D, Schütz H, Lang B. PowerTOST: Power and Sample Size Based on Two One-Sided t-Tests
357   (TOST) for (Bio)Equivalence Studies. R package version 1.4 7: 2018. https://cran.r-
358   project.org/package=PowerTOST.

359 13. European Medicines Agency, Committee for Medicinal Products for Human Use. Guideline on the
360   Investigation of Bioequivalence. London: European Medicines Agency; January 2010.
361   CPMP/EWP/QWP/1401/98 Rev. 1/ Corr **.

362 14. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug
363   Evaluation and Research, and Center for Biologics Evaluation and Research. Guidance for Industry:
364   Statistical Approaches to Establishing Bioequivalence. Rockville: January 2001.

365 15. Bonate PL. Pharmacokinetic–Pharmacodynamic Modeling and Simulation. 2nd ed. New York: Springer;
366   2011. p. 146.

367 16. Lee M, Strand V. Intravenous Immunoglobulins in Clinical Practise., Boca Raton: CRC Press; 1997.

368 17. Chow S-H, Liu J-p. Design and Analysis of Bioavailability and Bioequivalence Studies. 3rd ed. Boca
369   Raton: Chapman & Hall/CRC; 2009. p. 564.

370   18. Patterson S, Jones B. Bioequivalence and Statistics in Clinical Pharmacology. 2nd ed. Boca Raton:
371   Chapman & Hall/CRC; 2017. p. 234.

372   19. Duffull SB, Chabaud S, Nony P, Laveille C, Girad P, Aarons L. A pharmacokinetic simulation model for
373   ivabradine in healthy volunteers. Eur. J. Pharm. Sci. 2000;10:285-294. doi: 10.1016/s0928-
374   0987(00)00086-5.

375   20. Bergstrand M, Hooker AC, Wallin JE, Karlsson MO. Prediction-Corrected Visual Predictive Checks for
376   Diagnosing Nonlinear Mixed–Effects Models. AAPS J. 2011;13(2):143–151. doi:10.1208/s12248-011-
377   9255-z.

378   21. Bates D, Maechler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. J Stat
379   Softw. 2015;67(1):1–48. doi:10.18637/jss.v067.i01.

380   22. Bates D, Maechler M, Bolker B, Walker S, Christensen RHB, Singman H, Dai B, Scheipl F, Grothendieck
381   G, Green P, Fox J. Linear Mixed-Effects Models using 'Eigen' and S4. R package version 1.1 21: 2019.
382   https://cran.r-project.org/package=lme4.

383