Received: 13 March 2019 | Revised: 13 May 2019 | Accepted: 7 June 2019

DOI: 10.1002/bmc.4623

REVIEW

WILEY Biomedical Chromatography

Bioanalytical LC–MS/MS validation of therapeutic drug monitoring assays in oncology

Merel van Nuland[1,2] | Hilde Rosing[1] | Jan H.M. Schellens[3] | Jos H. Beijnen[1,2,3]

[1] Department of Pharmacy & Pharmacology, The Netherlands Cancer Institute, Amsterdam, The Netherlands

[2] Division of Pharmacology, The Netherlands Cancer Institute, Amsterdam, The Netherlands

[3] Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands

Correspondence
Merel van Nuland, MSc, Department of Pharmacy & Pharmacology, Antoni van Leeuwenhoek – Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX. Amsterdam, The Netherlands.
Email: m.v.nuland@nki.nl

Abstract

Therapeutic drug monitoring (TDM) has shown to benefit patients treated with drugs of many drug classes, among which is oncology. With an increasing demand for drug monitoring, new assays have to be developed and validated. Guidelines for bioanalytical validation issued by the European Medicines Agency and US Food and Drug Administration are applicable for clinical trials and toxicokinetic studies and demand fully validated bioanalytical methods to yield reliable results. However, for TDM assays a limited validation approach is suggested based on the intended use of these methods. This review presents an overview of publications that describe method validation of assays specifically designed for TDM. In addition to evaluating current practice, we provide recommendations that could serve as a guide for future validations of TDM assays.

KEYWORDS

bioanalysis, guideline, LC–MS/MS, oncology, therapeutic drug monitoring, validation

1 | INTRODUCTION

Therapeutic drug monitoring (TDM) is the clinical practice of measuring drug concentrations in biological fluids to individualize drug dosing. The goal of this intervention is to prevent drug failure by achieving adequate drug levels while also reducing toxicity by preventing overexposure. Some important requirements for drugs to be considered for TDM are long-term therapy, availability of a sensitive bioanalytical method, high inter-patient variability and low intra-patient variability, a narrow therapeutic window, an established dose–response and/or dose–toxicity relationship and a feasible strategy for individualized dosing (de Jonge, Huitema, Schellens, Rodenhuis, & Beijnen, 2005). Many anticancer drugs fit the above described prerequisites, and therefore, TDM of anticancer drugs is becoming an important tool in treatment of patients with cancer, especially with increased use of oral anticancer drugs with highly variable bioavailability (B. Gao et al., 2012; Herbrink et al., 2018; Lankheet et al., 2014; Paci et al., 2014; Widmer et al., 2014; Yu et al., 2014). Consequently, TDM has been shown to be a valuable intervention to optimize dosing of anticancer drugs, resulting in effective treatment (Groenland et al., 2019; Paci et al., 2014; Verheijen et al., 2017; Widmer et al., 2014; Yu et al., 2014).

A fundamental requirement for the implementation of TDM is the availability of bioanalytical assays to reliably measure drug concentrations, and concentrations of relevant metabolites. Different analytical techniques can be used, such as immunoassays and liquid-chromatography methods with UV (LC–UV), fluorescence or mass detection (LC–MS/MS) (Adaway & Keevil, 2012; Schellens, McLeod, & Newell, 2005). Although all four are used in clinical practice, implementation of LC-coupled techniques gained popularity for routine measurements as immunoassays show lack of specificity and precision and show high variability between manufacturers (Dasgupta, 2016). Furthermore, immunoassays could be plagued by cross-reactivity of structural analogous and generally have a shorter linear calibration range (Zhang & Zhang, 2018). LC–MS/MS methods, on the other hand, can be applied for simultaneous quantification of drugs and their metabolites with high sensitivity and selectivity and are therefore superior to LC–UV (Adaway & Keevil, 2012; Miura & Takahashi, 2016). Bioanalytical assays for TDM are used for routine clinical care and should therefore be fast and easy to implement, with high accuracy, precision and selectivity (Adaway & Keevil, 2012). LC–MS/MS assays can offer this by short run times and fast pretreatment procedures.

The focus of TDM assays should be on developing and validating a robust and high-throughput method for routine measurements, while the focus of assays for pharmacokinetic and toxicokinetic studies (PK–TK studies) should be on generating quantitative concentration data in a wider concentration range. Guidelines for bioanalytical method validation, issued by the European Medicines Agency (EMA) and US Food and Drug Administration (FDA), provide valuable assistance for the purpose of assay validation in clinical PK–TK studies (European Medicines Agency, 2011; US Food and Drug Administration (FDA), 2018). These guidelines are, however, comprehensive for TDM assays because drug concentrations determined for TDM purpose are generally reported as being below or above a target concentration and, therefore, not the exact concentration but target attainment is of interest. Together with the increasing demand for TDM in oncology, owing to the use of oral anticancer drugs with highly variable bioavailability, assay validation ought to be simple and straightforward, while still offering confidence in the data quality obtained with the validated method. There is a need for more concise guidelines specifically designed for the validation of TDM assays. In addition to more consise validation procedures, the analysis of study samples should have a rapid turnaround by implementation of a short analytical run. This review aims to present an overview of publications that describe LC–MS/MS assays which have been validated specifically for application in TDM. In addition to evaluating current practices, we provide recommendations that could serve as a guide for future validations and analysis of study samples for TDM purposes.

2 | LITERATURE SEARCH

PubMed was searched on 12 February 2019 using the following terms: "Therapeutic drug monitoring AND validation". We chose not to specify oncology in the search, to evaluate the number of bioanalytical validation papers for therapeutic drug monitoring in other fields. In addition, citation snow-balling was used. Inclusion was limited to bioanalysis in humans and full-text articles available in the English language.

The search identified 941 papers, of which 36 were found to be eligible for inclusion. Figure 1 shows a flow-chart of the inclusion process. Validation papers for therapeutic drug monitoring were identified by studying full-length articles for the aim of the study: assays that were developed specifically for implementation in therapeutic drug monitoring were included, while assays developed for the bioanalytical support of clinical studies and a potential application to TDM were excluded. Furthermore, only full validation articles were included in this review. This review focuses on the 36 published validation papers in the field of oncology; however, recommendations may be applicable to other fields. Results of the literature survey are summarized in Table 1.

