Journal of Pharmaceutical and Biomedical Analysis 51 (2010) 557–564

Contents lists available at ScienceDirect

Journal of Pharmaceutical and Biomedical Analysis

journal homepage: www.elsevier.com/locate/jpba

Efficient and economic HPLC performance qualification

Lukas Kaminski[a], Matthias Degenhardt[b], Joachim Ermer[c], Claus Feußner[d], Heidemarie Höwer-Fritzen[e], Peter Link[f], Bernd Renger[g], Martin Tegtmeier[h,i], Hermann Wätzig[a,*]

[a] Institute of Pharmaceutical Chemistry, Technical University Braunschweig, Beethovenstrasse 55, D-38106 Braunschweig, Germany
[b] Abbott GmbH & Co. KG, D-67061 Ludwigshafen, Germany
[c] Sanofi-Aventis Deutschland GmbH, D-65926 Frankfurt am Main, Germany
[d] Nycomed GmbH, Byk-Gulden-Strasse 2, D-78467 Konstanz, Germany
[e] W. Spitzner Arzneimittelfabrik GmbH, D-76227 Karlsruhe, Germany
[f] LAZ GmbH & Co. KG, Dischingerweg 15, D-72070 Tübingen, Germany
[g] Vetter Pharma-Fertigung GmbH & Co. KG, D-88212 Ravensburg, Germany
[h] Schaper & Brümmer GmbH & Co. KG, Bahnhofstraße 35, D-38259 Salzgitter, Germany
[i] Institute of Experimental and Clinical Pharmacology and Toxicology, University of Luebeck, Ratzeburger Allee 160, D-23538 Luebeck, Germany

ARTICLE INFO

Article history:
Received 16 June 2009
Received in revised form 28 August 2009
Accepted 8 September 2009
Available online 15 September 2009

Keywords:
HPLC
Performance Qualification
System suitability test
Quality assurance
Fisher test

ABSTRACT

Analytical instrument qualification (AIQ) is a prerequisite for any analytical method validation and thus must be considered as a vital basis of analytical data integrity and quality in pharmaceutical analysis. There is a well-established system of qualification phases—Design Qualification, Installation Qualification (IQ), Operational Qualification (OQ) and Performance Qualification (PQ). As HPLC systems are "off the shelf" equipment, Design Qualification may be disregarded here. IQ establishes that the instrument is received as designed and that it is properly installed. OQ is carried out modularly with the intention to ensure that the specific modules of the system and the whole system are operating according to the defined specifications. PQ as the last step of the initial qualification is supposed to ensure continued satisfactory performance of an instrument under actual running conditions over the anticipated working range during daily use. However, PQ is not a one time exercise, but is currently repeated regularly independently from routine use of the analytical system using standard reference test condition. But this approach, which is time consuming and expensive only provides a snapshot of system performance. As HPLC procedures generally require a system suitability test (SST) prior and/or after test, it might be far more reasonable and robust to use these SST data for a continuous PQ. The work presented here demonstrates that, under certain circumstances, satisfactory instrument performance assessment can be derived from system suitability tests and performance data from daily use as well. A generally accepted qualification list, consisting of only twelve critical parameters, was compiled in a first step. Some parameters such as injector or thermostatting accuracy were considered redundant while others were successfully incorporated in the proposed holistic approach. System suitability test data as well as OQ/PQ data were provided from different sources and evaluated. The promising results confirmed our concept of ongoing/continuous PQ as a major improvement in AIQ. This approach will not only help to reduce time and effort in the daily laboratory routine without losing data quality, but also avoid the critical re-evaluation of numerous analytical tests once a routine PQ fails.

1. Introduction

1.1. The importance of analytical instrument qualification

Analytical procedures in pharmaceutical analysis are subjected to highly formalised validation procedures in order to demonstrate that they are suitable for the intended use. As a consequence, prior to method validation it is necessary to assure that the equipment or analytical test system itself is adequately designed, maintained, calibrated and tested. This process is called analytical instrument qualification (AIQ).

Abbreviations: AIQ, analytical instrument qualification; AUC, area under the curve; DQ, Design Qualification; EQ, Equipment Qualification (equates to AIQ); ICH, International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use; IQ, Installation Qualification; OQ, Operational Qualification; Ph.Eur., European Pharmacopoeia; PQ, Performance Qualification; QC, quality control check; RSD%, relative standard deviation; $\hat{\sigma}$, standard deviation; $\hat{\sigma}^2$ (Var), variation; SST, system suitability test; t_R, retention time; URS, user requirement specifications; USP, United States Pharmacopoeia; U(L)WL, upper (lower) warn limit; U(L)SL, upper (lower) stop limit.

* Corresponding author. Tel.: +49 531 391 2764; fax: +49 531 391 2799.
E-mail address: H.Waetzig@tu-bs.de (H. Wätzig).
URL: http://www.pharmchem.tu-bs.de/waetzig.html (H. Wätzig).

0731-7085/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2009.09.011

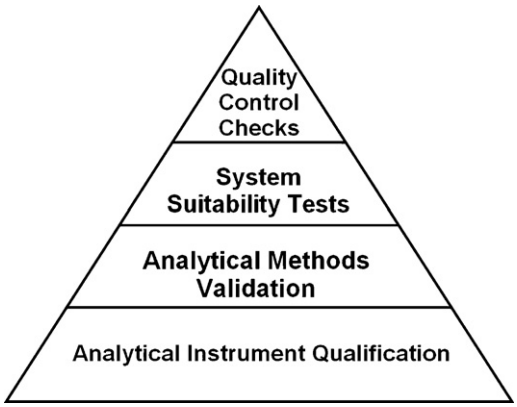

Fig. 1. Data quality triangle according to Bansal et al.

