Incorporating random effects in biopharmaceutical control strategies

Thomas Oberleitner[1] , Thomas Zahel[2], Marco Kunzelmann[3], Judith Thoma[3] and Christoph Herwig[4*]

Abstract

Objective Random effects are often neglected when defining the control strategy for a biopharmaceutical process. In this article, we present a case study that highlights the importance of considering the variance introduced by random effects in the calculation of proven acceptable ranges (PAR), which form the basis of the control strategy.

Methods Linear mixed models were used to model relations between process parameters and critical quality attributes in a set of unit operations that comprises a typical biopharmaceutical manufacturing process. Fitting such models yields estimates of fixed and random effect sizes as well as random and residual variance components. To form PARs, tolerance intervals specific to mixed models were applied that incorporate the random effect contribution to variance.

Results We compared standardized fixed and random effect sizes for each unit operation and CQA. The results show that the investigated random effect is not only significant but in some unit operations even larger than the average fixed effect. A comparison between ordinary least squares and mixed models tolerance intervals shows that neglecting the contribution of the random effect can result in PARs that are too optimistic.

Conclusions Uncontrollable effects such as week-to-week variability play a major role in process variability and can be modelled as a random effect. Following a workflow such as the one suggested in this article, random effects can be incorporated into a statistically sound control strategy, leading to lowered out of specification results and reduced patient risk.

Keywords Biopharmaceutical manufacturing, Process validation, Process characterization study, Random effects, Mixed-effects model, Likelihood model

Introduction

Biopharmaceutical manufacturers have a regulatory need to accurately describe production processes and to underpin design choices and control strategies with reports based on sound science. The registration application for new drug substances and their corresponding products includes a detailed description of the manufacturing process and a justification for the proposed control strategy (ICH, 2011; ICH, 2008). The ICH defines a control strategy as follows:

> *A planned set of controls, derived from current product and process understanding, that assures process performance and product quality. The controls can include parameters and attributes related to drug substance and drug product materials and components, facility and equipment operating conditions, in-process controls, finished product specifications, and the associated methods and frequency of monitoring and control.*

*Correspondence:
Christoph Herwig
christoph.herwig@tuwien.ac.at
[1] Competence Center CHASE GmbH, Vienna, Austria
[2] Körber Pharma Austria GmbH, PAS-X Savvy, Vienna, Austria
[3] Boehringer Ingelheim Pharma GmbH & Co. KG, Analytical Dev. Biologicals, Biberach an der Riss, Germany
[4] TU WIEN Research Area Biochemical Engineering, Vienna, Austria

Oberleitner *et al. AAPS Open* (2023) 9:4

Page 2 of 13

The development of this control strategy, or *process characterization*, constitutes a major part in the first stage of the FDAs process validation guideline (FDA, 2011) and is often the most elaborate and critical phase of development for a new drug product. Ideally, this process should yield detailed knowledge about the individual parts of the process, i.e., critical quality attributes (CQA), impact of process parameters (PP), and sources of variability. The level of understanding of the product and its production process also affects the regulatory process, as stated in the ICH Q8 guideline (ICH, 2017):

> *A greater understanding of the product and its manufacturing process can create a basis for more flexible regulatory approaches. The degree of regulatory flexibility is predicated on the level of relevant scientific knowledge provided in the registration application.*

The FDA's 2011 guide defines process validation as "the collection and evaluation of data" over the life cycle of the product, from product development to commercial production, in order to establish scientific evidence that "a process is capable of consistently delivering quality product." Part of this is detecting and understanding different sources of variation affecting the production process (FDA, 2011). This is especially important in stage 1 of the process validation activities, i.e., in the design phase where the effects of process parameters (PPs)/material attributes (MAs) and their impact on product quality are quantified. Design of experiments is recommended as an effective tool to achieve the following:

> *Design of Experiment (DOE) studies can help develop process knowledge by revealing relationships, including multivariate interactions, between the variable inputs (e.g., component characteristics or process parameters) and the resulting outputs (e.g., in-process material, intermediates, or the final product).*

DOE followed by linear regression for modelling relationships between PPs, MAs, and CQAs are common tools employed in biopharmaceutical development (ICH, 2017). The assumption of linear relationships, i.e., models being linear in their parameters and not necessarily linear in their prediction of a factor, is generally valid for sufficiently small regions around a known working point (Montgomery, 2017). However, this should be carefully evaluated, e.g., by performing residual analysis of the derived models.

Process parameters are generally modelled as fixed effects that are assumed to be distributed around the true parameter value, i.e., $E\left(\hat{\beta}_j\right) = \beta_j$. Typically, any-

thing actively controlled by an operator might be considered a fixed effect, e.g., the temperature within a reactor, pH values, and feeding rates.

Random effects, in contrast, are parameters that are not controllable in such a way. Their future setting in an experiment or run cannot be predicted beforehand. However, they can still impact product quality and should therefore be considered when identifying possible sources of variability. MAs can fall within this category. Examples for random effects are changing raw material attributes, transition conditions, or biological variability in seed trains or even variability introduced by different operators.

The FDA guide (FDA, 2011) states the following:

> *The functionality and limitations of commercial manufacturing equipment should be considered in the process design, as well as predicted contributions to variability posed by different component lots, production operators, environmental conditions, and measurement systems in the production setting.*

In most cases, a random effect affects a group of runs, which is called a block and the random effect a "block factor" (Montgomery, 2017). A typical example could be an experimental setup using a specific raw material lot for a set of runs. Statistical analysis is used to investigate the impact of those blocking effects. The resulting measure of variance is sometimes called *inter-block variability*, as opposed to *intra-block variability, which constitutes* the residual error term of the individual observations.

When processes are modelled in silico as regression models, block factors are usually incorporated as deviation-encoded fixed effects. In its most common form, deviation encoding describes individual blocking effects by their distance from the overall mean of the response (Alkharusi, 2012). When no block information is provided for a prediction, its numeric value is set to zero, and the mean response for the "average" block is computed, which is often the desired behavior. This, however, does not account for inter-block variability, and thus, the overall variability of the model is underestimated. As estimators of variability are used to compute proven acceptable ranges (PARs), this underestimation can have a large effect on the accuracy of such measures.