3 | BIOANALYTICAL METHOD VALIDATION GUIDELINES

Guidelines on bioanalytical method validation are provided by the FDA and EMA (European Medicines Agency, 2011; US Food and Drug Administration, 2018). Although there is an overlap in experiments and acceptance criteria for all validation parameters, some differences are apparent when these guidelines are compared. Table 2 gives a brief overview of the validation experiments and acceptance criteria as described by the FDA and EMA guidelines. Of the 36 included papers, 27 articles refer to EMA and FDA guidelines for validation procedures. Other guidelines are occasionally used for recommendations on specific validation parameters, such as the Clinical & Laboratory Standards Institute (CLSI) guidelines (Clinical and Laboratory Standards Institute, 2003, 2006). Furthermore, matrix effect and recovery were investigated by a variety of methods described in the literature (Bonfiglio et al., 1999; King et al., 2000; Marchi et al., 2010; Matuszewski et al., 2003; Taylor, 2005).

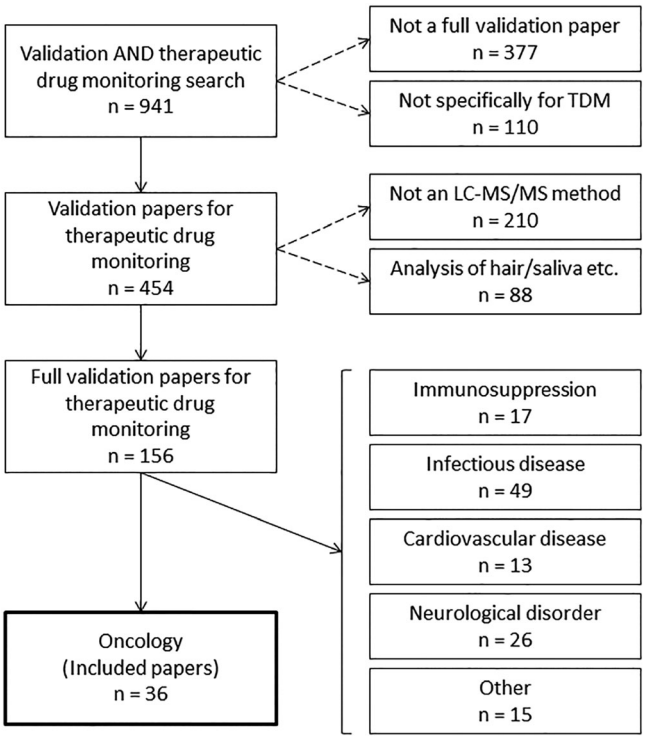

FIGURE 1 Flowchart that shows the step-by-step process of inclusion to generate a final number of studies for analysis in the review

In this review we will discuss the following aspects of method validation: calibration model, accuracy and precision, lower limit of quantification (LLOQ), carryover, selectivity (endogenous and exogenous), dilution integrity, matrix effect, recovery and stability. Furthermore, aspects on the analysis of study samples will be evaluated. For each parameter, recommendations from relevant guidelines for assays supporting PK–TK studies will be summarized, followed by results from the literature search and recommendations specifically for TDM assays. We aim to provide guidance and criteria for TDM method validation and on the application of these validated methods in the routine analysis of study samples.

4 | CALIBRATION MODEL

The calibration model shows the relationship between instrument response and nominal analyte concentrations. For assays in clinical

TABLE 1 Overview of bioanalytical LC–MS/MS assays for therapeutic drug monitoring in oncology

	Calibration		Accuracy and precision		LLOQ	Selectivity		Carryover	Dilution integrity
Analyte(s)	Levels	Range (-fold)	Levels	Replic.	S/N	Endogenous	Exogenous	*n* samples	Levels replic.
Osimertinib	7	1000	4	Intra: 6 Inter: 18		6 blanks 6 LLOQ 6 zeros		3 blanks	
Afatinib Inter:15 Axitinib Dabrafenib Dasatinib Erlotinib Gefitinib Ibrutinib Imatinib Lapatinib Nilotinib Pazopanib Regorafenib Ruxolitinib Sorafenib Sunitinib Trametinib Vandetanib Vemurafenib	6	100	3	Inter: 18 Intra: 8	5	6 blanks			
Busulfan	5	300	3	Intra: 20 Inter: 20			OTC, DO A		2 levels
Vincristine	8	1000	3	Intra: 5 Inter: 4		1 blank (pooled)	OTC, AC D		
Pemetrexed	7	64	3	Intra: 10 Inter: 30					
Sorafenib	6	1000	4	Intra: 4 Inter: 6		6 blanks	ACD		2 levels 5 replic.
Everolimus	7	80	5	Intra: 10 Inter: 10			IMS		1 level 10 replic.
5-Fluorouracil	8	1000	4	Intra: 6 Inter: 18	5	6 blanks	ACD		1 level 3 replic.
Methotrexate	7	500	3	Intra: 5 Inter: 15	5	6 blanks 6 LLOQ			1 level 5 replic.
Busulfan	5	333	3	Intra: 10 Inter: 28			Hemo, Lipi. Icte	3 Low	3 levels 1 replic.