In addition, wherever instrumental analytics is employed the equipment has to be re-qualified regularly. Therefore the process of qualification has also an enormous economical relevance, mainly due to instrument downtime. It is thus surprising, that up to now, only relatively few articles concerning this topic [1–3,5,6,8] have been published.

1.2. Terms and definitions

The process of collecting documented evidence is called either Equipment Qualification (EQ) or analytical instrument qualification (AIQ). AIQ builds the basis of data quality which is completed by analytical method validation, system suitability tests (SSTs) and quality control checks (QCs) (Fig. 1).

After years of inconsistent terminology and differing interpretations of the extend of qualification to be performed deduced and adopted from more general recommendations [1–3], the new General Chapter <1058> of USP 31/NF 28 [4] – effective since August 2008 – now forms a solid basis for analytical instrument qualification and codifies the term "qualification" which erroneously sometimes was called validation. Although, General Chapters with numbers larger than <999> are only recommendatory in nature, this monograph was well received by the pharmaceutical industry as it forms an official regulatory basis for the qualification of equipment in pharmaceutical quality control.

The General Chapter proposes three different categories of instruments with differing qualification effort to be applied:

Group A (simple equipment like stirrers)
Group B (e.g. thermometers, pH meters, refractometers)
Group C (mostly computer-based devices like HPLC, GC, NIR, etc.)

In addition the new chapter recommends to follow the well-established qualification phases also for analytical instrument qualification:

- Design Qualification
- Installation Qualification
- Operational Qualification
- Performance Qualification

Interestingly, Performance Qualification is not limited to a one time exercise, but includes the periodic checks of the instrument like regular calibration activities, preventive maintenance and necessary repairs over the whole life cycle of the individual piece of equipment.

DQ is the part where the design and technical characteristics of an instrument are predefined taking the "user requirement specifications" (URS) into account, unless the design is already in place for a commercial-off-the-shelf-system. In this case the users should ensure that the instrument is suitable for their designated applications.

Before the ordered instrument is delivered the user must verify that the installation site meets all vendor-specified environmental requirements. This is where the IQ part begins. Furthermore it comprises all assembly steps of the equipment at the users' site and is completed by going on line for the first time to run initial diagnostics.

The OQ part is carried out initially and after major modifications or repairs of the instrument. It contains a number of instrument function tests and shall verify that the instrument operates within the manufacturer specified and user approved parameters. Even though it is often performed at modular level, some OQ tests can be carried out holistically as well, making it very difficult to differentiate between OQ and PQ. Actually, AIQ experts [1,3,5,6], the USP [4] and the European Commission [7], as regulatory authorities, state that there is no sharp cut and particular tests of OQ and PQ are interchangeable. Anyhow, both OQ and PQ have to be performed as they serve a different purpose.

PQ is the last of the "four Qs". It shall ensure continued satisfactory performance during routine use. Holistical testing is most suitable here, so interactions between particular modules can be taken into account.

As outlined by the new General Chapter <1058> of the USP, Performance Qualification includes also the regularly activities of preventive maintenance, re-calibration and performance checks. One main challenge when defining acceptable frequencies of these activities was balancing between costs, effort and system availability on one side and the threat of a failing PQ on the other side. Any failing routine PQ would require enormous efforts to reassess and justify all analytical results derived from this piece of equipment starting from the last passing PQ. In many cases passing system suitability tests were used as evidence for compliant system performance.

At this point the question comes up, whether system suitability tests generally can provide supplementary or all needed information about system performance and can be used as an indicator for system failure [1–4]. In this context Bedson and Rudd even introduced an interesting concept of the initial PQ, performed subsequent to OQ, and the ongoing PQ which they equalize with SST [3]. Others do the PQ without additional information from SSTs by providing simple and close to laboratory reality PQ methods which can easily be performed in short time intervals [6,8]. However, all experts agree to one point. SSTs are useful and they often provide additional information about system performance. However, until now it was not believed that SSTs could substitute a PQ to a major extent.

This work will not deal further with Design and Installation Qualification as they both were exhaustively described in the above mentioned works and do not offer possibilities for economization. Instead it will concentrate on the OQ/PQ complex and its economical potential.

1.3. Continuous Performance Qualification: more by less

In this paper we demonstrate that SSTs can do far more than only provide suitability information for particular methods and supplemental PQ information. We present a sound scientific approach to confirm that under certain circumstances a thorough qualification of an HPLC instrument is possible by just using method specific SSTs to obtain a continuous PQ. Using this concept, only slight extensions (effort approximately 1 h) will result in a full PQ with the

Table 1
The revised PQ list consists of only 12 parameters which are necessary to qualify an HPLC instrument thoroughly. It combines PQ parameters (Table 3) and modular parameters of the OQ (Table 2). It can be processed holistically.

Module	Parameter	Procedure	Tolerance	Finding
Injector	Precision of injection volume	Was determined by measuring the RSD% of peak areas	<1.0% RSD	0.99%[a]
	Linearity of injection volume	Was determined by stepping up the injection volume successively (1, 10, 20, 50, 100 μl) and measuring the increase of the peak areas	$R^2 \geq 0.999$	0.9998[b]
	Injection carryover	Was determined by running a blank test directly after an analysis and measuring possible absorption	Method specific	No carryover
Autosampler	Thermostatting precision	Measurement of temperature over a set period of time. Only suitable for autosamplers with temperature control	±2 °C	Not measured
Solvent delivery system	Flow rate accuracy	Was determined by measuring the volumetric flow rate of mobile phase through the column over a set period of time (1.0 ml/min for 10 min, 2.0 ml/min for 5 min and 2.5 ml/min for 10 min)	Expected volume ± 3%	Expected volume ± <0.5%
	Mobile phase proportioning	Can be surveyed continuously with the aid of retention times and their RSD%. If unexpected discrepancies occur a classic gradient test is advisable.		
	Flow rate precision	Was determined by measuring the RSD% of retention times	<1.0% RSD	0.88%[a]
Detector	Wavelength accuracy	Was determined by measuring the spectrum of one substance of the test sample	Specific maxima ±2 nm	Specific maxima ± 0.4 nm
	Noise	Was determined by carrying out a dynamic measurement with mobile phase for 15 min	$<1 \times 10^{-3}$ AU (for dynamic noise)	5×10^{-4} AU
	Drift	Was determined by carrying out a dynamic measurement with mobile phase for 1 h	$<5 \times 10^{-3}$ AU/h	3.5×10^{-3} AU/h
	Linearity of detector response	Was determined in the same manner as linearity of injection volume	$R^2 \geq 0.999$	0.9998[b]
Column oven	Thermostatting precision of column oven	Was determined by measuring the RSD% of retention times	<1.0% RSD	0.88%[a]