To illustrate this point, consider the method for calculating the PAR shown in Fig. 1. The plot shows CQA values as function of the process parameter screening range. The slope of the line in the center indicates the effect of the parameter on the CQA value. A measure of variability around the predicted mean is given by a statistical interval (confidence, prediction, or tolerance interval) shown as dashed lines. The range of the PAR (gray area) can be calculated by finding intersection points between the
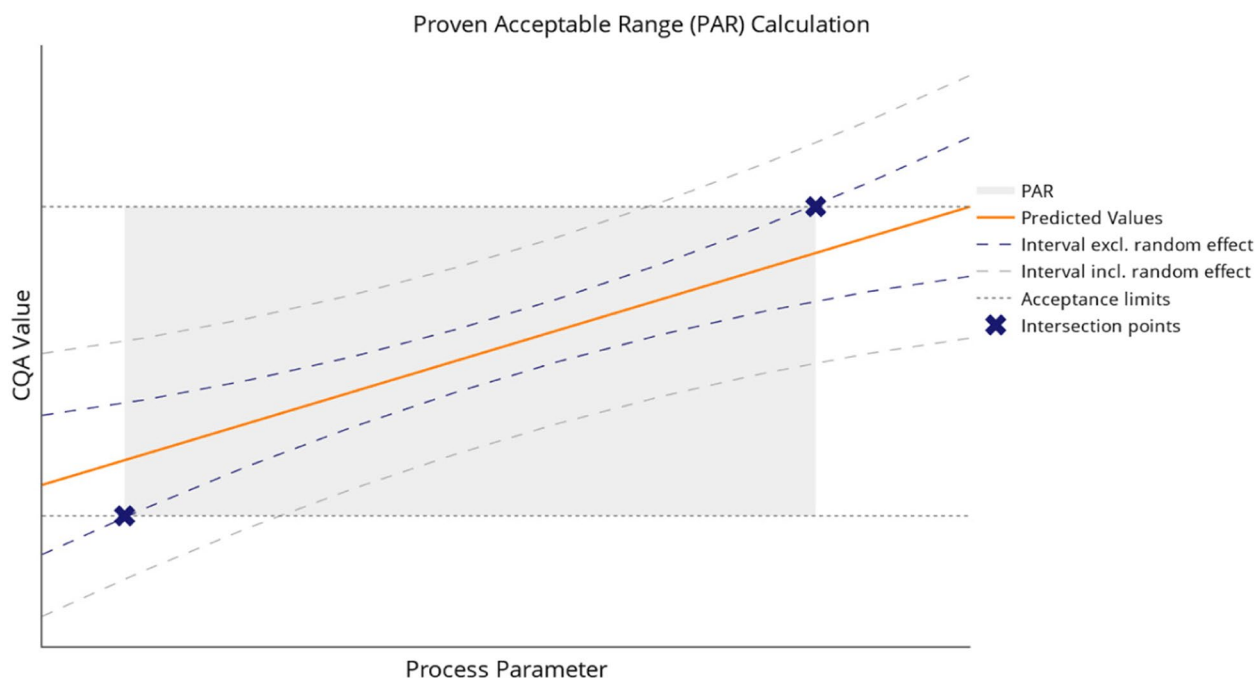
Proven Acceptable Range (PAR) Calculation

Fig. 1 An example for how the PAR of a process parameter can be calculated. The intersection points of a statistical interval and the CQA acceptance criteria define the lower and upper boundary. Note that the PAR that ignores the random effect is larger than the one that incorporates the random nature of the effect

acceptance limits and the statistical interval. When taking into account the additional variance introduced by random effects, the interval gets wider, and consequently, the PAR gets smaller. To find PARs for each critical process parameter in relation to the CQA acceptance criteria in drug substance, either each unit operation can be analyzed individually or integrated process modelling, enabling a holistic control strategy (Zahel et al., 2017), can be employed.

To accurately describe the random nature of those blocks, linear mixed models (LMM) can be employed to incorporate multiple sources of variation. In particular, the variation of random blocks can be computed separately and added to the overall variation of the model's prediction. Burdick et al. briefly illustrated the statistical methods behind LMMs and how they could be applied in process validation stage 1 (process design) in general (Burdick et al., 2017).

Goos et al. contrasted the conclusions drawn from OLS and LMM models in industrial split-plot designs and provides some guidance on analyzing experiments that involve random effects. The paper gives a motivating example and highlights some technical details of the method, like the proper choice of degrees of freedom (Goos et al., 2006). While pitfalls in improper experimental design and analysis are explained in a general manner, our work puts, for the first time, the problem in the

context of biopharmaceutical process development and provides a detailed workflow that considers characteristics specific to the domain.

Usually, the potential impact of fixed effects such as pH, temperature, or oxygen concentration is assessed in risk assessment such as failure mode and effects analysis (FMEA) and then investigated experimentally. The main contribution of this article is to illustrate in a case study how strong random effects can be in comparison with these fixed effects. Moreover, we introduce a workflow to incorporate random effects into process characterization in a statistically sound manner.

Following a workflow such as the one proposed is important in order to increase knowledge for the next round of risk assessment and experimental planning. Therefore, we aim to investigate the relative importance of random effects using a real-world data example of a full process characterization data set generated at Boehringer-Ingelheim over multiple unit operations of a drug substance production process including upstream and downstream.

Materials and methods
Models
To contrast ordinary least squares (OLS) and linear mixed model (LMM)-based approaches for creating a control strategy, both model types were fit to the process

characterization study (PCS) data. For OLS models, the random effect was treated as categorical fixed effect. While mean predictions are equivalent in this application, only LMMs decompose variance into a random and residual part, enabling the calculation of more accurate statistical intervals. This is especially important in the context of control strategies where patient safety could be at risk as a consequence of intervals that are too narrow, i.e., optimistic. A detailed description of OLS models and LMMs can be found in (Montgomery et al., 2021) and (SAS Institute Inc., 2010), and key differences and formulas are summarized in Additional file 4.

Statistical intervals

Statistical intervals represent an important and widely used tool to calculate and visualize uncertainty in data, estimators, or predictions in regression models. The most well-known type of interval is the confidence interval, which expresses uncertainty around the models' response, i.e., a confidence boundary around the predicted mean that contains the true population mean to a nominal level of confidence. Prediction intervals expand on this idea and add a standard deviation to the interval to define the region where a single, new observation is expected to fall within. To cover a nominal percentile of the actual distribution rather than a single observation, a third type of interval is used: the tolerance interval. Tolerance intervals cover the area that contains a predefined proportion of the true distribution of a response, often called coverage, to a nominal level of confidence. As we are interested in this true distribution of a modelled response, here in the form of critical quality attributes, tolerance intervals are used for the definition of the control strategy. Different techniques can be found in

literature to account for variance components in the calculation of intervals. Here, we use the method proposed by Franzq et al. (Francq et al., 2019) to include the random contribution to variance estimated by linear mixed models. See Additional file 4 for a more detailed description of intervals and formulas used in the evaluation of the case study data.

Note that a tolerance interval converges to a prediction interval as the degrees of freedom increase. Figure 2 illustrates this effect while comparing the widths of the different types of intervals on simulated data. For this example, OLS-based intervals were computed. However, the relative widths and the effect of the degrees of freedom are the same when using LMM-based intervals.