TABLE 1 Overview of bioanalytical LC–MS/MS assays for therapeutic drug monitoring in oncology

Analyte(s)	Matrix effect Levels blanks replic.	Recovery Levels blanks replic.	Internal standards	Short-term stability Levels	 Conditions	Reference
Osimertinib	2	3 levels 4 replic.	Pazopanib	3	RT 4 h F/T 3 FE 24 h	(Rood, van Bussel, Schellens, Beijnen, & Sparidans, 2016)
Afatinib Inter:15 Axitinib Dabrafenib Dasatinib Erlotinib Gefitinib Ibrutinib Imatinib Lapatinib Nilotinib Pazopanib Regorafenib Ruxolitinib Sorafenib Sunitinib Trametinib Vandetanib Vemurafenib	FE 96°C 6 blanks 1 replic.		[$^{2}H_3$]-Erlotinib [2H3]-Gefitinib [$^{2}H_7$]-Lapatinib [$^{2}H_3$]-Sorafenib			(van Dyk, Miners, Kichenadasse, McKinnon, & Rowland, 2016) 6 replic.
Busulfan	3 levels 3 replic. (Annesley, 2003)		5 blanks		RT 28 days F/T 6 cycles 4°C 28 days −70°C 28 days	(Danso, Jannetto, Enger, & Langman, 2015)
Vincristine		3 levels 3 replic.	Vinblastine	3	RT 15 h F/T 3 cycles FE 10 h	(Dennison, Renbarger, Walterhouse, Jones, & Hall, 2008)
Pemetrexed			Methotrexate	3	F/T 3 cycles 4°C 24 h −20°C 20 days	(Meesters et al., 2010)
Sorafenib	5 levels 9 blanks 1 replic.	5 levels 3 replic.	[^{13}C,$^{2}H_4$]-Sorafenib	4	RT 5h FE 24 h—30°C 1 week	(Bobin-Dubigeon et al., 2011)
Everolimus	2 levels 10 blanks 1 replic. (King, Bonfiglio, Fernandez-Metzler, Miller-Stein, & Olah, 2000)	5 levels 3 replic.	40-*O*-(3-hydroxy)propyl-rapamycin	3	FE 24 h	(Taylor, Franklin, Graham, & Pillans, 2007)

TABLE 1 (Continued)

	Calibration		Accuracy and precision		LLOQ	Selectivity		Carryover	Dilution integrity
Analyte(s)	Levels	Range (-fold)	Levels	Replic.	S/N	Endogenous	Exogenous	*n* samples	Levels replic.
Docetaxel Paclitaxel	7	100	4	Intra: 5 Inter: 5	10	6 blanks 6 LLOQ			
Vinblastine Vinorelbine Everolimus Sirolimus	6	19	8	Intra: 15 Inter: 15			IMS		Patient samples 1 replic.
Octreotide	9	50	4	Intra: 6 Inter: 15					
Imatinib	8	20	3		10	1 blank 1 LLOQ	Cross-IS	1 solvent	
Lapatinib	8	50	4	Intra: 5 Inter: 5	10	6 blanks	Cross- IS		
Letrozole	60	60	4	Intra: 6 Inter: 18	10	6 blanks		>1 blank	
Methotrexate	6	5000	3	Inta: 10 Inter: 20		6 spiked samples		1 blank	3 levels 3 replic.
6-Methylmer-captopurine 6-Thioguanine	6	100	3	Imprecision: 3 Inaccuracy: 12		10 blanks			
Everolimus	5	54	3	Intra: 3 Inter: 30				1 Low	
Cobimetinib Dabrafenib Pazopanib Regorafenib Trametinib Vemurafenib	8	500	3	Intra: 6 Inter: 6		8 blanks	Metabolites and other	>1 blank	
Dasatinib Erlotinib Gefitinib Imatinib Lapatinib Nilotinib Sorafenib SunitinibC	8	500	4	Intra: 15 Inter: 15		6 blanks 6 LLOQ	Cross-analyte/IS	2 blanks	1 level 5 replic.
Olaparib Pazopanib Ruxolitinib Vismodegib		250–1000		Intra: 25 Inter: 24		6 blanks	TKIs, ART, AFT, other		

TABLE 1 (Continued)

Analyte(s)	Matrix effect Levels blanks replic.	Recovery Levels blanks replic.	Internal standards	Short-term stability Levels	 Conditions	Reference
5-Fluorouracil		3 levels 3 replic.	[$^{15}N_2$]-5-Fluorouracil	3	RT 4 h F/T 3 cycles FE 96 h	(Kosovec, Egorin, Gjurich, & Beumer, 2008)
Methotrexate	3 levels 1 blank 5 replic.	3 levels 5 replic.	*p*-Aminoacetophenone	2	RT 7 h F/T 3 cycles FE 48 h	(Wu, Wang, Sun, Ouyang, & Qian, 2015)
Busulfan	1 level 3 blanks 3 replic. (Matuszewski, Constanzer, & Chavez-Eng, 2003)	3 levels. 3 replic.	[$^{2}H_3$]-Busulfan		4°C 1 week F/T 8 cycles	(French, Sujishi, Long-Boyle, & Ritchie, 2014)
Docetaxel Paclitaxel Vinblastine Vinorelbine Everolimus Sirolimus	3 levels 1 blank 5 replic. (Matuszewski et al., 2003) Post-column infusion (Bonfiglio, King, Olah, & Merkl, 1999)	3 levels, 5 replic.	Vindoline [^{13}C,$^{2}H_4$]- Everolimus	 3	RT 12 h F/T 3 cycles −20°C 1 month RT 12 h	(S. Gao et al., 2014) (Morgan, Brown, & Tredger, 2014)
Octreotide	3 levels 6 batches 3 replic. (Taylor, 2005)	3 levels 6 replic.	Triptorelin	4	RT 12 h F/T 3 cycles FE 3 days	(Capron, Destree, Maiter, & Wallemacq, 2014)
Imatinib		3 levels 3 blanks	[$^{2}H_8$]-Imatinib			(Rezende et al., 2013)
Lapatinib		3 levels 5 replic.	Sorafenib	3	RT 8 h F/T 3 cycles FE 10 h −70°C 21 days	(Escudero-Ortiz, Perez-Ruixo, & Valenzuela, 2013)
Letrozole	2 levels 6 blanks 1 replic.	3 levels 6 replic.	Anastrozole	3	RT 4 h F/T 3 cycles FE 24 h	(Shao et al., 2016)
Methotrexate	Post-column infusion (Bonfiglio et al., 1999)		[$^{2}H_3$]-Methotrexate	2	RT 4 h F/T 3 cycles FE 24 h 4°C 24 h	(Schofield et al., 2015)
6-Methylmercaptopurine 6-Thioguanine	1 level 1 blank 5 replic. (Matuszewski et al., 2003) 5 blanks	1 level 5 replic.	[$^{13}C_2$,^{15}N]-6-thioguanine [$^{2}H_3$]-6-Methylmercaptopurine		FE 24 h	(Kirchherr, Shipkova, & von Ahsen, 2013)
Everolimus	1 replic.		[$^{2}H_4$]-Everolimus			(Heideloff, Payto, & Wang, 2013) (Annesley, 2003)