[a] Value obtained from the worst peak in the routine method.
[b] In addition to R^2 a residue plot was performed to exclude any trend.

additional benefit of continuous performance surveillance. Representative data sets from various pharmaceutical companies will be presented and statistically evaluated. A revised PQ list will then be presented in Table 1.

2. Experimental

2.1. HPLC equipment used in the pilot scheme

The HPLC instrument used for the qualification tests in the pilot scheme was a Hitachi® MERCK® system consisting of a solvent pump (model L 6200 A), an autosampler (AS 2000A), a DAD (diode array detector L-7450), and an interface (D-6000). The data were collected and analyzed using the D7000 HSM software (Merck). The column oven (T1) was provided by TechLab® (Erkerode, Germany).

2.2. Chromatographic conditions for PQ and SST during the pilot scheme

All separations were performed on a Chromolith® Performance RP-18e column (100 mm × 4.6 mm, Merck®). For system suitability testing one of our in-house methods (further on referred to as routine method), a generic assay of glibenclamide, glimepiride and two degradation products (routine method mixture), dissolved in a mixture of phosphate buffer pH 7.4 and acetonitrile 20%:80% (v/v), was used. For closer information about this method see Ref. [9].

The injection volume of the routine method mixture was 10 μl and the monitoring wavelength was 228 nm. The chromatographic runtime was 1.5 min.

For PQ we used a method suggested by Merck® (PQ—method), the supplier of the monolithic column. It was a quite simple separation of progesterone and anthracene with thiourea as t_0 marker (PQ—mixture) dissolved in the mobile phase which consisted of acetonitrile and water 60%:40% (v/v). The injection volume was 10 μl and the monitoring wavelength was 254 nm. PQ—runtime was 4.5 min. All separation runs were performed at ambient temperature.

2.3. Statistical methods and equations

All datasets were evaluated with a two-sided Fisher test (*F*-test) with an error probability α of 5% according to the basic equations (1) and (2).

$$T_F = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}; \quad \hat{\sigma}_1^2 \geq \hat{\sigma}_2^2 \tag{1}$$

$$T_{F(\text{scaled})} = \frac{\text{RSD\%}_1^2}{\text{RSD\%}_2^2} \tag{2}$$

The critical *F*-values $F_{\text{crit}(\alpha;\, n-1;\, m-1)}$ were calculated using the Microsoft® Excel® command FINV (α; $n-1$; $m-1$). Small datasets were pooled to obtain significant statistical information (3).

$$\hat{\sigma}_{\text{pooled}}^2 = \frac{\sum_{i=1}^{n} df_i \cdot \hat{\sigma}_i^2}{\sum_{i=1}^{n} df_i} \tag{3}$$

3. 3 Results and discussion

3.1. Development of the revised OQ/PQ parameters list

Currently there are mandatory parameter listings neither for OQ nor for PQ. The USP General Chapter <1058> [4] agrees with most experts that users of analytical equipment are ultimately responsible for their instruments' operations and data quality.

Therefore, they shall develop instrument tests and acceptance criteria which are necessary for AIQ using their expertise. However, this was and still is a very unfavorable situation for users of analytical equipment. In 1996 the UK Instrumentation Workgroup chaired by Dr. Mike Sargent published a general approach to AIQ, which is applicable to most instruments used in analytical laboratories, as a supporting guideline for analytical chemists [10]. This approach was supplemented in 1999 by guiding documents specific to HPLC [3]. Although in this work Bedson et al. offered clearly arranged and already shortened parameter lists as well for DQ and IQ as for OQ and PQ, they still pointed out that their work only provided recommendations and not a mandatory proceeding in AIQ.

Nevertheless these works were the initial point for the development of our compact OQ/PQ parameter compilation which is supposed to be used as a straightforward guideline for instrument qualification.

In a first step Bedson's recommendations were compared to common parameter lists from our cooperation partners and HPLC equipment manufacturers. Thus two lists containing as few as possible but as many as necessary parameters became apparent (Tables 2 and 3). In order to verify these compilations they were used for detailed qualification procedures of our own equipment. Furthermore, they had been presented to expert groups from the pharmaceutical industry during the meeting of Qualified Persons (QP) in Northern Germany (HK-Nord) May 30, 2008, to the QC departments of Sanofi Aventis, Germany, June 26, 2008 and to the working group QA of the A.P.V., June 27, 2008. Their recommendations and our requalification experiments led to further refinement of the lists.