Manufacturing process

The case study was performed at Boehringer-Ingelheim as a PCS of a monoclonal antibody process. The process consists of typical steps such as fermentation of a cell culture, harvesting, protein A column, intermediate and polishing column, and an UFDF step. Generally, the pool (output) of a unit operation is used as the load (input) of the next unit operation, so that the overall production process can be seen a sequential chain of operations. The result of this chain is the actual drug substance, i.e., the product whose critical quality attributes are expected to fall within a predefined range, the so-called drug substance specifications.

Case study and data analysis

In a PCS conducted at Boehringer-Ingelheim, we investigated the impact of process parameters on 11–22 CQAs in each of the eight unit operations (UO). Models were created that regress a CQA in a unit operation onto
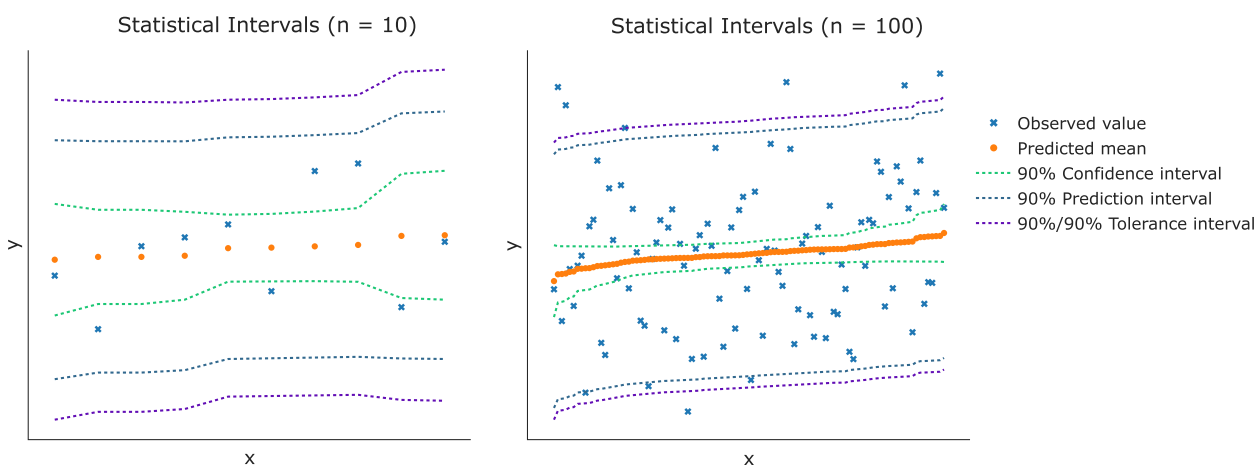

Fig. 2 Comparison of different statistical intervals and sample sizes. For conceptual and mathematical reasons, prediction intervals are always wider than confidence intervals when the same confidence level is assumed. Note that the tolerance interval converges to a prediction interval as the degrees of freedom are increased

factors found significant in the model selection process. The models were fit using data acquired in one-factor-at-a-time (OFAT) and design of experiments (DOE) runs using bench-scale experiments that are representative for the manufacturing process (see Additional file 1). Representativeness has been achieved via a pre-conducted small-scale qualification. During these activities, scale independent geometrical and engineering principles have been ensured as well as the absence of major performance differences between the scales.

The designs for DoE runs were planned in a way that minimizes aliasing and correlation between parameters (D-optimal), and that makes sure that effects can be detected with adequate statistical power. The DoE design and power analysis were conducted for fixed effects using the statistical software JMP (version 14.0.0). For the a priori power analysis, all runs (DoE + OFAT) were used. Moreover, a full model that includes all interaction and quadratic effects was assumed, which represents the worst possible case in terms of power. By convention, a power value of at least 0.8 is recommended. The power values for finding significant effects within two or three standard deviations of the residual error around the setpoint are reported in Table 1. A common critical value of $\alpha = 0.05$ was used as the significance threshold. Note that at the time the PCS was conducted, the power analysis was not explicitly conducted to incorporate random effects; see "Random effects in power analysis" section for an explanation of our approach. We performed variable selection (best subset selection or stepwise bidirectional) to eliminate nonsignificant effects and create more parsimonious models. Data for each unit operation consisted of one random effect describing the week-to-week variability across the experiments.

Good modelling practice was employed to check for model quality after variable selection. This was done by residual analysis to check for normality of residuals, inspecting model parameter *p*-values to determine if they exceed a threshold of 0.1, checking whether the RMSE is within expected reproduction variability and thereby mitigating the risk of overlooking effects as well as overfitting, leave-one-out cross-validation to exclude biasing the model via single runs. These measures increase the confidence that neither a substantial type 1 error (including effects that are not significantly different from 0) nor type 2 error (overlooking effects) has been made. The latter also implies that no aliasing is expected, which may bias a fixed/random effect. In general, we followed the approach to data analysis and model creation outlined as "workflow B" in the "Workflow B: modelling random effects using linear mixed models" section. The evaluation of effects in "Effect sizes and variances" section is summarized as a series of box plots that show the fixed and random effect sizes as well as variance ratios for each model (Figs. 3 and 4). To create a comparable measure of effect size, the original data used to fit the model were min-max normalized based on the parameter screening ranges. This means that all the input parameter values lie within the interval $[-1, 1]$, and their effects after fitting are comparable within the model.

$$x^* = 2\frac{x - \min(x)}{\max(x) - \min(x)} - 1 \tag{1}$$

As the response values in the training data were not normalized and used in their original scales, the effects were additionally divided by the root-mean-square error (RMSE) to make them comparable between models. For each model, an average measure of fixed and random effects was computed using their absolute values.

$$\overline{\beta}_{CQA} = \frac{1}{p-1}\sum_{i=1}^{p-1}\left|\frac{\hat{\beta}_i}{\hat{\sigma}_\epsilon}\right| \tag{2}$$

$$\overline{\gamma}_{CQA} = \frac{1}{m}\sum_{i=1}^{m}\left|\frac{\hat{\gamma}_i}{\hat{\sigma}_\epsilon}\right| \tag{3}$$

Table 1 Average statistical power over all effects to detect an effect within 2 or 3 standard deviations from the set point. Note that those values represent the worst case that assumes a full model, i.e., a model that includes all quadratic and interaction effects

UO	Power for 2 SD	Power for 3 SD	Runs DoE/OFAT	RE levels	Significant effects	Responses
UO 1	0.91	0.99	24/6	3	16	22
UO 2	1.00	1.00	27/24	4-5	30	20
UO 3	0.95	1.00	17/5	5	10	14
UO 4	0.71	0.93	18/3	6	10	11
UO 5	0.94	1.00	21/3	5	15	17
UO 6	0.99	1.00	42/11	3	36	13
UO 7	0.82	0.96	0/6	5	3	11
UO 8	0.89	0.98	0/13	5	8	15

where $p-1$ is the number of parameters minus the intercept, m the number of levels of the random effect investigated, and $\hat{\sigma}_\epsilon$ represents the estimator of the residual variance, i.e., the RMSE. The intercept is excluded in the calculation of $\overline{\beta}_{CQA}$ as only the effect of actual model parameters should be measured. A distribution of those values per unit operation is illustrated as box plots in Fig. 3. The figure highlights the random effects contribution to overall variance and compares it to the fixed effect contribution.

