Journal of Pharmaceutical Innovation (2019) 14:245–258
https://doi.org/10.1007/s12247-018-9350-0

ORIGINAL ARTICLE

Analytical Method Development Using Transmission Raman Spectroscopy for Pharmaceutical Assays and Compliance with Regulatory Guidelines—Part II: Practical Implementation Considerations

Julien Villaumié[1] · Darren Andrews[2] · Kris Geentjens[3] · Benoît Igne[4] · Gary McGeorge[5] · Andrew Owen[2] · Nicholas Pedge[6] · Vicki Woodward[6]

Published online: 3 October 2018

Abstract

Transmission Raman spectroscopy is a relatively new technique for quantitative analysis of pharmaceutical products, either during manufacturing or as a finished product test. As with any new analytical tool, several requirements need to be met for widespread application. These include assessment of technical capability, integration with quality and manufacturing processes and successful deployment in a quality-controlled environment. In the first paper of a two-part series, regulatory guidelines and method development were discussed for the creation of transmission Raman spectroscopic methods for content uniformity (CU), assay and drug product identity (ID) applications. In this part II, the practicalities of method development are addressed, and an example of the development of a quantitative method for the determination of drug content uniformity in individual tablet cores using partial least-squares is presented.

Keywords Transmission Raman · Method development · Method validation · Content uniformity · Assay

Introduction

The analysis of solid oral dosage forms is typically performed by an analytical method based on high-performance liquid chromatography. Transmission near-infrared spectroscopy and, more recently, transmission Raman spectroscopy (TRS) are techniques able to non-destructively analyse dosage forms quantitatively with successful deployments [1–3] and acceptance by global health authorities. In the accompanying paper [4], regulatory guidelines and method development were discussed for the creation of transmission Raman spectroscopic methods for content uniformity, assay and ID applications. In this part II, the practicalities of method development are addressed, and an example of the development of a quantitative method for the determination of drug content in individual tablet cores using partial least-squares (PLS) is presented. These results can be used for bulk batch assay, uniformity of content determination and drug product identification. Although a single example is used to illustrate the principles, other product-specific choices might be made depending on the circumstances, several of which are highlighted throughout the paper.

✉ Benoît Igne
benoit.x.igne@gsk.com

[1] Accord Healthcare, Ltd., Whiddon Valley, Barnstaple, Devon EX32 8NS, UK

[2] Agilent Technologies, Inc., 174 Brook Drive, Milton Park, Abingdon OX14 4SD, UK

[3] Janssen Pharmaceutical, Turnhoutseweg 30, 2340 Beerse, Belgium

[4] GlaxoSmithKline, 709 Swedeland Road, King of Prussia, PA 19406-0390, USA

[5] Bristol-Myers Squibb, 1 Squibb Drive, New Brunswick, NJ 08903, USA

[6] AstraZeneca, Silk Road, Macclesfield, Cheshire SK10 2NA, UK

Method Development Considerations

As with the development of any analytical method that follows the best practices of Quality by Design [5, 6], a structured process should be employed as outlined below:

- Design requirements and clearly defined acceptance criteria (performance requirements) for the method

- Risk assessment
- Method development
- Method validation
- Continuous monitoring and improvement (lifecycle)

A practical embodiment of this workflow is represented in Fig. 1, which may vary with the specifics of a particular product. At the outset of method development, the purpose and outcomes of the study should be determined, along with the risk factors that should be explored in the method development and the criteria required for successful validation. Since Raman spectra possess many features (peaks) and are rich in physical/chemical information that can be used for quantitative analysis, a single model can be developed that may determine several critical quality attributes. A single TRS model can be used to determine (a) drug product identity, (b) batch assay and/or (c) uniformity of content of tablets within a batch. All three of these attributes are required to release a batch and it is possible to employ TRS for some or all of these tests in a regulatory filing. Although the same chemometric model may be used, the sampling and statistical tests differ—assay has a relatively high precision and accuracy requirement (using an average value of multiple samples), CU uses the results from individual samples and ID requires only specificity to be determined. The validation criteria may include limits for accuracy, precision, linearity, etc., [2] and risk factors identified during the method risk assessment will inform the design of experiments for the feasibility and be revisited before method development and validation stages. As part of this process, the error of the reference method and its impact will be assessed and measured, and the criteria of equivalency of the spectroscopic method to the reference method will be defined.

The process outlined in Fig. 1 can broadly be viewed as three successive stages: method feasibility, development and validation.

Method Feasibility

The feasibility stage is a means of assessing the viability of a TRS application without venturing into a complete method development effort and also to investigate measurement parameters. Feasibility may be quickly assessed by overlaying an active pharmaceutical ingredient (API) spectrum with drug product, comparing placebo with drug product, a PCA dose-response analysis or other appropriate assessment of method feasibility. Measurement parameters such as sample presentation, laser power, spectral region of interest and measurement acquisition time can be varied to find suitable conditions. Sample-specific considerations and risk factors would be explored at this time, e.g. adequate specificity, influence of fluorescence contribution, any impact of differently coloured capsule ends, multiple sampling requirements for large samples and influence of layers in bi-layers.

For new drug products, the method feasibility may be carried out on a limited sample set. In this case, care must be taken to draw the right conclusions, especially if products are still in development and may ultimately vary during the product development process. For these reasons, spectroscopic methods are typically employed for assay and CU when the commercial formulation has been finalised. It is important when developing a reference-quality spectroscopic method to present sources of variability that will exist during commercial manufacture and avoid, e.g. large reference method errors or overly large batch to batch variability that may not be relevant to the commercial product.

Method Development

Method development takes the learnings from the feasibility work, along with the risk factor analysis and validation criteria requirements, to inform the design of experiments (DoE) for calibration and validation and sets the protocols for the spectroscopic method. At this stage, the instrument collection parameters are optimised, the sample sets are created and models are optimised (spectral ranges and pre-processing).

It is important to determine the acquisition time that is required to obtain a target precision early in the method development/feasibility in order to have a high confidence that the validation acceptance criteria will be met at the end. A useful approach is to determine the evolution of the signal-to-noise ratio as a function of the collection parameters. For products with a range of tablet strengths, it is likely that the acquisition time required to achieve a target %RSD will depend on the tablet strength.

To ensure robustness, it is advisable to include production samples (if available) if the bulk of the samples were made at laboratory or pilot scale to ensure that the calibration samples are representative of the finished product. Additionally, the physical shape or thickness used may change from development to commercial and such sources of variability would also need to be included in the calibration set (or demonstrated to not pose any issues during validation). More generally, a risk assessment should be conducted and the sources of variability that are expected to affect the model should be included in a design of experiments that varies API concentration(s), raw material variability, tablet mass, thickness and possibly the final film coating. The risk assessment is the mechanism that drives the creation of the design of experiment used for model development. The outcome should be that the method is suitable across all raw material and manufacturing process ranges. In other words, the method should be valid across the product design space.