3.1.1. Improvement of the parameter lists

The following parameters turned out to be expendable and were abandoned from the lists. *Accuracy of injection volume* is difficult to determine and in practice standards are commonly used so slight inaccuracy is not carrying much weight. *Autosampler temperature accuracy* must only be measured if required. *Linearity of injection volume* and *linearity of detector response* were combined as they can be determined in the same manner. By checking one of them the other is automatically checked, too. The coefficient of determination (R^2) may be used to verify an appropriate linear range e.g. $R^2 \geq 0.999$. Of course, a relatively high R^2 value may not always guarantee linearity. However, here the aim is not to demonstrate a linear response function, but to verify that the linear range of the detector is not exceeded. For this purpose, the coefficient of determination can be considered as a suitable numerical parameter as it will decrease in case of a systematic deviation from linearity. In order to be sure that the deviation is really due to a deviation from linearity, a residue plot can be performed in addition to confirm the assumed linearity and to exclude any trend of the calibration curve [23]. The *signal to noise ratio* measurement is rather important during method validation where it is often used to determine the limit of detection (LOD: $S/N \geq 3$) and the limit of quantification (LOQ: $S/N \geq 10$) [11]. In OQ/PQ determination of LOD and LOQ is not necessary. However, S/N affects precision and therefore it should be determined when precision gets worse. In practice related testings of our cooperation partners and manufacturers S/N was not measured.

3.1.2. Transfer of modular parameters into the holistic approach

In the next step of the development both parameter lists were merged into one holistic approach which is shown in Table 1. A holistic approach was favorable for our project as it combined easy handling, time saving and applicability of SSTs.

The following paragraph gives a clear overview of transferred parameters sorted according to the associated instrument parts.

3.1.2.1. Autosampler. Thermostatting precision—can be measured as hitherto with a temperature probe as this is not technically demanding and it can be done during routine analysis, whenever needed.

3.1.2.2. Solvent delivery system. Flow rate accuracy—in the classic, discontinuous OQ/PQ the volumetric flow rate is measured with HPLC grade water. A restrictive capillary or a dummy column is used to generate the required backpressure. However, changing columns, mobile phases and time consuming equilibrations make this proceeding inconvenient. For the holistic approach the analytical column remained connected to the system and the flow rate was measured with the mobile phase of the routine method. Acetonitrile/water (aqueous phosphate buffer) mixtures in the concentration range of 100%: 0% up to 20%:80% and methanol/water mixtures within the same concentration range were tested during these experiments as they are commonly used mobile phases in RP-HPLC. Different flow rates ranging from 1.0 up to 2.5 ml/min were applied as well to receive valid results. As expected from a previous work [9] in all cases the pump performed accurately within the limits of measuring uncertainty.

Mobile phase proportioning—in modular testing two different solvent channels connected to two different solvent reservoirs are used. One contains pure water while the other contains an acetone solution (0.1% m/m). Increasing the portion from the acetone channel (either linearly or incrementally) leads to an increase of UV absorption and hence the proportioning accuracy of the gradient mixer can be calculated.

It is not possible to transfer this test to the holistic approach as it is dependent on special solutions and test conditions. However, it is possible to conclude proper functioning of the gradient mixer from stable retention times in gradient mode analysis. If well known retention times of standard substances begin to fluctuate and no longer comply with their acceptance criteria, this could be among other things an indicator for a malfunction of the gradient mixer. A closer inspection in the form of the before mentioned test is necessary only in this case.

3.1.2.3. Detector. Wavelength accuracy—typically the spectrum of a built in holmium perchlorate or holmium oxide filter is measured. The detected wavelengths must not exceed the limits of ±2 nm [3]. The use of certain standard solutions such as caffeine or anthracene solution is also applicable [5,8]. In this regard it should be noted that absorption at particular wavelengths is an explicit physical characteristic of UV active chemical compounds. Hence in principle, in the holistic approach the UV spectrum of any UV active substance should be adequate to ensure *wavelength accuracy*. However, this is only true for compounds with defined absorption peaks. Substances with multiple absorption peaks covering the whole UV range are most suitable here. In the holistic approach we used the spectrum of glibenclamide from our routine method (Fig. 2). It can be seen that the measured absorption peaks match perfectly the literature values *A* and *B* [12,13]. Apart from that, these values were equal to archived data measured with the same instrument half a year, 1 year and even 6 years earlier. For these reasons one can absolutely conclude detector suitability. In contrast UV spectra with limited significance (e.g. most benzodiazepines) [12] can hardly be used for detector qualification as they have poorly defined absorption maxima. Instead in this case one should revert to the approved standards like caffeine.

Noise and drift—are typically measured statically (with the pump switched off) in classic OQ. The cleaned flow cell is filled with purified HPLC grade water and the UV detector signal is recorded for a

Table 2
This was a preselection from our point of view necessary modular OQ parameters. This list was altered while the project has been developed and some parameters were transferred to the holistic approach of the revised PQ list (Table 1).

Module	Parameter	Procedure	Tolerance	Finding
Injector	Accuracy of injection volume	Was determined by comparing peak areas received with autosampler injection and a calibrated dosage loop	Not yet defined	±1.3%
Autosampler	Thermostatting accuracy	Measurement with a proper temperature probe. Only suitable for autosamplers with temperature control	±2 °C	Not measured
	Thermostatting precision	Measurement of temperature over a set period of time. Only suitable for autosamplers with temperature control	±2 °C	Not measured
Solvent delivery system	Flow rate accuracy	Was determined by measuring the volumetric flow rate of water through a restrictive capillary over a set period of time	1.0 ml/min for 10.0 min ±3%	10.0 ml
	Mobile phase proportioning	Was determined by running acetone as additive through one solvent channel and measuring its concentration following a stepwise increase	±1% for low pressure gradient	<1% for each step
Detector	Wavelength accuracy	Was determined by measuring the spectrum of a holmium perchlorate filter	361 ± 2 nm	361.1 nm
	Noise (peak to peak)	Was determined by carrying out a static measurement for 15 min with HPLC grade water	$<5 \times 10^{-5}$ AU	3.4×10^{-5} AU
	Drift	Was determined by carrying out a static measurement for 1 h with HPLC grade water	$<5 \times 10^{-3}$ AU/h	2.4×10^{-3} AU/h