A similar approach was taken with the variance ratios in Fig. 4. Per CQA model, the ratio $\hat{\sigma}_\gamma / \hat{\sigma}_\epsilon$ was calculated, and the distribution of values is shown as box plots per unit operation.

Results

In the analysis of the case study data of real industrial data from a PCS, we contrast the OLS and LMM-based method to forming statistical intervals. As a random effects contribution to observed variance is proportional to its effect size, we first compare normalized estimators of fixed and random effects. This helps us to identify how strong random effects are in comparison with well-known fixed effects, such as pH and temperature. We then show how this random contribution increases the tolerance intervals and, in turn, reduces the PAR in an example picked from the case study.

Effect sizes and variances

The random effect investigated in the analysis of the case study data was week-to-week variability. To show effect sizes of the random effect predictors relative to fixed effects and their contribution to variability, LMMs were fit to the data. Figures 3 and 4 show the effects and variances per unit operation. It can be seen that the random effect is even greater than the fixed effect in some of the unit operations. Ignoring the random effects' impact on variability would underestimate the size of statistical intervals and result in an inappropriate control strategy. In four of the eight unit operations (UO3, UO5, UO7, and UO8), the median standardized random effect size was larger than the median standardized fixed effect size (see Fig. 3). In the other UOs, the random effect size is approximately the same as the fixed effect. Moreover, for six out of eight UOs (UO2, UO3, UO4, UO5, UO7, UO8), the median variance ratio of random versus residual variance is equal or larger to one (see Fig. 4). Effect sizes are of course dependent on the experimental design the data is based on. However, while variation of random effect estimates might be larger than those of fixed effects, a general trend is clearly discernible in our results (see Fig. 3), and significance of both random and fixed effects was checked in the model selection process.

LMM random effect predictors are often described as the empirical best linear unbiased predictors (EBLUP)
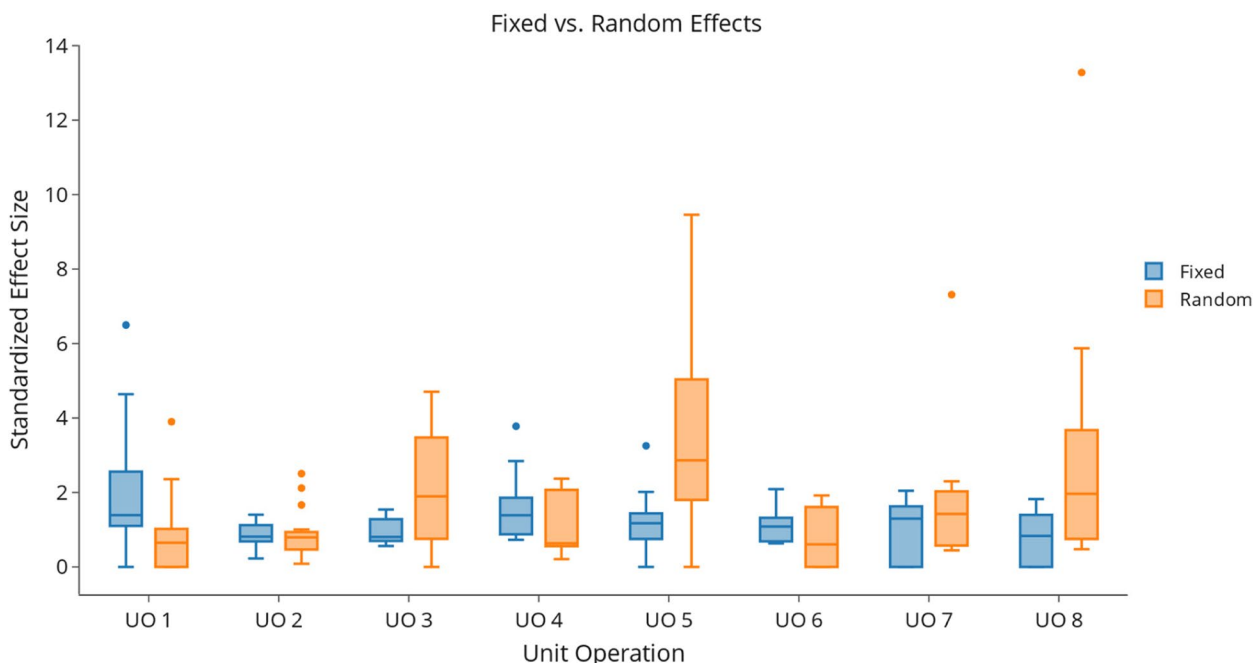

Fig. 3 Standardized fixed and random effect sizes are contrasted for each unit operation. A unit operation contains models for 11–22 CQAs, and their respective fixed and random effect distributions are shown as box plots. To create comparable measures of effect size, normalized data were used to fit the models, and the effects were divided by the RMSE. Note that for several unit operations, the median random effect is even larger than the median fixed effect
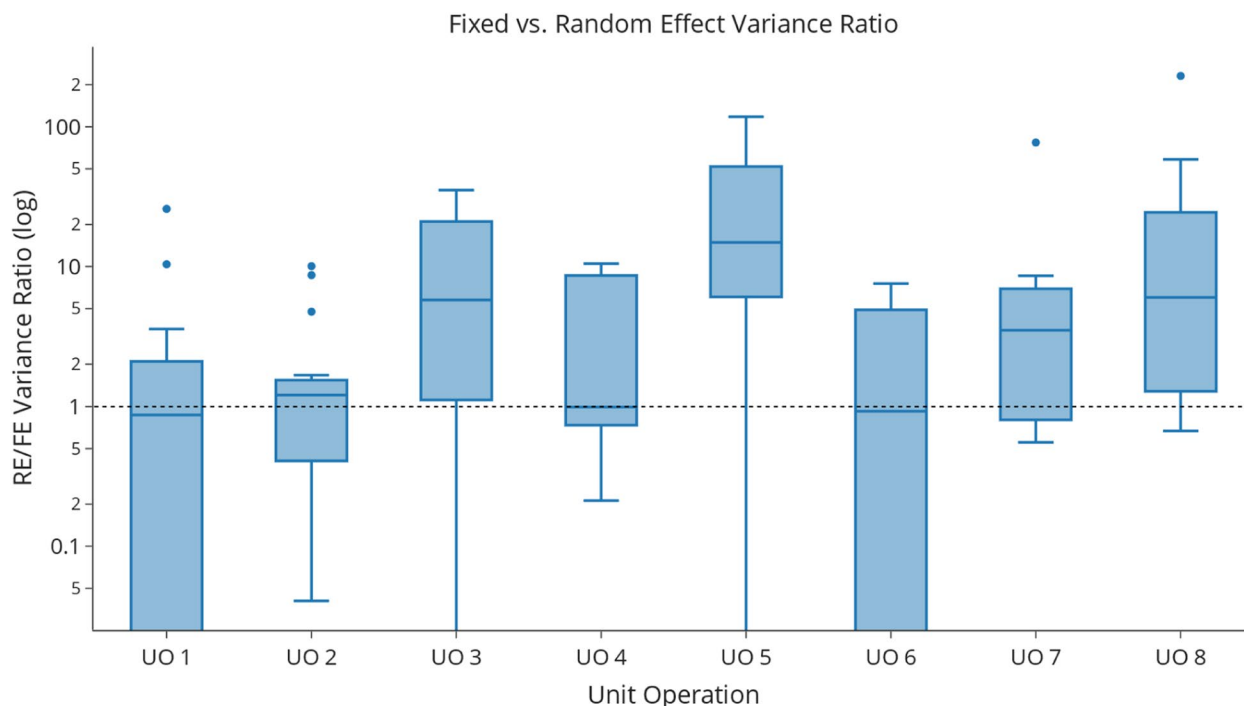
Oberleitner *et al. AAPS Open*   (2023) 9:4