Whilst HPLC is the most common method used to obtain reference values for the samples generated by the DoE and commercial manufacture, other methods are available,

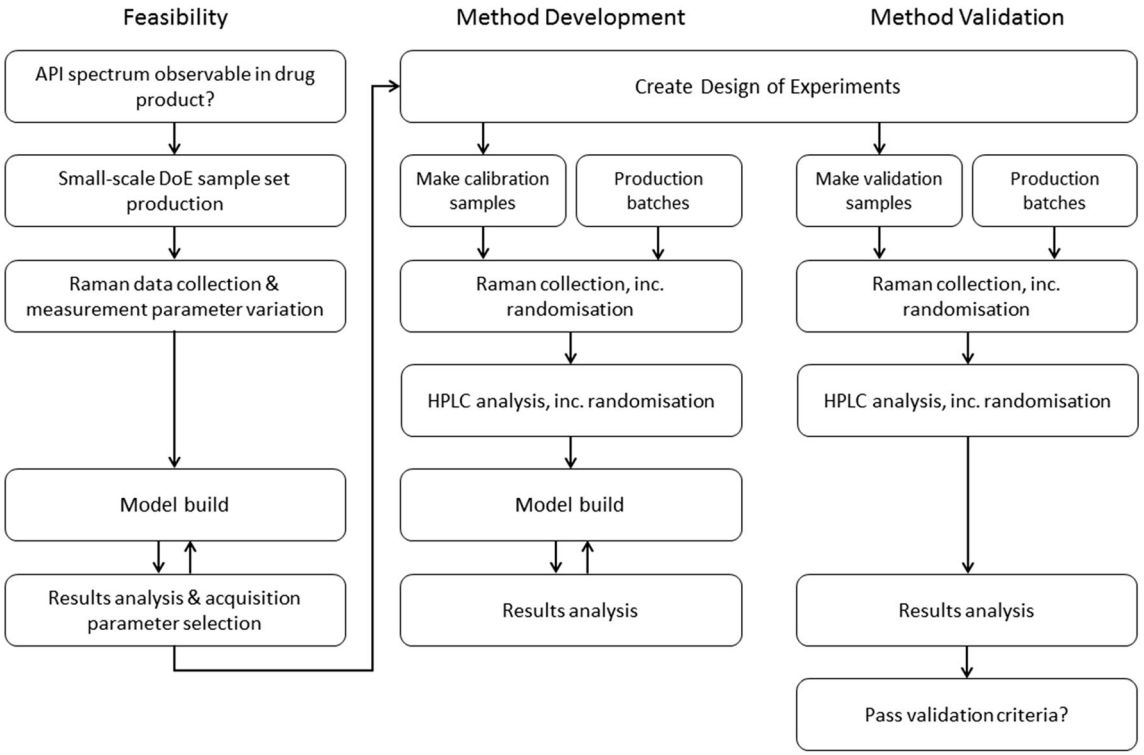

Fig. 1 Outline of a possible method development and validation process

including gravimetric values. Combination of gravimetric (i.e. for DoE samples) and HPLC analysis (for production samples) may be suitable.

Similar to common practices for the reference method, sample randomisation should be considered to account for true lab variances (different days, instruments, etc.) that will be essential in demonstrating successful validation. This can include splitting the samples into several groups so that each group is composed of a balanced selection of the different concentrations; also, the different groups can be analysed on different days and HPLC systems to average out reference method errors. Specifically, samples should not be analysed by order of increasing concentration or other systematic variable, which should result in the most robust model, since the best TRS model comes from having the most representative TRS and HPLC data.

At the time of writing, existing spectroscopic release testing guidance relates directly to near-infrared (NIR) spectroscopy and many of the considerations may be considered similar to NIR by reviewers, i.e. regulators may approach a TRS method review by looking for similar information or justification. TRS applications are obviously different, but until there is more evidence considering impact method factors, developments must leverage NIR-based literature. It is a good practice to discuss submissions with regulators and the intended approach for method validation (refer to the first paper of this series [4] for a discussion of the similarities and differences between the technologies).

Method Validation

Validation tests the spectroscopic model using a predetermined validation protocol. The comparison of the reference method data with the TRS model is made using the predetermined success criteria for method validation, following ICH Q2(R1), such as accuracy, linearity and precision [7]. The output of validation is a report containing the validation plan, data and conclusions based on the comparison of the results against the validation criteria.

Method validation requires the use of samples that are independent of the calibration samples to give results that are representative of the true method performance. The meaning of "independent" varies from application to application, but generally entails the following:

- At a minimum, samples from batches that are different from those used in the calibration, including different lots of active and excipients
- Preferably, samples from batches made in campaigns that are different from the campaigns where the calibration batches were made, so that manufacturing and raw material variability are exhibited
- Ideally, samples from batches that were made with raw material lots different from the lots used in the calibration batches, under a range of manufacturing conditions (e.g. equipment, settings) and with a range of analytical properties (e.g. highest and lowest available assay results)

Note that robustness is as important to the method lifecycle as accuracy is to model validation. If a model is built with minimum variability in raw materials or manufacturing conditions, the method may give an over-optimistic view of future performance. The model may subsequently fail to perform well when the natural variation of the process introduces additional sources of variability that are new to the model. As a consequence, method diagnostics must be in place to ascertain whether a predicted value can be trusted and procedures must be in place to update the model as part of the lifecycle management of the analytical method. A discussion of lifecycle management and model diagnostics was provided in the first paper of the series [4].

Practical Considerations

Method development and validation involve several factors:

(i) Risk analysis

Using a structured risk analysis methodology, the sources of variability impacting the TRS measurement and their potential impact on the model must be identified, ranked and managed. For instance, risks that can directly impact model performance should be mitigated (raw material variability, tablet weight/thickness variability, …); risks that can be avoided may be controlled by the analytical process, e.g. minimising water absorption by the sample before measurement (where water absorption is known to influence the product); risks that are identified should be documented. An example of risk analysis is provided in Fig. 2. The diagram is for illustrative purpose only and does not show all the risk factors that may be associated with different transmission Raman methods (readers should refer to the first article of the series for a discussion of the risk factors [4]).

(ii) Design of experiment

Since production samples are designed to exhibit minimal variation in active content, and spiking is not possible as in traditional wet chemistry methods, the preparation of samples with non-nominal drug concentrations is required. These off-target standards are generally prepared on a laboratory scale using a process that replicates the commercial process. However, if alternate suitable sample preparation techniques can be identified, then they could be used. For multivariate methods, like PLS, the preparation of a single placebo blend and subsequent addition of increasing API amounts is not a recommended approach as it would greatly restrict the spectral variance provided to the model and fail to build a robust model. Such practices create high correlation between the excipients and the active and can make the model specificity and robustness difficult to validate.

Designed experiments, where API concentrations and excipient concentrations are varied, are a superior alternative that simultaneously demonstrates the method's accuracy with regard to the API and the method's robustness to changing excipient concentrations and natural variations in excipient characteristics. It should be noted that whilst a reference method for the active ingredient will be available, it is unlikely that it will be possible to measure the concentration of the excipients. The advantage of inverse calibration methods such as PLS is that the concentration of the excipients is not required to build a model for prediction of the API.