TABLE 1 (Continued)

	Calibration		Accuracy and precision		LLOQ	Selectivity		Carryover	Dilution integrity
Analyte(s)	Levels	Range (-fold)	Levels	Replic.	S/N	Endogenous	Exogenous	*n* samples	Levels replic.
Dabrafenib Trametinib	8	100	4	Intra: 15 Inter: 15	5	6 blanks 6 LLOQ	Cross-analyte/IS	2 blanks	1 level 5 replic.
Dasatinib Erlotinib Gefitinib Imatinib Lapatinib Nilotinib Sorafenib Sunitinib	7	50–100	3	Intra: 5 Inter: 5	5				1 level 6 blanks 1 replic.
Abiraterone Enzalutamide	4	100	3	Intra: 15 Inter: 15	5	6 blanks 6 LLOQ	Cross-analyte/IS	2 blanks	1 level 5 replic.
Dasatinib Imatinib Nilotinib	8	533–2000	4	Inter: 6 Intra: 18	5	6 blanks 6 zeros	Cross-analyte/IS	>1 blanks	
Pazopanib	8	50	4	Intra: 15 Inter: 15	5	6 blanks 6 LLOQ	Cross-IS	2 blanks	1 level 5 replic.
Dasatinib Erlotinib Gefitinib Imatinib Lapatinib Nilotinib Pazopanib Sorafenib Sunitinib Vermurafenib	4	20	3	Intra: 15 Inter:15	5	6 blanks 6 LLOQ		2 blanks	1 level 5 replic.
Bosutinib Cobimetinib Dabrafenib Dasatinib Erlotinib Ibrutinib Imatinib Lapatinib Nilotinib Ponatinib Sorafenib Sunitinib Trametinib Vemurafenib	6–8	100–500	4	Intra: 25 Inter: 24		6 blanks	ART, AFT,other		

TABLE 1 (Continued)

Analyte(s)	Matrix effect Levels blanks replic.	Recovery Levels blanks replic.	Internal standards	Short-term stability Levels	 Conditions	Reference
Cobimetinib Dabrafenib Regorafenib Trametinib Vemurafenib	3 levels 7 blanks 1 replic.	3 levels 7 replic.	[$^{13}C_6$]-Cobimetinib [$^{2}H_9$]-Dabrafenib [^{13}C,$^{2}H_3$]-Pazopanib [^{13}C,$^{2}H_3$]-Regorafenib [$^{13}C_6$]-Trametinib [$^{13}C_6$]- Vemurafenib	3	RT 48 h F/T 3 cycles FE 24 h 4°C 48 h	(Cardoso et al., 2018)
Dasatinib Erlotinib Gefitinib Imatinib Lapatinib Nilotinib Sorafenib SunitinibC	2 levels 1 blank 3 replic.	2 levels 3 replic.	[$^{2}H_8$]-Dasatinib [$^{13}C_6$]-Erlotinib [$^{2}H_8$]-Gefitinib [^{13}C, $^{2}H_3$]-Imatinib [^{13}C, $^{2}H_7$]-Lapatinib [$^{2}H_3$]-Nilotinib [^{13}C, $^{2}H_3$]-Sorafenib [$^{2}H_{10}$]-Sunitinib	2	RT 48 h F/T 3 cycles FE 8 days	(Lankheet et al., 2013; Lankheet et al., 2013)
Olaparib Pazopanib Ruxolitinib Vismodegib	3 levels 1 blank (Matuszewski et al., 2003)	3 levels	[$^{2}H_8$]-Olaparib [$^{2}H_5$]-Pazopanib [$^{2}H_9$]-Ruxolitinib [$^{13}C_7$,$^{2}H_3$]-Vismodegib	3	RT 4 days F/T 4 cycles FE 24 h 4°C 4 days	(Pressiat et al., 2018)
Dabrafenib Trametinib	2 levels 6 blanks 1 replic.		[$^{2}H_8$]-Dabrafenib [$^{13}C_6$]-Trametinib	2	RT 24 h F/T 3 cycles 2–8°C 68 −20°C 20 days	(Nijenhuis, Haverkate, Rosing, Schellens, & Beijnen, 2016)
Dasatinib Erlotinib Gefitinib Imatinib Lapatinib Nilotinib Sorafenib Sunitinib	1 level 6 replic.	[$^{2}H_8$]-Gefitinib [$^{2}H_8$]-Imatinib [$^{13}C_2$,$^{15}N_2$]-Nilotinib [$^{2}H_{10}$]-Sunitinib	3	RT 1 day F/T 3	cycles 4°C 1 week	(Couchman et al., 2012)
Abiraterone Enzalutamide	2 levels 6 blanks 1 replic.	2 levels 3 replic.	[$^{2}H_4$]-Abiraterone [$^{2}H_6$]-Enzalutamide	2	RT 5 days F/T 3 cycles FE 5 days −20°C 1 month	(van Nuland, Hillebrand, Rosing, Schellens, & Beijnen, 2017)

TABLE 1 (Continued)