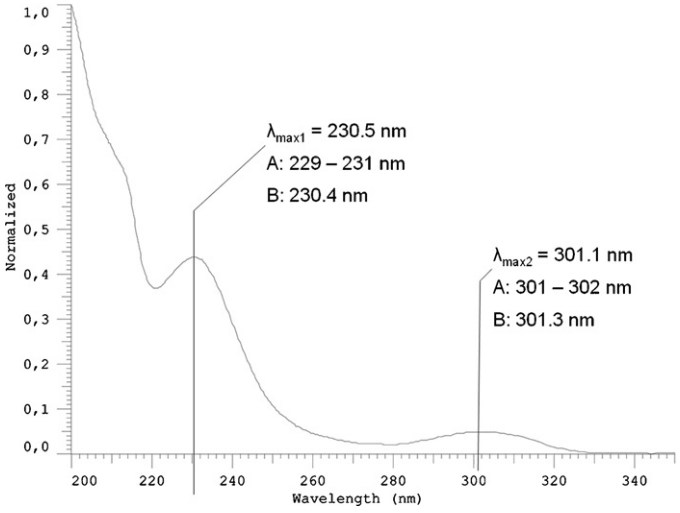

Fig. 2. UV spectrum of glibenclamide showing the accord of measured values with literature values *A* [12] and *B* [13] of λ_{max}.

defined time period (usually 15 min). The values must not exceed acceptance limits. These limits depend on the used detector. Such limits as 5×10^{-5} AU for detector *noise* (peak to peak noise) and 5×10^{-3} AU/h for detector *drift* have been suggested as default values by the manufacturer of our equipment [14]. However, note that manufacturer specified limits are often based on a best case scenario and can hardly be met under realistic working conditions [5]. At this point it is reasonable to question if one really needs to assure that the detector performs well under best conditions available. Possibly it is sufficient to assure that performance remains good under realistic conditions for the analyses to be performed.

Huber and Welebob [5] demonstrated in their work that testing a detector for acceptance criteria set by the manufacturer can have a huge impact on the instruments' maintenance costs. This can only be justified if a particular application requires such stringent limits. Otherwise, according to Huber et al. users should define less strict acceptance criteria with e.g. 2×10^{-4} AU as a general limit for static *detector noise*. In our holistic approach *noise* and *drift* were tested with the standard flow cell and the same mobile phase compositions already mentioned in the flow rate accuracy part of this work. As expected the increase of the organic modifier portion led to a higher baseline noise. However, in a static measurement series all received results stayed within the manufacturers' spec-

Table 3
A preselection of necessary holistic PQ parameters according to our point of view. This list was refined while the project has been developed. Most parameters were incorporated in the revised PQ list (Table 1).

Parameter	Procedure	Tolerance	Finding
Precision of injection volume	Was determined by measuring the RSD% of peak areas. Could also be performed at OQ level	<1.0% RSD	0.61%[a], 0.96%[b]
Linearity of injection volume	Was determined by stepping up the injection volume successively and measuring the increase of the peak areas	$R^2 \geq 0.999$	0.9998[c]
Injection carryover	Can be determined by running a blank test directly after an analysis and measuring possible absorption	Method specific	Not measured
Flow rate precision	Was determined by measuring the RSD% of retention times	<1.0% RSD	0.40%[a], 0.41%[b]
Thermostatting precision of column oven	Was determined by measuring the RSD% of retention times	<1.0% RSD	0.40%[a], 0.41%[b]
Linearity of detector response	Was determined in the same manner as linearity of injection volume	$R^2 \geq 0.999$	0.9998[c]
Signal to noise ratio	Can be determined by measuring a highly diluted test sample	Method specific	Not measured

[a] Worst value obtained with the PQ method.
[b] Worst value obtained with the routine method.
[c] In addition to R^2 a residue plot was be performed to exclude any trend.

ifications. In a dynamic measurement series (pump switched on) the turbulences in the flow cell increased the baseline *noise* [15] approximately by a factor of ten. Long term data evaluation of our own equipment [9,16,17] confirms the work of Cabooter et al. [15] and allows for the conclusion that even adverse conditions, such as different mobile phases and dynamic measurement increase the baseline *noise* not more than one order of magnitude, compared to a static measurement. From these works one can derive a new general limit for *detector noise*. A value of 1×10^{-3} AU for a dynamic noise level which corresponds to approximately 1×10^{-4} AU for static noise could be recommended as sufficient, even if better values can often be obtained. However, noise levels worse than these values would impair the performance of the used methods.

The *detector drift* was neither impaired significantly by different mobile phase compositions nor by dynamic measurement. Hence both parameters (*noise* and *drift*) can be determined holistically during routine analysis, although it is reasonable to check them while running a blank sample between two analyses.

3.2. OQ/PQ data in comparison with SST data

The other concern of this work was to demonstrate that SSTs can do more than only provide information about particular methods or complement OQ/PQ. Under certain circumstances they can not only replace PQ but even the whole OQ/PQ complex. Therefore it was necessary to demonstrate that the informative value of SSTs is at least as good as the one of OQ/PQ. In a first step this was accomplished with data sets from our own equipment in a pilot scheme.

3.2.1. Pilot scheme

An SST was derived from our routine method. Then the SST data was compared to OQ/PQ data which was acquired with the same instrument using the PQ method. Both, SST and OQ/PQ testings were carried out four times, each with five injections to obtain data sets suitable for statistical evaluation ($n = 20$).