A well-designed DoE can cover the whole calibration space whilst minimising the number of formulations to be made. Response surface designs are usually well suited to reduce the number of samples and provide desirable characteristics such as orthogonality. A widely used design in that category is the central composite design. More efficient designs can also be utilised, such as D-optimal designs. They are computationally derived designs and have demonstrated good performance [8–10].

(iii) Calibration/validation sample manufacture

It is generally desirable to make calibration samples using similar manufacturing processes as to the full-scale production to increase the assurance that the calibration set is valid. However, since small-scale granulation, rotary tablet presses, etc., require specialist equipment and a large amount of powder-blend material, it is often a financial detriment. Surrogate manufacturing processes are therefore often sought out for preparing calibration and some validation samples, which are generally done using laboratory or pilot-scale mixing and small-scale presses. Whilst advantageous, the suitability of the practice must be ensured during method validation and the augmentation of the calibration set with samples made at commercial scale is usually encouraged, e.g. using production samples from multiple batches.

(iv) Validation

Accuracy

Accuracy is determined by comparing the results from the Raman method with the results from the reference method over the method's intended API concentration. Since the reference method generally gives a result in milligram API and Raman methods measure the *%w/w* API, sample weight correction may need to be applied so that results are in the same units and can be compared

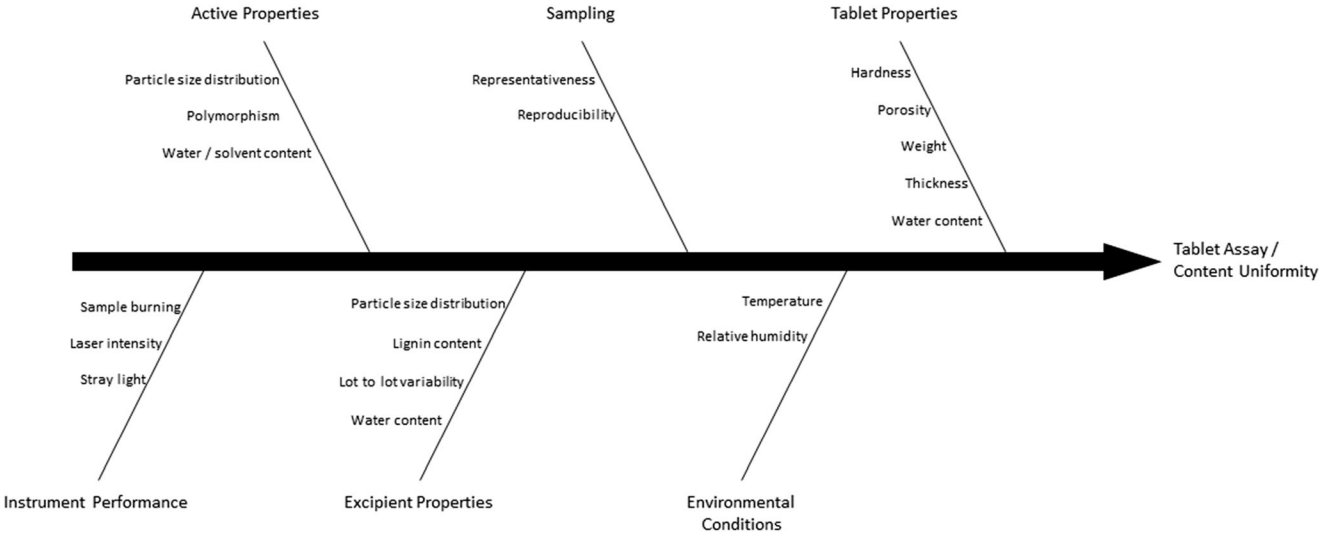

Fig. 2 Ishikawa diagram of potential risk factors associated to a tablet assay/content uniformity method

directly. It is also possible to build models normalised in the same units for comparison, i.e. % label claim.

The key metrics for accuracy are the standard error of prediction (SEP) and the bias, which are defined as

$$\mathrm{SEP} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}}$$

and

$$\mathrm{Bias} = \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)}{n}$$

where n is the number of validation samples, and y and $\hat{y}$ are, respectively, the reference method and Raman method values for validation sample i. These values should be compared to the calibration and cross validation errors to demonstrate the absence of over-fitting of the model.

Precision

Precision is determined by taking multiple Raman measurements and determining the amount of variation (e.g. %RSD) between these measurements. Since Raman spectroscopy is rapid and non-destructive, the same samples can easily be repeatedly analysed. Analysis by the reference method is not required to determine the %RSD of the Raman results, but precision samples can be the same samples as those used for accuracy.

Precision is typically a function of the stability of the instrumentation. For example, variability in the total signal intensity will change the signal to noise of the measurement. The repeatability of sample placement in the auto-sampler holder and the precision positioning performance of the auto-sampler are other examples that could be explored.

The precision should be determined, at a minimum, at the target/nominal concentration. Determining the precision at all the concentration levels tested for the purpose of accuracy/linearity by doing multiple Raman measurements is advisable, as the precision can vary with the concentration of API in the formulation. Repeat Raman measurements can be carried out for the same samples over multiple days by multiple analysts, with removal of the samples from the instrument and repositioning of the samples in the sampling system each time, to demonstrate intermediate precision. In situations where more than one TRS instrument will be used, inter-instrument precision needs to be demonstrated.

Specificity

Raman spectroscopy is characterised by sharp peaks that contain specific information about the chemical composition of the samples. For spectroscopic methods that use chemometrics, specificity relies on the comparison of the spectral characteristics of the analyte of interest and the latent variables/principal components used by the method to quantitate the analyte. This ensures the model is truly predicting active content and not that from any of the excipients or artefacts of the measurement.

Linearity and Range

For quantitative methods like bulk assay or uniformity of content, linearity can be demonstrated by

- Plotting the Raman results against the matching reference method results over the range of active concentrations, and determining the correlation coefficient
- Analysing the residuals (difference between the Raman results and the reference method results) over the range

Table 1 Blend composition for the design of experiment (X, Y, Z and T represent factors used to vary each component)