Page 7 of 13

Fig. 4 Variance ratios (random variance/residual variance) are shown per unit operation on a logarithmic scale. For each model in a unit operation, the ratio between random and residual variance is calculated and the resulting distribution illustrated as a box plot. As random effect size increases, so does its contribution to variance — in some cases, the random contribution to variance is many times as large as the residual variance

in literature and yield more accurate effect sizes when compared to those obtained by modelling them as fixed effects using OLS (Govaerts et al., 2020). Due to the way they are calculated in mixed models (see Additional file 4), they tend to be closer to zero. This should be considered when comparing random and fixed effect sizes in Fig. 3. The amount by which their effect size is "shrunk" is inversely proportional to the associated variance component, i.e., the smaller the random effects variance, the larger the amount of shrinkage and vice versa. Figure 4 shows that in our case study, random effect variance is quite large relative to that of fixed effects, indicating that effect sizes based on EBLUPs should not differ substantially from those of obtained from modelling random effects as fixed effects using OLS. Moreover, our overall message that random effects are equally or more influential in a representative process characterization would be even more pronounced calculating out the shrinkage.

Tolerance intervals

Modeling a random effect as a categorical, fixed effect using OLS models is an often-employed practice in biopharmaceutical manufacturing. Here, we show the implications of this approach in a representative example picked from a real-world case study. Assuming a normal distribution of residuals, the chosen tolerance interval

should contain at least 90% of observations in 50% of repeated samplings. However, this was not the case. As illustrated in Fig. 5, in extreme cases, the OLS tolerance interval almost never included the value observed in the runs. This was due to variability introduced by different blocking factors in the production process. The larger the blocking effect, the larger its influence on variability — a quantity that is ignored in the OLS case. Incorporating random effect variability by employing LMM models and appropriate interval calculation methods solves this problem, which can be seen in the outer interval in Fig. 5.

Further analysis revealed that this observation was not an exception, but that the data for most CQAs included significant blocking effects that would result in a tolerance interval too narrow when ignored. Table 2 gives an overview of interval width ratios $r = (TI_{LMM, upper} - TI_{LMM, lower})/(TI_{OLS, upper} - TI_{OLS, lower})$ for the most common CQAs at setpoint conditions. Depending on the random effect size, the LMM intervals can be several multiples as wide as their OLS counterparts, when ignoring the random effect.

PAR and control strategy

The general increase in tolerance interval width when incorporating random effects and LMMs reported in the previous section can have a considerable impact on
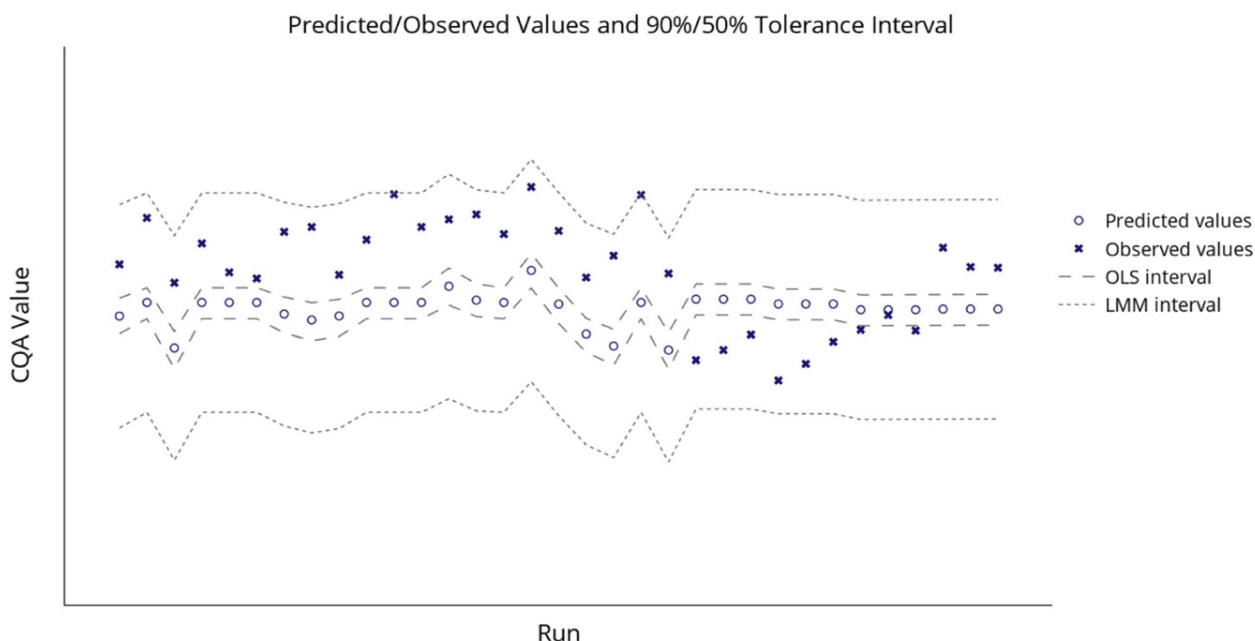
Oberleitner *et al. AAPS Open*   (2023) 9:4

Page 8 of 13

Fig. 5 A 90%/50% tolerance interval is created around the mean. By definition, it should include 90% of the data in 50% of cases, which is obviously not the case when using an OLS model. However, the interval computed using variance information from the LMM model does indeed cover at least 90% of observations

Table 2 LMM/OLS tolerance interval width ratios for 6 of the most common CQAs per unit operation. Due to the strong random effect, the LMM interval is generally much wider than the OLS interval

	UO 1	UO 2	UO 3	UO 4	UO 5	UO 6	UO 7	UO 8
CQA 1	1.85	3.80	2.67	3.18	3.01	1.57	1.17	1.08
CQA 2	1.62	1.65	6.15	3.36	4.59	3.37	2.38	5.15
CQA 3		1.70	4.99	3.43	4.06	2.43	2.34	1.41
CQA 4	1.85	1.64	1.77	1.12	4.38		8.95	2.84
CQA 5		1.80	4.58	1.37	8.34		2.92	3.10
CQA 6		1.67		1.25	9.02		3.46	2.98

the control strategy. Again, a representative example is selected from the case study in the form of a process parameter. For the chosen parameter, both an OLS model and LMM was fit, and tolerance intervals were calculated using the corresponding methods (for models and data, see Additional file 3). As can be seen in Fig. 6, when using the intersection points of the interval with the upper acceptance limit, the resulting PAR for OLS is indeed wider than the one based on the LMM.