	Calibration		Accuracy and precision		LLOQ	Selectivity		Carryover	Dilution integrity
Analyte(s)	Levels	Range (-fold)	Levels	Replic.	S/N	Endogenous	Exogenous	*n* samples	Levels replic.
Z-Endoxifen	4	25	3	Intra: 15 Inter: 15	5	6 blanks 6 LLOQ	Cross-IS	2 blanks	
Binimetinib Cobimetinib DabrafenibC Trametinib Vemurafenib	7–9	250–1000	3	Intra: 5 Inter: 15		6 blanks 6 zeros	Cross-analyte/IS		
Afatinib Axitinib Ceritinib Crizotinib Dabrafenib Enzalutamide Regorafenib Trametinib	4	100	3	Intra: 15 Inter: 15	5	6 blanks 6 LLOQ	Cross-analyte/IS	2 blanks	1 level 5 replic.
Sunitinib	8	200	4	Intra: 15 Inter: 15		6 blanks 6 LLOQ	Cross-IS	2 blanks	1 level 5 replic.
Imatinib	6	200	3	Intra: 5 Inter: 10	10	5 blanks 5 LLOQ			
Methotrexate	7	1000	4	Intra: 30 Inter: 30		10 blanks 10 LLOQ		1 blank	2 levels
Abiraterone Anastrozole Bicalutamide Enzalutamide Exemestane Letrozole Z-Endoxifen	4	20–200	3	Intra: 15 Inter: 15	5	6 blanks 6 LLOQ		2 blanks	

Abbreviations: S/N, signal-to-noise ratio; replic., replicates; RT, room temperature; F/T, freeze–thaw; FE, final extract; LLOQ, lower limit of quantifications; IS, internal standard; OTC, over the counter; Hemo, hemolytic; Lipi, lipidemic; Icte, icteric; DOA, drugs of abuse; ACD, anticancer drugs; IMS, immunosuppressants; antibiotics; TKI, tyrosine kinase inhibitors; ART, antiretroviral therapy; AFT, antifungal therapy.

TABLE 1 (Continued)

Analyte(s)	Matrix effect Levels blanks replic.	Recovery Levels blanks replic.	Internal standards	Short-term stability Levels	 Conditions	Reference
Dasatinib Imatinib Nilotinib	3 levels 6 blanks 1 replic.	3 levels 6 replic.	[$^{2}H_{8}$]-Dasatinib [$^{2}H_{8}$]-Imatinib [^{13}C,$^{2}H_{3}$]-Nilotinib	2	RT 48 h F/T 3 cycles FE 24 h 4°C 24 h	(Wojnicz et al., 2017)
Pazopanib	2 levels 6 blanks 1 replic.	2 levels 5 replic.	[^{13}C,$^{2}H_{3}$]-Pazopanib	2	RT 5 days F/T 3 cycles FE 70 days	(Verheijen et al., 2018)
Dasatinib Erlotinib Gefitinib Imatinib Lapatinib Nilotinib Pazopanib Sorafenib Sunitinib Vermurafenib			[$^{2}H_{8}$]-Dasatinib [$^{13}C_{6}$]-Erlotinib [$^{2}H_{3}$]-Gefitinib [^{13}C]-Imatinib [^{13}C]-Lapatinib [$^{2}H_{3}$]-Nilotinib [^{13}C]-Pazopanib [^{13}C]-Sorafenib [$^{2}H_{10}$]-Sunitinib [$^{13}C_{6}$]-VermurafenibC	2	RT 48 h F/T 3 cycles FE 8 days −20°C 1 month	(Herbrink et al., 2016)
Bosutinib Cobimetinib Dabrafenib Dasatinib Erlotinib Ibrutinib Imatinib Lapatinib Nilotinib Ponatinib Sorafenib Sunitinib Trametinib Vemurafenib	3 levels 1 blank (Matuszewski et al., 2003)	3 levels	[$^{2}H_{9}$]-Bosutinib [^{13}C]-Cobimetinib [$^{2}H_{9}$]-Dabrafenib [$^{2}H_{8}$]-Dasatinib [$^{13}C_{6}$]-Erlotinib [$^{2}H_{5}$]-Ibrutinib [$^{2}H_{8}$]-Imatinib [^{13}C,$^{2}H_{7}$]-Lapatinib [^{13}C,$^{2}H_{3}$]-Nilotinib [$^{2}H_{8}$]-Ponatinib [^{13}C,$^{2}H_{3}$]-Sorafenib [$^{2}H_{10}$]-Sunitinib [$^{13}C_{6}$]-Trametinib [$^{13}C_{6}$]-Vemurafenib	3	RT 48 h F/T 4 cycles FE 24 h 4°C 48 h	(Huynh et al., 2017)
Z-Endoxifen	6 blanks 1 replic.		[$^{2}H_{5}$]-Z-endoxifen	2	RT 7 days F/T 3 cycles FE 7 days −20°C 7 days 2–8°C 7 days	(de Krou, Rosing, Nuijen, Schellens, & Beijnen, 2017)
Binimetinib Cobimetinib DabrafenibC Trametinib Vemurafenib	1 level 6 blanks 1 replic.	1 level 3 replic.	[$^{13}C_{2}$,$^{2}H_{4}$]-Binimetinib [$^{13}C_{6}$]-Cobimetinib [$^{2}H_{9}$]-Dabrafenib [$^{13}C_{6}$]-Trametinib [$^{13}C_{6}$]-Vemurafenib	2	RT 24 h 4°C 3 days F/T 3 cycles −20°C 1 month	(Rousset et al., 2017)

TABLE 1 (Continued)