Condition	API	Excipient 2	Excipient 3	Excipient 4	Excipient 1 (main excipient)
1	Nominal − 2 * X	Nominal − 2 * Y	Nominal − Z	Nominal + T	To 100% *w/w*
2	Nominal − 2 * X	Nominal − Y	Nominal + 2 * Z	Nominal − 2 * T	To 100% *w/w*
3	Nominal − 2 * X	Nominal + Y	Nominal − 2 * Z	Nominal + 2 * T	To 100% *w/w*
4	Nominal − 2 * X	Nominal + 2 * Y	Nominal + Z	Nominal − T	To 100% *w/w*
5	Nominal − X	Nominal − 2 * Y	Nominal − 2 * Z	Nominal − T	To 100% *w/w*
6	Nominal − X	Nominal − Y	Nominal + Z	Nominal + 2 * T	To 100% *w/w*
7	Nominal − X	Nominal + Y	Nominal − Z	Nominal − 2 * T	To 100% *w/w*
8	Nominal − X	Nominal + 2 * Y	Nominal + 2 * Z	Nominal + T	To 100% *w/w*
9	Nominal + X	Nominal − 2 * Y	Nominal + Z	Nominal − 2 * T	To 100% *w/w*
10	Nominal + X	Nominal − Y	Nominal − 2 * Z	Nominal + T	To 100% *w/w*
11	Nominal + X	Nominal + Y	Nominal + 2 * Z	Nominal − T	To 100% *w/w*
12	Nominal + X	Nominal + 2 * Y	Nominal − Z	Nominal + 2 * T	To 100% *w/w*
13	Nominal + 2 * X	Nominal − 2 * Y	Nominal +2 * Z	Nominal + 2 * T	To 100% *w/w*
14	Nominal + 2 * X	Nominal − Y	Nominal − Z	Nominal − T	To 100% *w/w*
15	Nominal + 2 * X	Nominal + Y	Nominal + Z	Nominal + T	To 100% *w/w*
16	Nominal + 2 * X	Nominal + 2 * Y	Nominal − 2 * Z	Nominal − 2 * T	To 100% *w/w*

of API concentrations and ensuring that there is no practical impact on quadratic curvature

- Plotting the norm of the net analyte signal for each sample against its reference value. The net analyte signal is the multivariate contribution to a spectrum that is useful for the prediction of y. This plot demonstrates that the magnitudes of the net analyte signals for a set of samples are correlated to the reference value it is predicting.

One aspect of linearity that is inherently different from HPLC methods is that in an HPLC method, one is simply assessing that the detector response produces a linear response for the reference material at different concentrations. Whilst for a spectroscopic method, it is the comparison between the two methods that are being assessed.

System Suitability

A spectral quality test is necessary to ensure the similarity between the sample spectrum and the calibration spectra and evaluate whether the recorded spectrum is located within or outside the model space. Such spectral check can be performed by using Hotelling's T^2 and Q residual values that can be calculated by the chemometric software used for the quantitation. Hotelling's T^2 expresses how extreme the spectrum is relative to the spectral library (typical spectra will have low T^2 values). The Q residuals express how much of the spectral features are not explained by the chemometric model (unexpected peaks from a contaminant, unexpected peaks due to a source of variation not included in the spectral library, or excessive noise in the data will increase the Q residuals).

The spectral check must be able to accept spectra from correct samples that conform to the method's spectral library and also be able to reject spectra from incorrect samples that do not conform to the method's spectral library, e.g.:

- Samples with an excipient not intended for the product's formulation
- Placebo
- Samples made with an incorrect grade of a material, e.g. lactose anhydrous instead of lactose monohydrate or the wrong API polymorph

Robustness

The extent of robustness testing is application dependent, but should cover the following sources of variation at a minimum:

- Chemical variation (e.g. concentration of API and concentration of excipients, which can be addressed at the same time with designed experiments)
- Raw material variation
- Sampling (e.g. presentation/orientation of the sample in the instrument)
- Sample preparation, if any

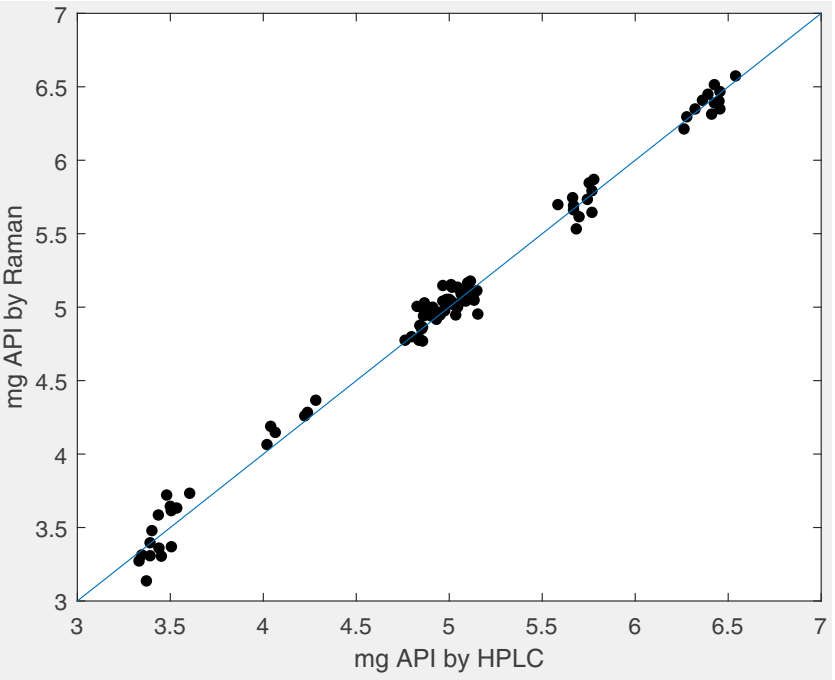

Fig. 3 Predicted vs. HPLC reference plot for the calibration set

In addition, robustness to the spectral acquisition pa rameters may need to be demonstrated by deliberately changing the acquisition time and/or laser power. This is equivalent to testing the robustness of a HPLC method by introducing small changes in flow rate or mobile phase composition. The data normalisation typically performed as part of spectral pre-processing should eliminate absolute intensity variation caused by optical throughput variations of the samples. If absolute intensity variation is of interest for the application, then data normalisation should not be included, and the method will not be robust to changes in data acquisition parameters, nor will it be robust to other variables affecting signal intensity (e.g. tablet thickness).

(v) Submission to a regulator

It is recommended to discuss the potential submission, particularly the first one by a company, with the relevant authorities before commencing work. By agreeing the principles of the method development and validation criteria in advance, the development process can be tailored to address any specific concerns for the filing variation.

Case Study Example

In this case study, a Raman method was developed as a replacement for the reference HPLC method used to quantify the API in a 100-mg tablet manufactured by direct compression. The tablets nominally contained 5 mg of active. The product was frequently manufactured by the site, leading to high volumes of HPLC testing following the manufacturing campaigns. Transmission Raman testing, which avoided sample preparation, would be expected to be a quicker alternative to liquid chromatography resulting in time-saving for quality control personnel.

Table 2 Example of high-level validation criteria

Validation category	Validation criteria
Accuracy	Fit for purpose based on the specifications and requirements of the method
Precision	Repeat measurements error is acceptable with and without repositioning, by operators, at different concentration levels.
Specificity	The model is specific to the parameter of interest.
Linearity	Model predictions are linear over the range of interest.
Range	Covers the range ± 30% of target
Robustness	Robust to batch variability as evaluated by predicting validation samples from pilot and commercial batches The model is able to identify samples different from the calibration set.