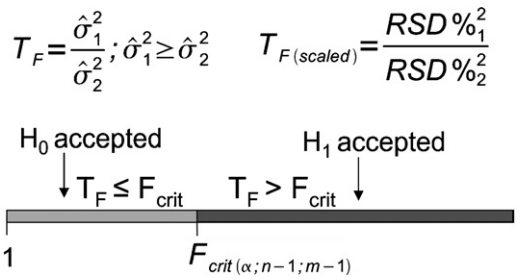

Fig. 3. The ratio of the squared standard deviations T_F is the test statistic. If $T_F = 1$ the variances are all the same. For $T_F \leq F_{crit}$ the variances have no significant difference. For $T_F > F_{crit}$ the variances differ significantly with the probability of $1 - \alpha$. In our calculations $\alpha = 5\%$ and $n - 1$; $m - 1$ were degrees of freedom.

If the standard deviations of retention times and peak areas, related to the routine and the PQ method, do not differ significantly, the informative value of both methods should not differ significantly either. In this case it does not matter which method is used to determine the performance parameters and hence to assess the system qualification. For this reason a two-sided Fisher test was used. This is a statistical method which can be applied to decide if the variances ($\hat{\sigma}^2$) of two samples out of two populations differ significantly with a certain error probability α. Therefore a null hypothesis ($H_0 : \hat{\sigma}_{SST} = \hat{\sigma}_{PQ}$) is set up which can either be accepted or nullified in favor of an alternative hypothesis ($H_1 : \hat{\sigma}_{SST} \neq \hat{\sigma}_{PQ}$) (Fig. 3). Note that an acceptance probability ($1 - \alpha$) is only given for nullifying H_0. There is no level of significance for accepting this hypothesis. This can only be achieved with equivalence tests such as these used in bioequivalence studies or analytical method transfers [18–20]. However, these tests were not required in this work as here it was not necessary to prove that the variances were equal. It was sufficient to know that they were not significantly different.

In the pilot scheme relative standard deviations of t_R values as well as of AUC values from each peak in the SST data were compared to each peak in the PQ data (Table 4). In none of these cases

Table 4
An overview of the substances used in the pilot scheme. Each row represents a series of five consecutive injections.

Series no: PQ data	Substance	Mean area	σ area	RSD% area	Scaled variance	
26	Progesterone	40.8154	0.2507	0.6141	0.3772	
27	Progesterone	40.5820	0.0421	0.1038	0.0108	
28	Progesterone	40.5870	0.1733	0.4269	0.1823	$VAR_{pool} = 0.1630$ (AUC)[a]
29	Progesterone	40.5964	0.1162	0.2863	0.0820	$RSD\%_{pool} = 0.4038$
26	Anthracene	51.2164	0.2362	0.4611	0.2126	
27	Anthracene	51.5506	0.1880	0.3648	0.1331	
28	Anthracene	51.4448	0.3447	0.6701	0.4490	$VAR_{pool} = 0.2189$
29	Anthracene	51.4554	0.1464	0.2844	0.0809	$RSD\%_{pool} = 0.4679$

Series no: SST data	Substance	Mean area	σ area	RSD% area	Scaled variance	
48	Related substance (a)	23.3150	0.1044	0.4476	0.2004	
49		23.1490	0.0646	0.2789	0.0778	
50	Related substance (a)	23.1724	0.0760	0.3279	0.1075	$VAR_{pool} = 0.3281$ (AUC)[a]
51		23.1362	0.2227	0.9627	0.9267	$RSD\%_{pool} = 0.5728$
48	Related substance (b)	21.4570	0.0915	0.4263	0.1818	
49		21.6444	0.0841	0.3886	0.1510	
50	Related substance (b)	21.5222	0.1460	0.6785	0.4604	$VAR_{pool} = 0.2712$
51		21.5898	0.1166	0.5400	0.2916	$RSD\%_{pool} = 0.5207$
48	Glibenclamide	29.6072	0.1142	0.3856	0.1487	
49	Glibenclamide	29.5890	0.1289	0.4356	0.1898	
50	Glibenclamide	29.5002	0.1623	0.5501	0.3026	$VAR_{pool} = 0.2704$ (AUC)
51	Glibenclamide	29.4504	0.1955	0.6638	0.4406	$RSD\%_{pool} = 0.5200$
48	Glimepiride	25.7408	0.2362	0.9177	0.8422	
49	Glimepiride	25.6398	0.1047	0.4083	0.1667	
50	Glimepiride	25.8250	0.0741	0.2870	0.0824	$VAR_{pool} = 0.3012$
51	Glimepiride	25.8436	0.0871	0.3369	0.1135	$RSD\%_{pool} = 0.5488$

[a] The variances were used for the calculated example (Fig. 4).

$$T_{F(scaled)} = \frac{RSD\%_1^2}{RSD\%_2^2} = \frac{0.3281}{0.1630} = 2.0123$$

$$F_{crit(0.05;16;16)} = 2.3335$$

$$T_{F(scaled)} < F_{crit} \rightarrow H_0 \text{ accepted}$$

Fig. 4. In this calculation example 0.3281 is the pooled and squared RSD% of the AUC of related substance a. 0.1630 is the pooled and squared RSD% of the AUC of progesterone (Table 4).

H_1 was accepted so one can assume that the variances were not significantly different (Table 5). A calculation example is given in (Fig. 4).

3.2.2. Data sets of cooperation partners

The pilot project provided satisfying results hence it was expanded. SST and OQ/PQ data sets from cooperating pharmaceutical companies and analytical laboratories were collected. Pretreatment of these data was necessary as in practice SST data sets are larger than PQ data sets simply because SSTs are carried out more often. Therefore OQ/PQ data were pooled and compared to SST data from a certain time period to obtain $n_{SST} \approx n_{PQ} \approx 10–20$ and hence better comparability.

After this pretreatment, the data sets were evaluated in the same manner as the pilot scheme. Most data sets provided the results, that H_0 was accepted. In some cases H_1 was accepted which signalized different variances. However, a closer look at the data showed that in most of these cases SST data were significantly better and hence provided even better performance information than OQ/PQ did. This is only surprising at first glance but it just reflects the users experience with routine methods and the high quality of today's method development.