Analyte(s)	Matrix effect Levels blanks replic.	Recovery Levels blanks replic.	Internal standards	Short-term stability Levels	 Conditions	Reference
Afatinib Axitinib Ceritinib Crizotinib Dabrafenib Enzalutamide Regorafenib Trametinib	2 levels 6 blanks 1 replic.		[$^{13}C_6$]-Afatinib [^{13}C,$^{2}H_3$]-Axitinib [$^{2}H_7$]-Ceritinib [$^{13}C_2$]-Crizotinib [$^{2}H_9$]-Dabrafenib [$^{2}H_6$]-Enzalutamide [^{13}C,$^{2}H_3$]-Regorafenib- [$^{13}C_6$]-Trametinib	2	RT 48 h F/T 3 cycles FE 48 h −20°C 1 month	(Herbrink et al., 2018)
Sunitinib	2 levels 1 blank 3 replic.	2 levels 3 replic.	[$^{2}H_{10}$]-Sunitinib	2	RT 72 h F/T 3 cycles FE 7 days	(Lankheet, Hillebrand, et al., 2013; Lankheet, Steeghs, et al., 2013)
Imatinib	1 level 4 blanks 1 replic.	3 levels	[$^{2}H_8$]-Imatinib			(Arellano, Gandia, Lafont, Jongejan, & Chatelut, 2012)
Methotrexate	4 levels 6 blanks 3 replic.	4 levels 3 replic.	[$^{2}H_3$]-Methotrexate	4	F/T 3 cycles −20°C 48 h 4°C 72 h −80°C 34 days	(Mei et al., 2018)
Abiraterone Anastrozole Bicalutamide Enzalutamide Exemestane Letrozole Z-Endoxifen			[$^{2}H_4$]-Abiraterone [$^{2}H_{12}$]-Anastrozole [$^{2}H_4$]-Bicalutamide [$^{2}H_6$]-Enzalutamide [$^{2}H_3$]-Exemestane [$^{2}H_4$]-Letrozole [$^{2}H_5$]-Endoxifen	2	RT 5 days F/T 3 cycles 4°C 5 days FE 5 days −20°C 21 weeks	(van Nuland et al., 2019)

TABLE 2 Recommendations for bioanalytical method validation as given by the European Medicines Agency (EMA), the US Food and Drug Administration (FDA) and proposed recommendations specifically for therapeutic drug monitoring (TDM) assays

Validation parameter	Experiments/ criteria	EMA	FDA	TDM
Calibration model	Experiments	Consists of a blank sample, a zero sample and 6–8 calibration standards (incl. LLOQ)	Consists of a blank sample, a zero sample and 6–8 calibration standards (incl. LLOQ)	Consists of a blank sample, a zero sample and 4 calibration standards (incl. LLOQ)
	Acceptance criteria	85–115% 80–120% for LLOQ 75% should meet the criteria, including LLOQ and ULOQ	85–115% 80–120% for LLOQ 75% should meet the criteria, including LLOQ and ULOQ	85–115% 80–120% for LLOQ 75% should meet the criteria, including LLOQ and ULOQ
LLOQ	Experiments	Lowest calibration standard level	Lowest calibration standard level	Lowest calibration standard level
	Acceptance criteria	80–120% ≥5 S/N	80–120% ≥5 S/N	80–120% ≥10 S/N
Carryover	Experiments	Blank sample injected after a high sample	Should be monitored during analysis	At least 2 blank samples injected after the ULOQ
	Acceptance criteria	≤20% of LLOQ ≤5% of IS	≤20% of LLOQ	≤20% of LLOQ ≤5% of IS
Accuracy and precision	Experiments	4 concentration levels 5 samples per level	4 concentration levels 5 samples per level	3 concentration levels (LLOQ, mid = target concentration, ULOQ), 5 samples per level
	Acceptance criteria	85–115% 80–120% for LLOQ	85–115% 80–120% for LLOQ	85–115% 80–120% for LLOQ
Dilution integrity	Experiments	Dilute sample > ULOQ (n = 5) with blank matrix	Dilute sample > ULOQ (n = 5) with blank matrix	Not applicable
	Acceptance criteria	85–115% 80–120% for LLOQ	85–115% 80–120% for LLOQ	
Endogenous Selectivity	Experiments	6 batches, blank samples	6 batches, blank samples and at LLOQ	6 Batches, blank samples and at LLOQ
	Acceptance criteria	≤20% of LLOQ ≤5% of IS	≤20% of LLOQ ≤5% of IS	≤20% of LLOQ ≤5% of IS
Exogenous Selectivity	Experiments	Potential interfering substances should be tested separately	Cross-interference when >1 analyte in the assay	If applicable: interference of structural analogs
Stability	Experiments	Low and high concentrations: stock solutions, working solutions, F/T, short-term at RT, long-term If applicable: dry extract, autosampler stability	Low and high concentrations: stock solutions, working solutions, F/T 3 cycles, short-term at RT, long-term If applicable: dry extract, autosampler stability	LLOQ and ULOQ concentrations: Stock solutions, Working solutions, F/T 3 cycles, short-term at RT, prolonged at RT during transport, long-term, influence of exposure to light If applicable: dry exactly, authosampler stability
	Acceptance criteria	85–115%	85–115%	85–115%
Matrix effect	Experiments	6 batches of blank matrix, low and high samples	Matrix effect should be evaluated	Not applicable if a stable isotopically labeled internal standard is used co-eluting with the analyte
	Acceptance criteria	CV of IS-normalized should be <15%		
Recovery	Experiments	Not applicable	Extracted compared with unextracted at 3 concentration levels	Not applicable

*Percentage of nominal concentration unless otherwise specified.

Abbreviations: ULOQ, upper limit of quantification; CV, coefficient of variance.

PK–TK studies, FDA and EMA guidelines have reached consensus on the experiments and acceptance criteria for the calibration model. The matrix of the calibration standards should, if possible, represent the matrix in study samples and fresh calibration standards should be prepared prior to each validation run. The number of calibration standards should be anticipated on the validation range with a minimum of six standards, including an LLOQ sample. Additionally, each set of calibration standards should include a blank sample (processed matrix sample without analyte and without internal standard) and a zero sample (processed matrix sample without analyte). These samples are not included in the calculation of the regression line. The EMA recommends the analysis of calibration standards on three occasions in duplicate (total $n = 6$) to evaluate linearity of the calibration model. Acceptance criteria for calibration standards are 85–115% of the nominal concentration, and 80–120% for the LLOQ. At least 75% of calibration standards should meet these criteria, including the LLOQ (and the upper limit of quantification (ULOQ) in EMA guidelines).