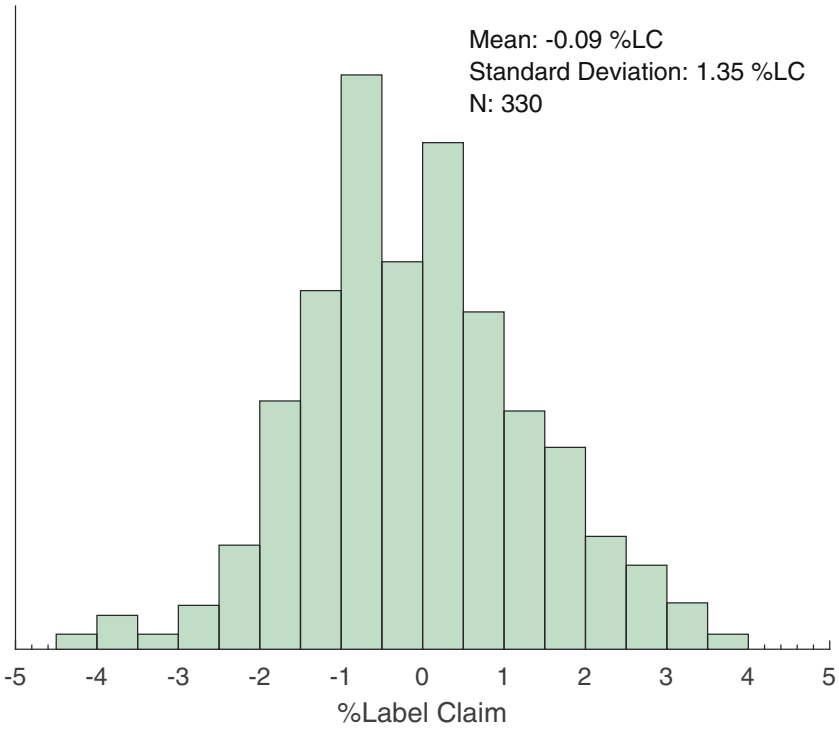

Fig. 4 Distribution of the %LC difference for each sample

Note that in this example, the choices of measurement parameters, tests and analysis that were made are related to this product and the desired outcome—other products and methods are likely to require a modified approach.

Method Development

In order to develop the method, several commercial batches were selected, after review of the warehouse records, such that a wide variation in the raw materials (lots of active and excipients, and, where possible, suppliers) was encompassed in their composition; part of the retain samples placed in storage as is commonly done for commercial product was used to provide samples for the Raman method development. Some of the selected batches were assigned to the calibration data set, and others to the validation data set. Commercial batches, even when selected in that fashion, exhibited very little variation around the nominal active content as expected.

To introduce additional systematic variation into the composition of the calibration sample, a design of experiment was used to make calibration blends on a small scale (75 g each) (Table 1). The DoE was a fractional factorial with 16 different compositions in which the concentrations of the API relative to the excipients and the concentrations of excipients relative to each other were varied. The API concentration was varied to mimic the expectations for a traditional method validation (e.g., HPLC), between 70 and 130% of nominal as per ICH Q2(R1). Excipients' concentrations were varied to make the model more robust and widen its design space, and for logical reasons, an API concentration of, e.g. 70% nominal in a tablet, where the method is intended to work, can

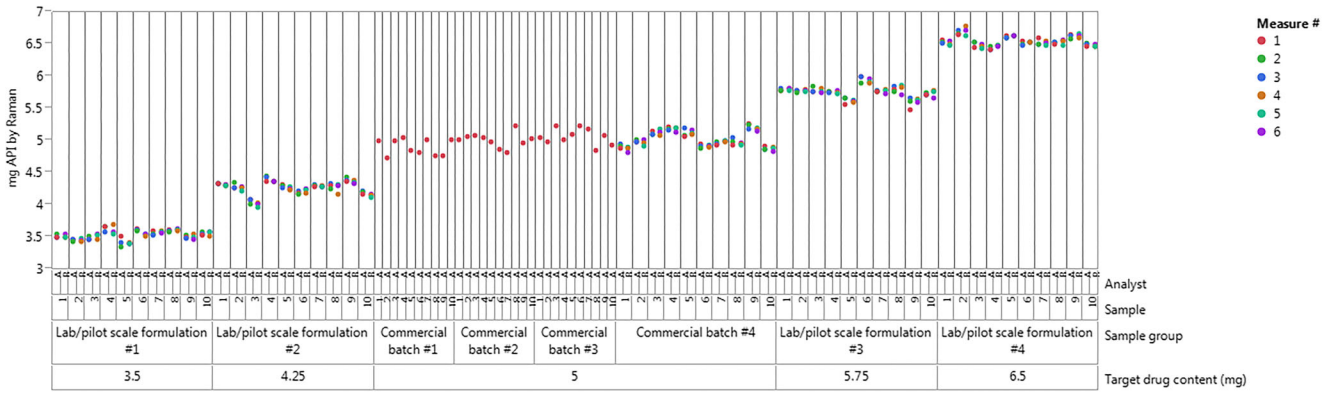

Fig. 5 Example of experimental plan and results obtained to determine method precision and intermediate precision