Table 5
The pooled AUC variance of each peak in the SST chromatogram was compared to the pooled AUC variance of each peak in the PQ chromatogram. No variance was accepted as significantly different. The variances of t_R were tested in an analogous manner. Since these results were conclusive as well, they are not presented here.

Related substance (a): progesterone		
$T_F = 2.0123$	$F_{crit} = 2.3335$	H_0 = accepted
Related substance (a): anthracene		
$T_F = 1.4989$	$F_{crit} = 2.3335$	H_0 = accepted
Related substance (b): progesterone		
$T_F = 1.6631$	$F_{crit} = 2.3335$	H_0 = accepted
Related substance (b): anthracene		
$T_F = 1.2388$	$F_{crit} = 2.3335$	H_0 = accepted
Glibenclamide: progesterone		
$T_F = 1.6586$	$F_{crit} = 2.3335$	H_0 = accepted
Glibenclamide: anthracene		
$T_F = 1.2354$	$F_{crit} = 2.3335$	H_0 = accepted
Glimepiride: progesterone		
$T_F = 1.8474$	$F_{crit} = 2.3335$	H_0 = accepted
Glimepiride: anthracene		
$T_F = 1.3760$	$F_{crit} = 2.3335$	H_0 = accepted

SST data worse than OQ/PQ data also occurred sometimes but could be traced by our cooperation partners to instrument malfunctions at that time.

3.3. Control charts

Control charts are widely used in the field of quality control to determine whether a manufacturing process is in a state of statistical control or not. It seems very appropriate to use them also in the concept of continuous PQ to survey performance stability over time. Much data is accumulated during scheduled SSTs which can most appropriately be used to generate a long term

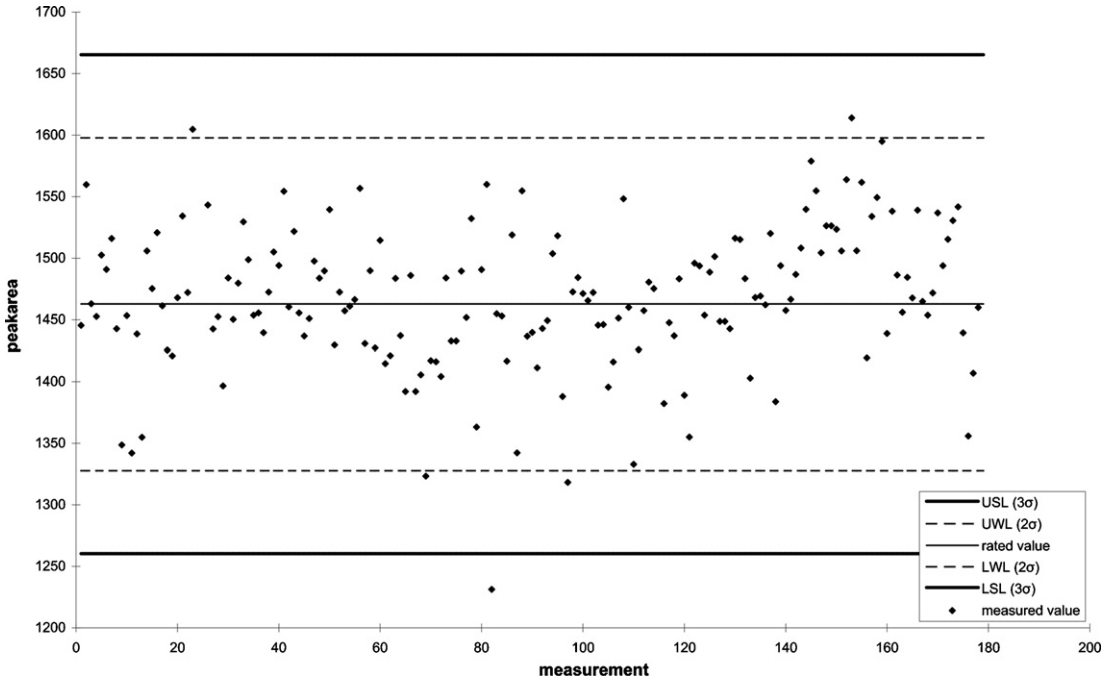

Fig. 5. This is an example of a classic control chart. It helps to recognize a possible performance drift in time. However, users must customize such control charts for their own needs.

instrument performance history. It is thinkable to customize the control charts to special needs of continuous PQ but this concept is not fully developed yet (Fig. 5).

4. Concluding remarks

Ongoing or continuous PQ is superior to "classic" OQ/PQ in many points but it is not trivial to establish. Regulatory bodies must be convinced with a sound scientific approach. Therefore in this work a reliable set of LC parameters was presented which can easily be determined holistically. It consists of only 12 parameters which are sufficient for a thorough qualification of an HPLC instrument. This setting comprises successfully transferred OQ and already established PQ parameters in one holistic approach. The advantage for the user is a straightforward checklist and the possibility to qualify an instrument without changing any parts such as flow cell, column and mobile phase reservoir. It has also been shown that this revised PQ list can at least be applied for any RP-LC method using a UV/DA detector. Thus time consuming sample and mobile phase preparations in addition to long equilibration procedures for extra OQ/PQ methods can be avoided.

Classic SSTs as they are prescribed by Ph.Eur. [11], USP [21] and by the ICH [22] are an integral part of many analytical methods. In general SSTs are performed directly before and between routine analytical series. They are method specific and based on the concept that the equipment (including software and analytical procedures) constitutes an integral system that can be evaluated as such. Thus they are very similar to the holistic approach although they do not provide all relevant information for qualification. However, these SSTs can easily be extended by applying the parameters list presented here. It was demonstrated that SSTs possess a comparable informative value to OQ/PQ so they can definitely be used in the concept of ongoing or continuous PQ with all its benefits. Note that some routine methods might be inappropriate and thus associated SSTs are incapable of providing all important information even if the introduced parameters list is applied. However, for economical reasons in most cases more than one method is run on an instrument. Then it is absolutely sufficient that only one SST per instrument is capable of using the concept of continuous PQ. Imagine the following scenario: methods A, B and C are run alternately on one instrument. The SSTs of A and C do not provide suitable data for PQ but the SST of B does. Hence this instrument will be continuously qualified whenever method B is used.