4.1 | Calibration standards

Included analytical assays all used a matrix similar to study samples for preparation of calibration standards. Most papers do not describe whether calibration standards were produced freshly before each validation run. The number of calibration standards ranged from four to nine, with a median number of seven and a median calibration range of 100-fold. The median number of calibration standards is in line with the recommended guidelines. However, nine articles use fewer than six calibration standards for the calibration model. When dividing the calibration range by the number of calibration standards, a median of 24 nominal concentration units per standard is calculated. Based on this median, a 100-fold calibration range would only need four calibration standards. Reducing the analysis time by using fewer calibration standards ($n = 3$) was previously demonstrated by Lankheet et al. in a method comparison of LC–MS/MS assays for the quantification of sunitinib (Lankheet, Hillebrand, et al., 2013; Lankheet, Steeghs, et al., 2013). Reducing the number of calibration standards from six to three increased the turnaround while preserving accuracy and precision. To further investigate this concept, we performed a similar experiment in our laboratory using data from previously published TDM assays for quantification of pazopanib, trametinib and dasatinib (Nijenhuis et al., 2016; Verheijen et al., 2018). QC samples at LLOQ, low, mid and high level ($n = 15$) were analyzed using both the original method with eight calibration standards and an adjusted method with four calibration standards. The results of the method comparison are shown in Figure 2. A regression test was performed for all three analytes and showed no significant constant error (*y*-axis intercept 95% confidence intervals contained zero: −0.0719–0.0607 for pazopanib, −0.0599–0.0433 for trametinib and −5.55–9.03 for dabrafenib). Furthermore, the slopes of the regression lines were not significantly different from 1 for pazopanib and dabrafenib (95% confidence interval: 0.996–1.00 and 0.997–1.00, respectively). Although the regression line for trametinib was found to be significantly different from 1 by reducing the number of calibration standards, with a slope of 1.01 (95% confidence interval: 1.01–1.02), the accuracy and precision improved compared with the original method from ±4.3 to ±3.8% and from ≤5.6 to ≤3.0%, respectively. These data suggest that reducing the number of calibration standards from eight to four when using calibration ranges of 100-fold or less does not affect the accuracy and precision of the method.

From a clinical point of view, target attainment is the final objective for decision making in TDM. Therefore, a one-level calibration could be considered with a calibration point being the target concentration. In a previous study, bias and precision of multiple-point and one-point calibration were compared. One-point calibration with a calibration close to the center of the complete calibration range (e.g. proposed target) shows bias and precision within the acceptance criteria for the majority of drugs (Peters & Maurer, 2007). However, dose adjustments following TDM may depend on the quantitative determination of the concentration of an anticancer agent; patients with a concentration around the target could receive minimal or no dose adaptations, while large deviations from the target may ask for other interventions. Therefore a concentration range should be chosen per analyte depending on the decision making in TDM. A calibration range that spans two orders of magnitude using four calibration standards is in most cases sufficient for these purposes and this reduction in the number of calibrations standards increases the turnaround time of TDM assays and has no impact on the quality of the reported data, as demonstrated in Figure 2. Furthermore a one-level calibration

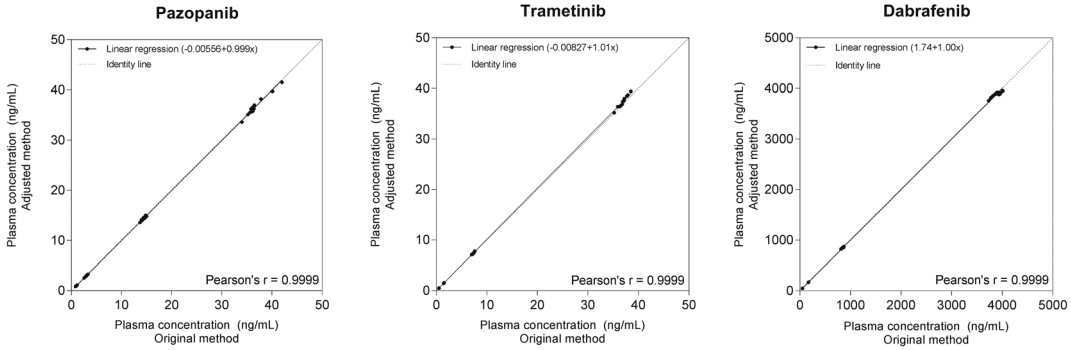

FIGURE 2 Scatter plots of method comparison showing plasma concentrations of four quality control (QC) levels ($n = 15$) measured with the original method (eight calibration standards) and the adjusted method (four calibration standards). The black line represents the linear regression line and the dotted line represents the line of identity

assumes a linear model and a variance independent of the analyte concentration (no weighting factor is applied). This assumption is in most MS methods not justified. Therefore the calibration model should be established in each analytical run by analyzing four calibration standards in the chosen, validated range.

4.2 | Regression line

All but one paper (Cardoso et al., 2018) used a weighted linear regression ($1/x$ or $1/x^2$) for the calibration model. A linear relationship is the simplest mathematical relationship with a constant accuracy over the complete range in contrast to quadratic fitting (Clinical and Laboratory Standards Institute, 2003). Therefore, linear regression is the preferred mathematical method for calibration of analytical methods (Gu, Liu, Wang, Aubry, & Arnold, 2014). A weighting factor of 1, $1/x$ or $1/x^2$ is selected if the standard deviation of the instrument response is proportional to x (Gu et al., 2014). Weighted regression of $1/x$ or $1/x^2$ should be used when the absolute variance is not constant for all observations, which is generally the case with a calibration range covering over one order of magnitude (Gu et al., 2014). Therefore, $1/x$ or $1/x^2$ weighting may be used to improve the accuracy at lower concentrations. If a quadratic fit is chosen to compensate for saturation of the ion detector, the method could be de-optimized to reduce saturation, or the MRM channel could be modified (+1 or + 2) to monitor the m/z values of isotopes and thereby avoid signal saturation (Liu, Lam, & Dasgupta, 2011). All but five articles report a determination coefficient (R^2) and a minimum of 0.99 is generally strived for. However, deviations from the nominal concentrations provide more information about linearity of the calibration model. Therefore, back-calculated concentrations should be reported instead of R^2. Acceptance criteria for the back-calculated calibration standards are provided by 24 papers, being 85–115% of the nominal concentration (80–120% for the LLOQ).