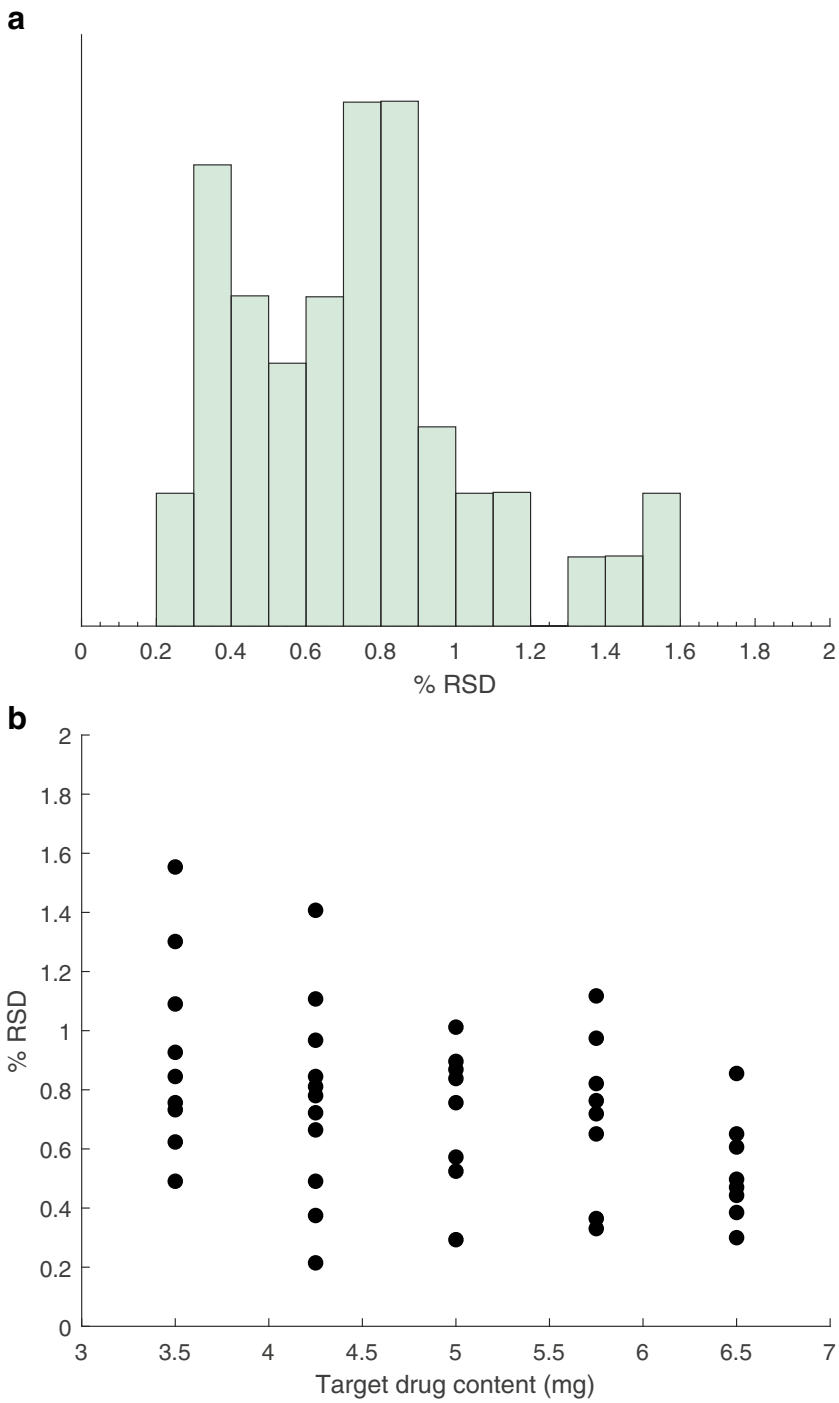

Fig. 6 Summary of a precision study (left), overall distribution of the %RSD values; (right), breakdown of the %RSD values by drug content

only occur if the amount of API relative to excipients is not nominal and, in that situation, one would expect the excipients to also be non-nominal relative to each other.

For the API and all the excipients except the major one, the *%w/w* of each component was varied at four different levels centred around the nominal *%w/w*. For instance, the amount of API could set at one of the following levels; 70%, 85%, 115% or 130% of the nominal *w/w* composition. The nominal *%w/w* of 5% (equivalent to 100%LC) was already present in commercial product, and therefore not repeated in the DoE. Similarly, for the other excipients, four discrete levels centred around nominal were used; nominal concentration was not included in the DoE. For the API and those excipients, five concentrations were therefore possible: four non-nominal from the DoE and one nominal from commercial batches.

The design of the DoE was balanced and orthogonal such that, between the 16 blends, each concentration (other than nominal) for each component represented an equal number of times

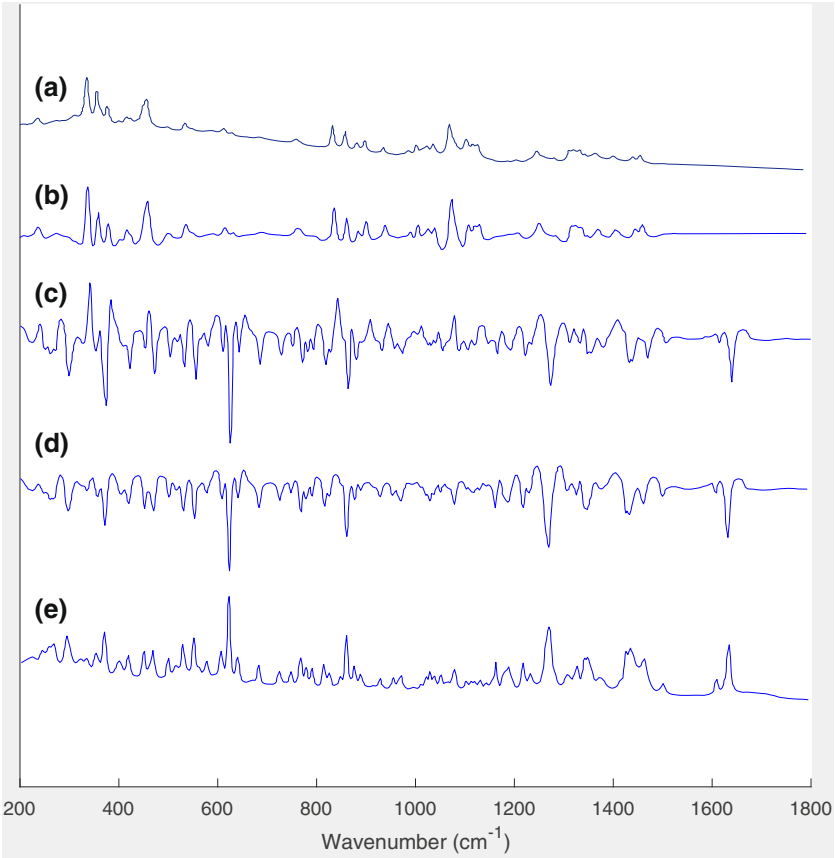

Fig. 7 Spectral comparison of the placebo tablet (raw (a); pre-processed (b)), first latent variable of the model (c) and active ingredient (pre-processed (d); raw (e)). The green shaded areas show feature similarity between the active ingredient that model loading

(4) and that for each concentration of each component, the other components would be an equal number of times at each of their possible concentration levels. The main excipient was used to reach 100% *w/w*. No two blends had the same composition.

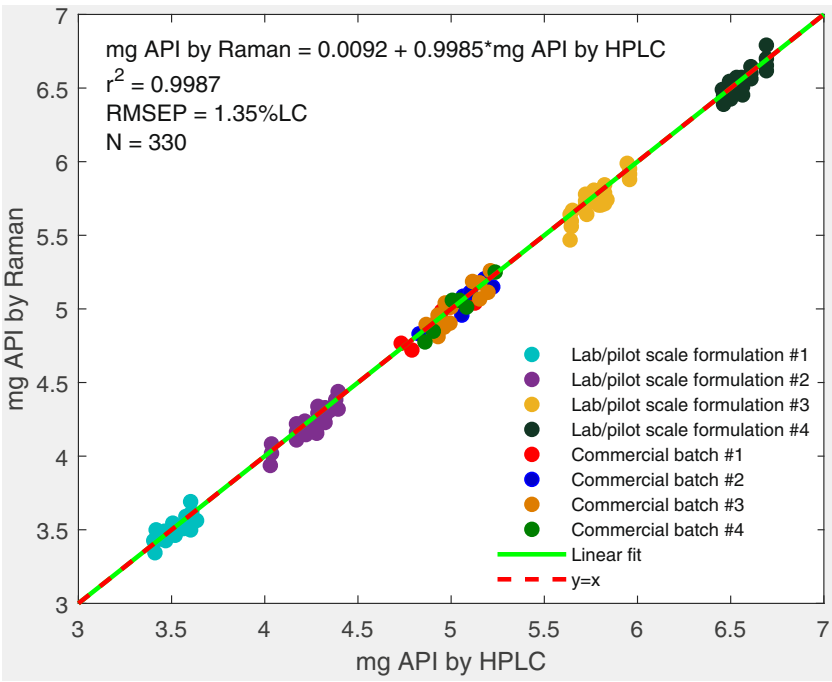

Fig. 8 Example of linearity plot and statistics

Table 3 Example of high-level validation criteria with results from the case study