Depending on the size of the fixed effect and the chosen acceptance limits, the reduction of the PAR might be more or less severe. Generally, for PARs formed with the method illustrated, its size can only decrease with the increase of the interval width as indicated in Fig. 6.

Discussion

Workflows to establish a control strategy

As shown in the case studies presented in the "Results" section, random effects can have a large effect on statistical intervals, the PAR, and consequently the control strategy. Here, we propose a workflow for establishing a control strategy that incorporates random effects at various stages. We first suggest an OLS-based workflow typically used in the industry and then contrast this strategy with one that incorporates random effects using linear mixed models.
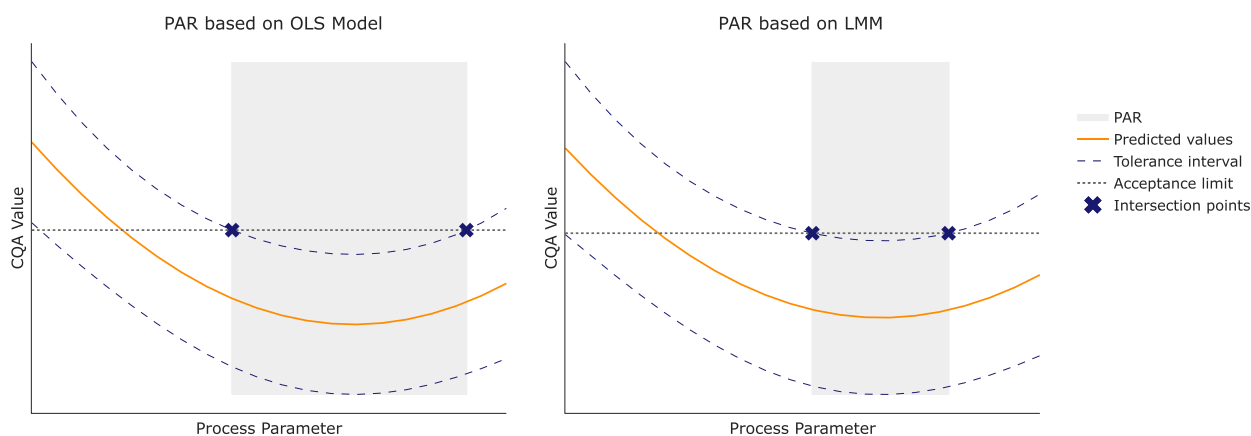
Fig. 6 PAR for a randomly picked parameter calculated from case study data. Due to the contribution of the random effect, the interval based on LMM variance components (right) is wider than its OLS counterpart (left). This results in smaller PAR (gray area) and a more conservative control strategy. In this example from the case study, the OLS PAR is 72% larger than the more conservative LMM PAR

Workflow A: modelling random effects as fixed effects using OLS

In the first step, the data that constitutes the basis for the regression model is acquired. We assume that data originates from a design that ensures desired properties for analysis, such as minimal correlation, minimal aliasing, and maximal power. In this phase, random blocking factors are identified alongside all other factors that might influence the response of the process, and their values are aggregated into a single data source that enables convenient analysis. This is followed by a preprocessing step where those data are cleaned up and response data possibly transformed to a form that satisfies OLS model assumptions (normality of residuals). Random effects are treated as categorical fixed effects and deviation-encoded so that the reference for the individual block coefficients is the grand mean of the response. This enables to set the blocks to zero for predictions, which results in a "mean block" prediction of the response. At this stage, a "full model" can be created by adding quadratic and interaction effects for each main effect. Given the available number of observations, use case, or preference, this full model can be used directly. Alternatively, the list of effects can be used as the input for a variable selection procedure to find a parsimonious model that explains the response while eliminating insignificant parameters. Such procedures are commonly based on estimators of prediction error, for example, the Akaike information criterion (AIC), or on *p*-values of model parameters. The implementation of such estimators depends on the type of model used, as they are different for OLS and LMM. Blocking factors might be found insignificant in the variable selection process and removed from the model. In the last step, either the full or optimal model

is used to compute the predicted values for the training data, whereby the predictor variables for the block are set to zero. Around those predictions, a tolerance interval is formed that contains a proportion of the population (coverage) with a certain probability (confidence). This should be reflected by the observed values of the training data contained in the interval. The PAR of the parameter is formed by the intersections of the interval with the acceptance criteria (see Fig. 1). However, such a tolerance interval based on OLS models does not incorporate the variance introduced by random effects correctly and might lead to control strategy that is too optimistic.

Workflow B: modelling random effects using linear mixed models

As it was the case in the first workflow, the LMM-based procedure starts by identifying both fixed and uncontrollable, random effects. Special attention is given to the latter as often multiple random factors are involved, which can be nested or crossed, both of which influences variance calculations in different ways. In addition to correlation analysis of fixed effects, some data prerequisites specific to LMMs should be checked to make sure the likelihood optimization converges, though this depends on the statistical software or library employed. Statistical significance of individual blocks potentially affects convergence and can be examined beforehand by deviation encoding them as described in in workflow A and investigating their effects using *p*-values obtained from an OLS fit. The levels of the random block variable as well as the number of intra-block observations are also factors in the optimization algorithm as highly imbalanced blocks can be the source of convergence problems. After making sure that the data meets all the criteria for applying a

LMM, a full model that includes quadratic and interaction effects can be created. Again, this full model can be used directly or as the input for variable selection where insignificant model parameters are eliminated in each step of the algorithm until the optimal model is found. In this workflow, variable selection can be performed in two different ways: one option is to deviation-encode random blocks and fit OLS models which are then used for evaluation in each step — essentially the same process as in the first workflow. The fixed effects from the final model are then used to transform the OLS model into an LMM. This can be a sensible workaround in situations where one is constrained by software lacking variable selection procedures that incorporate random effects. However, note that this approach might not be possible in some experimental designs. The second option is to use LMM-specific evaluation criteria in each step of the variable selection process. While this might be the most obvious approach, it is also not universally applicable, depending on the algorithm, criteria, or performance constraints. After a satisfactory model is found, predictions can be computed. For LMM, this means that only the matrix of fixed effects needs to be provided for the prediction as the model automatically assumes the "mean block" for the results. This is different from OLS models where blocks need to be set to zero explicitly for the prediction. The notable difference between the models is in

how model variance is computed and partitioned, which is important in the next step: the calculation of statistical intervals. Here, mixed models include a measure of variance of both fixed and random effects which, depending on the magnitude of the random effect, can widen the interval and therefore reduces the acceptable range of the process parameter.