Although this concept is referred to as continuous, we recommend that only parameters from the list which are related to retention times or peak areas are to be determined each time a method is run since it can be done automatically e.g. by using validated Microsoft® Excel® spreadsheets. The determination of the remaining parameters takes a little bit of extra time (approximately 1 h). However, instrumental downtimes and extra manpower are not required any longer. These extra parameters may be determined at the currently defined PQ frequency which would combine low effort with a continuous evaluation.

Considering all aspects continuous PQ is a major improvement. It avoids extra working time. The continuous survey of critical instrument parameters enhances analytical certainty and hence the overall data quality as it provides not only a snapshot of system performance but also an ongoing performance history. We suggest the use of control charts for documenting and archiving purposes. This is also the best way to detect a performance drift and to take appropriate counter measures in time.

Acknowledgements

We would like to thank Ms. S. Redweik, Ms. F. Würch, Ms. S. Grotefend and Mr. T. Hahne for their commitment in the qualification experiments. Furthermore we are grateful to Ms. S. Bollhorst (Schaper & Brümmer GmbH), Mr. A. Groh (Nycomed GmbH) and Dr. W. Barillas Porras (Vetter Pharma-Fertigung GmbH & Co. KG) for their enduring and helpful discussions considering their SST/PQ data.

References

[1] S.K. Bansal, T. Layloff, E.D. Bush, M. Hamilton, E.A. Hankinson, J.S. Landy, S. Lowes, M.M. Nasr, P.A. St Jean, V.P. Shah, Qualification of analytical instruments for use in the pharmaceutical industry: a scientific approach, AAPS PharmSciTech 5 (2004) 1–8.
[2] V. Grisanti, E.J. Zachowski, Operational and performance qualification, LC–GC N. Am. 20 (2002) 356–362.
[3] P. Bedson, D. Rudd, The development and application of guidance on EQ of analytical instruments: high performance liquid chromatography, Accred. Qual. Assur. 4 (1999) 50–62.
[4] USP, United States Pharmacopoeia 31 NF-26 S1 General Chapter <1058>, 2008, Analytical Instrument Qualification.
[5] L. Huber, L. Welebob, Selecting parameters and limits for equipment operational qualification, Accred. Qual. Assur. 2 (1997) 316–322.
[6] J. Crowther, J. Dowling, R. Hartwick, B. Ciccone, Performance qualification of HPLC instrumentation in regulated laboratories, LC–GC N. Am. 26 (2008).
[7] European Commission, EU Guide to Good Manufacturing Practice, Annex 15, European Commission, 2001.
[8] G. Hall, J.W. Dolan, Performance qualification of LC systems, LC–GC N. Am. 20 (2002) 842–848.
[9] L. Kaminski, S. El Deeb, H. Watzig, Repeatability of monolithic HPLC columns while using a flow program, J. Sep. Sci. 31 (2008) 1745–1749.
[10] P. Bedson, M. Sargent, The development and application of guidance on equipment qualification of analytical instruments, Accred. Qual. Assur. 1 (1996) 265–275.
[11] Ph.Eur., European Pharmacopoeia 6.4. Chapter 2.2. 46, 2009, Chromatographic separation techniques.
[12] H. Auterhoff, K.A. Kovar, Identifizierung von Arzneistoffen, WVG Stuttgart, 1998.
[13] K. Kumasaka, T. Kojima, H. Honda, K. Doi, Screening and quantitative analysis for sulfonylurea-type antidiabetic agents in adulterated health food using TLC and HPLC, J. Health Science 51 (2005) 453–460.
[14] Merck SOP: Operational Qualification of the DAD L-7450, 1997.
[15] D. Cabooter, F. Lynen, P. Sandra, G. Desmet, Turbulence as a source of excessive baseline noise during high-speed isocratic and gradient separations using absorption detection, Anal. Chem. 80 (2008) 1679–1688.
[16] S. El Deeb, L. Preu, H. Watzig, Evaluation of monolithic HPLC columns for various pharmaceutical separations: method transfer from conventional phases and batch to batch repeatability, J. Pharm. Biomed. Anal. 44 (2007) 85–95.
[17] S. El Deeb, U. Schepers, H. Watzig, A strategy to develop fast RP-HPLC methods using monolithic silica columns, J. Sep. Sci. 29 (2006) 1571–1577.
[18] W.J. Westlake, Use of confidence intervals in analysis of comparative bioavailability trials, J. Pharm. Sci. 61 (1972) 1340–1341.
[19] U. Schepers, H. Watzig, Application of the equivalence test for analytical method transfers: testing precision using the United States Pharmacopoeia concept <1010>, J. Pharm. Biomed. Anal. 41 (2006) 290–292.
[20] U. Schepers, H. Watzig, Application of the equivalence test according to a concept for analytical method transfers from the International Society for Pharmaceutical Engineering (ISPE), J. Pharm. Biomed. Anal. 39 (2005) 310–314.
[21] USP, United States Pharmacopoeia 31 NF-26 S1 General Chapter <621>, 2008, System Suitability.
[22] ICH, Guideline Q2 (R1): Validation of Analytical Procedures: Text and Methodology, 2005.
[23] K. Baumann, Regression and calibration for analytical separation techniques. Part II. Validation, weighted and robust regression, Process Contr. Qual. 10 (1997) 75–112.