Validation category	Validation criteria	Results
Accuracy	Fit for purpose based on the specifications and requirements of the method	With an error of 1.35%LC, the model is fit for purpose for the requirements of the method.
Precision	Repeat measurements error is acceptable with and without repositioning, by operators, at different concentration levels.	The %RSD is acceptable over the range of the calibration set.
Specificity	The model is specific to the parameter of interest.	Features of the active in the first latent variables are present, indicating specificity.
Linearity	Model predictions are linear over the range of interest.	The predictions are linear over the range of the calibration.
Range	Covers the range ± 30% of target	The range is covered by the calibration and validation sets.
Robustness	Robust to batch variability as evaluated by predicting validation samples from pilot and commercial batches The model is able to identify samples different from the calibration set.	No statistical difference could be observed between different pilot and commercial batches, indicating robustness.

Each blend was prepared in a small-scale blender and for each DoE blend, a small number of tablets (6) was made using a lab-scale tablet press; the compression force was varied between three levels to have some of the six tablets with a low hardness, some with nominal hardness and some with a high hardness to span the specification range of hardness. This was to ensure that differences in tablet porosity were captured in the calibration set.

Similarly, for the validation, blends at different API concentrations were made (only 1 per API level, e.g. at 70% nominal, 85% nominal, 115% nominal and 130% nominal) and the excipient concentrations were also varied to validate that the Raman method worked within its intended design space. Commercial batches were used for the nominal API concentration. The calibration and validation blends had different formulations, and the batches used for calibration and validation were different to ensure that the samples were completely independent.

The Raman analysis utilised a Cobalt Raman TRS100 with an auto-sampler tray. A laser spot size of 4 mm and a medium collection optic was used. Instrument settings were evaluated to ensure that the precision was sufficiently high. With a measurement time too short, the optimum RMSEP would not be achieved because the precision would be compromised by the limiting noise, in this case photon shot noise. If the optical illumination and sampling areas were too low, the RMSEP would not be optimised because of subsampling. The total accumulation time was 32 s per sample (32 scans of 1 s each). It was determined to provide a

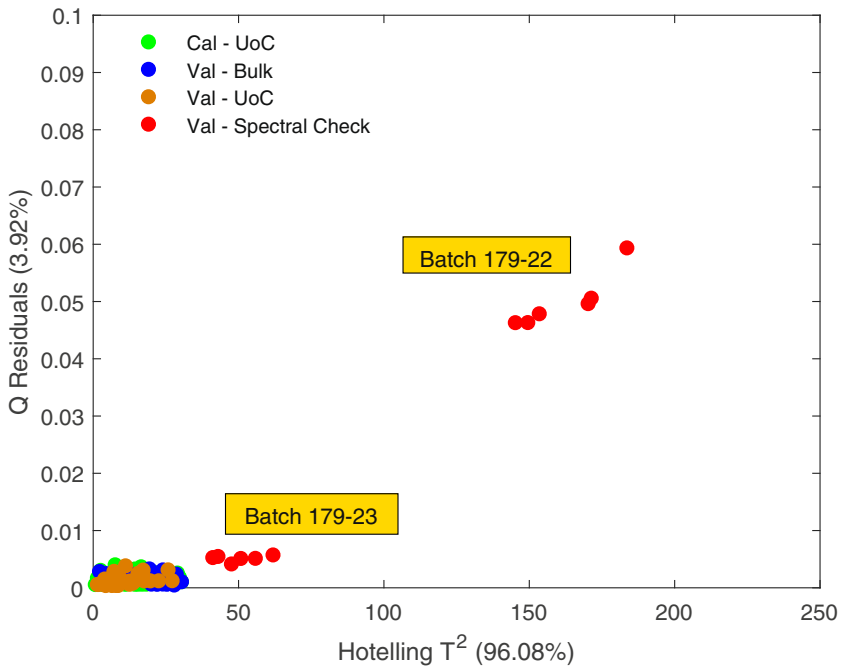

Fig. 9 Hotelling's T^2 and Q residuals are one possible way to detect compositionally incorrect samples, and samples that are not in the scope of the method

relative standard deviation under 2% for precision. The Raman data acquisition for the method (with the appropriate acquisition time and optical configuration) was deliberately randomised with replications on multiple days, using API concentration, batch, type (lab-made/production), day and operator as the randomised variables, to help the model account for instrumental effects such as systematic spectral variation due to instrument temperature changes. Potential instrumental variation may be latent additional variables in the model or will produce latent variables that more accurately describe the sample properties related to API concentration. A well-trained model can account for and compensate instrumental effects, giving a better precision.

Subsequently, analyses by the reference HPLC method were carried out on the same tablets, keeping individual samples separate from each other to allow 1:1 TRS:HPLC comparison. The HPLC method was a validated method used for batch release and was therefore validated as per the applicable guidelines. Nevertheless, to minimise the risk of building in any bias from the reference method, the samples were purposely split into several stratified groups to each be analysed independently by HPLC (i.e. not all on the same day by the same analyst). Each group contained a balanced selection of calibration and validation samples from the different concentrations/blends and commercial batches. Having each HPLC run dedicated to a certain API concentration was specifically avoided. New HPLC standards were prepared each day and tablets analyses were performed by different HPLC runs. This ensured that, even if an HPLC run did have some systematic variation (e.g. a run 1% too high or too low), the error would be spread between all the concentrations and not introduce a bias into the chemometric model.

The PLS model was developed over the frequency range 200 to 1800 cm^{-1}. The spectra were pre-treated with second derivative, normalised to unit area and mean-centred. Five (5) latent variables were selected. Figure 3 shows the predicted vs. HPLC reference plot.

Method Validation

Table 2 presents a high-level example of what a validation protocol could look like for the present case study. There can be many more items to demonstrate that the method is meeting all the required validation criteria. However, it purposefully does not contain statistical criteria because each criterion will be unique to each application and targeted use of the method. (Readers should refer to the first paper of the series [4], ICH [7], EMA [11] and FDA [12, 13] documents for examples of criteria and ASTM standards [14–16] for details about the calculations of some of these statistics.) Concepts of analytical Quality by Design can also be used to set these criteria to develop a fit for purpose method [17, 18].