Figure 7 summarizes the OLS- and LMM-based approaches to establishing a control strategy and provides an overview of their main differences.

The workflows outlined here represent two methods for computing PARs using common data-science techniques. Workflow A shows a common approach employed in biopharmaceutical manufacturing, while workflow B represents our proposal for an extended version that incorporates random variance correctly into statistical intervals. Given the unlikely scenario of a process being not affected by random effects at all, workflows A and B would result in the same control strategy.

Modelling scale impact

For the process characterization study described in this article, only data from bench-scale DoE and OFAT experiments were used, as no large-scale data was available at that point in time. Typically, manufacturing data is supplemented in the analysis to investigate the effects of scale. In regression models, this can be done by simply
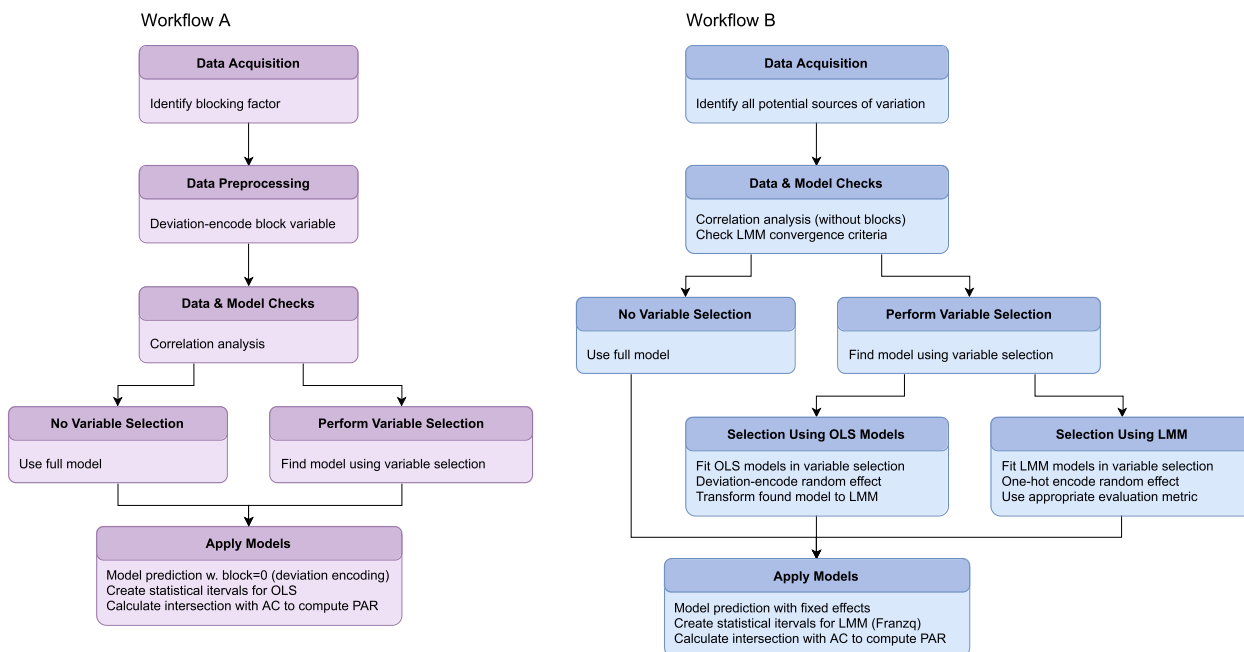

Fig. 7 Workflows for creating control strategies based on regression models. The left column describes an approach that uses OLS models for the estimation of PARs. A mixed-model-based workflow is summarized on the right. The differences in the steps involved are subtle but generally result in a more realistic estimation of variance and therefore a more robust control strategy

adding a categorical factor to the model with one level per data source (e.g., "large scale" and "DoE"). As a regular, fixed effect, such a factor can be subject to variable selection and might be removed from the model when deemed insignificant. Relationships between scale and other effects in the model can be explored by creating scale-interaction effects prior to variable selection, provided enough degrees of freedom are available to detect them. Admittedly, this requires off-setpoint runs at large scale which are unlikely to be available in a data set.

Random effects in power analysis

At the time the experimental runs for the PCS were planned and the power values in Table 1 were calculated, the importance of random effects was not known to its full extend. Therefore, a priori power analysis that considers the random variance component explicitly was not performed. Simulation-based power calculation methods that incorporate a random variance contribution are available in some software packages. This might be considered in future experimental planning. However, how does this affect our claim that random effects are strong throughout most UOs?. When considering the random effect levels in Table 1, one could argue that the number of levels might not be sufficient in terms of statistical power to detect all active random effects. However, it should be noted that the effect sizes shown in Fig. 3 have been obtained using the variable selection method described in workflow B ("Workflow B: modelling random effects using linear mixed models" section), which controls via a *p*-value threshold for the false-positive rate/type 1 error, even though actual effect sizes might be smaller due to the shrinkage effect described in "Effect sizes and variances" section. The random effect was then checked for significance in the resulting models using variance ratio tests (Nakagawa & Schielzeth, 2013). In only 18 of the 123 models, the random effect was found to be not significant, in which case a value of zero was used for the data points shown in Fig. 3. Moreover, averages of random effect predictor sizes found by LMM over all models are strong throughout all unit operations. This supports our finding that, overall, the random effect is oftentimes larger than the fixed effects. While the lack of statistical power might lead to overly conservative tolerance intervals, the PARs of this study have been found to be practically acceptable for manufacturing.

Implications for the biopharmaceutical industry

Workflow B proposed in the "Workflow B: modelling random effects using linear mixed models" section puts the method for considering random effects in process design (stage 1) proposed by Burdick et al. (Burdick et al., 2017) into the context of a workflow that includes variable selection. This aligns with the ICH8 recommendations for including all potential sources of variation into the computation of the control strategy (ICH, 2017).

Ignoring a random effect or modelling it as a fixed effect can change effect and variance estimates notably. Goos et al. (Goos et al., 2006) demonstrated in a simulation study that this is the case for improperly analyzed split-plot designs, and our results show that it holds true for the analysis of a process characterization data in biopharmaceutical manufacturing. The statistical implications of inappropriately choosing an OLS model over an LMM for the calculation of intervals is shown in "Tolerance intervals" section.

Large tolerance intervals and pronounced random effect sizes indicate that an effect affecting the process is poorly understood, and its true root cause should be investigated. By identifying the source of random variation and controlling it, it can essentially be resolved into a fixed effect.

For example, vendor-to-vendor variability of a raw material might lead to a large random effect, i.e., an unexpected random source of variation. Consequently, a set of experiments can be conducted to identify the true root cause of this variation, e.g., a supplement of the raw material. Provided the manufacturer is able to control this supplement, it can be incorporated into a model as a fixed effect. If this is not feasible or planned for a later point in time, LMM tolerance intervals can be used to estimate the distribution of critical quality attributes more accurately and to find a conservative control strategy for the fixed effects, thus reducing out-of-specification events.

In general, we recommend the following:

- Investigate the practical significance of the random effect (e.g., does its variance take up a large fraction of the CQA acceptance limits/drug substance specification and hence is a risk to the patient?).
- If feasible, conduct experiments to identify causes of random variation and re-evaluate experimental data.
- If it turns out that the effect can be modelled and controlled as a fixed effect, implement changes in the process to control the root cause.
- Uncontrollable effects can still be modelled as random effects in LMMs for conservative tolerance intervals.

Note that process validation is a risk-based approach starting with risk assessment conducted to identify potentially impacting factors (fixed and random effects). Of course, if this initial step overlooks one of the important factors, they will not be assessed experimentally. In case those underrated factors from the risk assessment

are not controlled well in manufacturing, the control strategy established through a PCS might be insufficient. In that case, stage 3 of process validation (continued process verification — CPV) steps in and aims at identification of special cause variation possibly raised from one of the underrated factors. When special cause variation can be detected, it may trigger a new round of risk assessment and experimental planning and analysis, bringing birth to a true life cycle, which FDA proposes in its 2011 PV guideline.

Conclusion

In this article, the role and impact of random effects on setting the control strategy of a biopharmaceutical process were investigated in a real-world case study conducted at Boehringer Ingelheim. Data from a production process comprised of eight up- and downstream unit operations were analyzed in a case study. Although this contribution is based on an extensive process characterization of a monoclonal antibody process and the results are believed to be representative for similar processes, we encourage researchers to conduct similar case studies with other processes and random variables. Here, interweek batch variability was chosen as the random effect. Such an effect, if not ignored entirely, is commonly incorporated in an OLS model as a categorical fixed effect. For the case study, however, the factor was modelled as a random effect using linear mixed models where the segmentation of variance components into random and fixed components enables a more accurate calculation of statistical intervals. The results show that the random effect not only increases the width of the statistical intervals used to compute PARs, but also exceeds even in several unit operations the average size of the fixed effect. Those findings are confirmed by the number of observations contained within the tolerance interval, which agrees with the nominal coverage level for LMMs but not for OLS models. As random effects might have such a strong impact and even stronger impact than fixed effects, they should be incorporated into risk assessments and included into experimental studies. If tolerance intervals derived from LMM models are too large, further investigations should be performed to resolve random effects into fixed effects, e.g., by identifying the underlying root cause of the variation and controlling it. However, until this state is reached, the random variance should at least be accounted for in the model prediction uncertainty as described in this contribution.

Furthermore, we presented a workflow commonly used for creating a control strategy using OLS models. In this workflow, one of the standard implementations of tolerance intervals in a multiple regression setting is utilized, and the intersection points with acceptance criteria are computed to arrive at the acceptable range for each process parameter. This constitutes the control strategy for the process. As an alternative, we proposed an LMM-based workflow that performs similar actions but touches upon certain characteristics of random effects and mixed models. This mainly manifests in the variable selection process and in the computation of statistical intervals where the variance introduced by fixed and random effects is incorporated appropriately. We suggest the use of tolerance intervals based on the sum of expected mean squares proposed by Franzq et al. (Francq et al., 2019). Depending on the group structure and the available degrees of freedom, the interval produced by this method tends to be wider than its OLS counterpart.

Identifying and incorporating random effects are vitally important when defining the control strategy of a process and adjacent tasks like experimental planning and risk assessment. Employing methods described in the proposed workflow, e.g., linear mixed models and corresponding tolerance intervals, leads to a more conservative and appropriate control strategy, which ultimately facilitates more robust processes, patient safety, and fewer out-of-specification events.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s41120-022-00070-5.

Additional file 1. Anonymized PCS data.

Additional file 2. Estimated effect sizes and variance ratios.

Additional file 3. Motivating Example for PAR Calculation. As an introductory example for how different models and interval calculation methods affect PAR calculation, this plot was shown in the article.

Additional file 4. Models and Intervals.

Availability of data and materials
All data generated or analyzed during this study are included in this published article and its supplementary information files.

Declarations

Competing interests
The authors declare that they have no competing interests.

References
Alkharusi H (2012) Categorical variables in regression analysis: a comparison of dummy and effect coding. Int J Educ 4:202–210
Burdick RK, LeBlond DJ, Pfahler LB, Quiroz J, Sidor L, Vukovinsky K, Zhang L (2017) "Process design: stage 1 of the FDA process validation guidance," in Statistical Applications for Chemistry, Manufacturing and Controls (CMC) in the Pharmaceutical Industry, Springer, pp 115–154
FDA, Process validation: general principles and practices, 2011.
Francq BG, Lin D, Hoyer W (2019) Confidence, prediction, and tolerance in linear mixed models. Stat Med 38(30):5603–5622
Goos P, Langhans I, Vandebroek M (2006) Practical inference from industrial split-plot designs. J Qual Technol 38(2):162–179
Govaerts B, Francq B, Marion R, Martin M, Thiel M (2020) The essentials on linear regression, ANOVA, general linear and linear mixed models for the chemist. Ref Module Chem Mol Sci Chem Eng:431–463
ICH, ICH pharmaceutical quality system Q10, 2008.
ICH, ICH guideline Q11 on development and manufacture of drug substances, 2011.
ICH, ICH guideline Q8 (R2) on pharmaceutical development, 2017.
Montgomery DC (2017) Design and analysis of experiments. Wiley
Montgomery DC, Peck EA, Vining GG (2021) Introduction to linear regression analysis. Wiley
Nakagawa S, Schielzeth H (2013) A general and simple method for obtaining R2 from generalized linear mixed-effects models. Methods Ecol Evol 4:133–142
SAS Institute Inc. (2010) SAS/STAT® 9.22 User's Guide. SAS Institute Inc., Cary
Zahel T, Hauer S, Mueller EM, Murphy P, Abad S, Vasilieva E, Maurer D, Brocard C, Reinisch D, Sagmeister P (2017) Integrated process modeling - a process validation life cycle companion. Bioengineering 4(4):86

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.