(i) Accuracy

Figure 4 shows the distribution of the % label claim from the HPLC reference values. This includes the error from the laboratory (operators, instrument and other variances) by appropriately setting the calibration testing protocol so no systematic bias can affect the results. The TRS-predicted values cannot be smaller than the error of the HPLC since the comparison with the reference values will always include this error.

The SEP uses the squared value of each difference to summarise the lack of agreement between the Raman and reference methods; the wider the distribution of differences, the higher the SEP. The model showed a RMSEP of 1.35%LC.

(ii) Precision

The precision was determined by calculating the %RSD between repeat API content measurement by Raman for a number of samples (here 6) shown in Fig. 5. Precision can vary with the drug content; therefore, determining the precision over the range of the method is required by the guidelines [10]. If the repeat measurements were taken by different analysts on different days, then the intermediate precision can be determined. For each sample analysed multiple times by Raman spectroscopy, a %RSD value can be determined, giving as many %RSD values as there are samples. Figure 6 shows an example of output from a precision study. A total of six measurements were performed per sample.

(iii) Specificity

Figure 7 shows the comparison between the spectra of the active ingredient, the placebo tablet and the first latent variable of the model. The specificity of the method can be easily proven by examining the spectral features in the green shaded areas.

(iv) Linearity and range

Figure 8 shows a plot of reference API concentration values vs. TRS prediction values. The linearity was determined by the coefficient of determination (r^2).

(v) Robustness

Figure 8 summarises the fit between the HPLC and TRS data. It included validation data from commercial production samples with variability in the raw material lots. If there were unexpected differences between the validation samples and production values, it would indicate a robustness issue and may require further developing the method at the calibration stage (e.g. adding the required sources of variance). However, the figure shows that within normal operating conditions, the commercial data are well predicted by the model.

(vi) Validation summary

Table 3 presents the criteria outlined in Table 2 outlining if each category was successfully validated.

System Suitability

Once the model is built and validated, a system suitability test is needed to ensure that the instrument is performing as expected and that the model is suitable for predicting the new samples (refer to the first paper of the series for a discussion on system suitability). For the present case study, Fig. 9 shows how Hotelling's T^2 and Q residuals could be used to monitor the suitability of the model. Batches 179-22 and 179-23 were deliberately adulterated at a pilot scale by changing excipient grades to show the ability of the diagnostics to identify known deviations. This was performed with placebo tablets as well (not shown on the plot).

Whilst shown here in a static configuration where the diagnostic statistics are used to differentiate between normal and intentionally different batches, they could be used with statistical limits to demonstrate that as a function of time, the samples are within a relevant (i.e. 95%) confidence interval with respect to the calibration set. In the present case study, the statistical limits of the diagnostics were set after a prolonged production period and as part of the method lifecycle to ensure their robustness and the suitability of the variability included in the calibration data. Cogdill presents a good example of such application of diagnostics [19]. One should also note that these outlier diagnostics can also be used for the basis of an identity method if demonstrated to be suitable.

Conclusions

In this paper, a description of the steps involved in method development and validation was presented for the creation of a transmission Raman model for the prediction of the active ingredient concentration in final oral solid dosage forms. A case study was used to illustrate the concepts and demonstrate the suitability of transmission Raman as a fast and practical technology for replacing wet chemistry methods with robust and accurate measurements.

References

1. Villaumié J, Jeffreys H. Revolutionising Raman with the transmission technique. Eur Pharm Rev. 2015;20(3):41–5.
2. Peeters E, Tavares da Silva AF, Toiviainen M, Van Renterghem J, Vercruysse J, Juuti M, et al. Assessment and prediction of tablet properties using transmission and backscattering Raman spectroscopy and transmission NIR spectroscopy. Asian J Pharm Sci. 2016;11(4):547–58.
3. Casian T, Reznek A, Vonica-Gligor AL, Van Renterghem J, De Beer T, Tomuță I. Development, validation and comparison of near infrared and Raman spectroscopic methods for fast characterization of tablets with amlodipine and valsartan. Talanta. 2017;167:333–43.
4. Andrews, D., Geentjens, K., Igne, B., McGeorge, G., Owen, A., Pedge, N., Villaumié, J., Woodward, V. Analytical method development using transmission Raman spectroscopy for pharmaceutical assays and compliance with regulatory guidelines – Part I: transmission Raman spectroscopy and method development (2018). J Pharm Innov, 13(2) pp 121–132.
5. Borman P, Nethercote P, Chatfield M, Thompson D, Truman K. The application of Quality by Design to analytical methods. Pharm Technol. 2007;31(12):142–52.
6. Reid GL, Morgado J, Barnett K, Harrington B, Wang J, Harwood J, et al. Analytical Quality by Design (AQbD) in pharmaceutical development. Am Pharm Rev. 2013.
7. ICH Q2 (R1) Validation of analytical procedures: text and methodology, International Conference on Harmonization (ICH).
8. Brereton RG. Multilevel multifactor designs for multivariate calibration. Analyst. 1992;122:1521–9.
9. El-Hagrasy, A.S., D'Amico, F., Drennen, III J.K. A process analytical approach to near-infrared process control of pharmaceutical powder blending. Part I: D-optimal design for characterization of powder mixing and preliminary spectral data evaluation. (2006). J Pharm Sci, 95 pp392–406.
10. Bondi Jr., R.W., Igne, B., Drennen III, J.K., Anderson, C.A. Effect of experimental design on the prediction performance of calibration models based on near-infrared spectroscopy for pharmaceutical applications (2012). Appl Spectrosc, 66 (12), pp. 1442–1453.
11. EMEA Guideline: Guideline on the use of near infrared spectroscopy by the pharmaceutical industry and the data requirements for new submissions and variations, January 2014. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2014/06/WC500167967.pdf
12. FDA Guidance for industry on analytical procedures and methods validation for drugs and biologics: July 2015 Pharmaceutical Quality/CMC. http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm386366.pdf
13. FDA Draft guidance for industry on development and submission of near infrared analytical procedures, March 2015. http://www.fda.gov/ucm/groups/fdagov-public/@fdagov-drugs-gen/documents/document/ucm440247.pdf
14. ASTM E2617-17. Standard practice for validation of empirically derived multivariate calibrations.

15. ASTM 2891-13. Standard guide for multivariate data analysis in pharmaceutical development and manufacturing applications.
16. ASTM E1655-17. Standard practices for infrared multivariate quantitative analysis.
17. Borman P, Chatfield M, Nethercote P, Thompson D, Truman K. The application of quality by design to analytical methods. Pharm Technol. 2007;31(12):99 142–52.
18. Schweitzer M, Pohl M, Hanna-Brown M, Nethercote P, Borman P, Hansen G, et al. Implications and opportunities of applying QbD principles to analytical measurements. Pharm Technol. 2010;34(2):52–9.
19. Cogdill, R.P., Anderson, C.A., Drennen, J.K. Process analytical technology case study, Part III: calibration monitoring and transfer. (2005) AAPS PharmSciTech, 6 ppE284-E297.