4.3 | Quantitation range

The quantitation range of bioanalytical assays should be chosen on the basis of concentrations expected in clinical samples. TDM assays are developed to determine whether individual concentrations are above or below a certain target and, therefore, the concentration range should be built around this target concentration. A median calibration range of 100-fold was used in included assays for TDM purpose. The calibrations range should be as narrow as possible for high accuracy and precision, covering the concentration of the majority of samples as seen in the clinic, from the minimum reported concentration to the maximum reported concentration after drug intake. Accordingly, the range will depend on inter-patient variability of anticipated concentrations. In our experience, a range of 20- to 100-fold is in most cases sufficient.

4.4 | Accuracy and precision

Accuracy of the LC–MS/MS method describes the closeness of mean measured concentrations to the nominal concentrations of the analyte and is expressed as a percentage, while the precision of the method describes the closeness of repeated measurements of an analyte. For PK–TK assays, both parameters should be assessed using quality control (QC) samples, i.e. spiked samples at known concentrations. QC samples are generally produced at LLOQ, low (within 3 times the LLOQ), mid (in the midrange) and high (approaching the top end, >75% of ULOQ, of the calibration range) level. Accuracy and precision can be further subdivided into within-assay and between-assay accuracy and precision. According to the EMA and FDA, within-assay accuracy and precision should be determined by measuring a minimum of five samples at a minimum of four concentration levels (LLOQ, low, mid, high). Furthermore, between-assay accuracy and precision should be assessed by measuring four concentration levels in at least three runs or batches on at least two different days. Mean concentrations should be 85–115% of the nominal values for QC samples, except for the LLOQ for which 80–120% is considered acceptable. It is recommended by the EMA to demonstrate accuracy and precision over at least one of the runs in a size equivalent to a prospective analytical run containing study samples.

Included analytical papers for TDM purpose determine accuracy and precision, with a minimum of three concentration levels. Although a variable number of QC samples was used for determining accuracy and precision, all papers included at least five samples to determine within-assay accuracy and precision and a minimum of three runs were performed for between-assay accuracy and precision. Only seven papers did not give acceptance criteria for accuracy and precision, while other papers reported acceptance criteria in line with FDA and EMA guidelines. These results suggest that recommendations in FDA and EMA guidelines are generally acceptable for determining accuracy and precision of TDM assays. Regarding the short calibration range of TDM assays and the aim for a fast turnaround, we believe that a minimum of three concentration levels (LLOQ, medium or target concentration an ULOQ) is sufficient. As most papers do not provide information on how accuracy and precision were calculated, we recommend using the following equations (Rosing, Man, Doyle, Bult, & Beijnen, 2000):

$$\text{Within-assay accuracy}\,(\%) = 100\%\,(\text{mean measured concentration per run} - \text{nominal concentration})/(\text{nomimal concentration}) \quad (1)$$

$$\text{Between-assay accuracy}\,(\%) = 100\%(\text{overall mean measured concentration} - \text{nominal concentration})/(\text{nominal concentration}) \quad (2)$$

$$\text{Within-assay precision}\,(\%) = 100\%(\text{SD of the measured concentration per run})/(\text{mean measured concentration per run}) \quad (3)$$

$$\text{Between-assay precision}\,(\%) = \frac{\sqrt{\left(\frac{s^2_{\text{overall}}/((n_1+\ldots+n_a-1)-((n_1-1)s_1^2+\ldots+((n_a-1)s_a^2)))}{a-1}\right)-\left(\frac{(n_1-1)s_1^2+\ldots+(n_a-1)s_a^2}{n_1+\ldots+n_a-a}\right)}}{\left(\frac{n}{a}\right)}\times 100\% \quad (4)$$

Mean of runs

where SD = standard deviation, $s^2_{overall}$ = overall SD^2, S^2_x = variance (SD^2) of mean of replicates on a concentration level for run x, a = number of runs and n = number of replicates.

5 | LOWER LIMIT OF QUANTIFICATION

For assays for PK–TK studies, the LLOQ is the lowest level of the calibration standards which can determined with an accuracy and precision of ≤20% of the nominal concentration. Both EMA and FDA guidelines state that the LLOQ should be at least 5 times the signal of a blank sample.

The LLOQ in TDM assays is the lowest level of the calibration standards; however, it is generally not the lowest concentration of an analyte which can be quantified reliably as the concentration range is higher. Therefore, the LLOQ in TDM assays is rather a lower limit of the measuring interval (LLMI). In 34 of 36 papers of TDM assays, the LLOQ was defined as the lowest level of the calibration range, with an acceptance criterion of 80–120% of the nominal concentration. Based on these results, a maximum of ±20% deviation from the nominal concentration seems to be accepted in TDM assays. The signal-to-noise (S/N) ratio was provided in 19 papers, being at least 5 (n = 12) or 10 (n = 7). Furthermore, the limit of detection was determined in 10 papers, with a S/N ratio of at least 3 (n = 8), five (n = 1) or 6 (n = 1). To set an example, in our laboratory, we perform weekly TDM measurements of *Z*-endoxifen with a validated LC–MS/MS assay (de Krou et al., 2017), and we have recorded the S/N ratio of the LLOQ (1 ng/mL) for 49 weeks. Figure 3 displays the S/N ratio of the LLOQ to range from 20 to 200 in this time period, demonstrating a factor 10 interoccasion variability when the method is applied for a longer period. Although the EMA and FDA recommend a S/N ratio of at least 5 for the LLOQ, we believe this limit should be increased for TDM assays regarding the between-assay variability of the LC–MS/MS signal over a long time period. Therefore, we aim for a S/N ratio of at least 10 instead of 5. Increasing the S/N ratio is supported by the CLSI guidelines, in which a S/N ratio of at least 20 is recommended (Lynch, 2016). Moreover, TDM assays are developed to measure steady-state drug concentrations and while choosing calibration standards to cover concentrations in a 20- to 100-fold range, the LLOQ will generally exceed an S/N ratio of 10. For example, enzalutamide is known to have a mean trough concentration at steady state (at a 160 mg dose) of 11.4 mg/L (Gibbons et al., 2015). A validated method in our laboratory showed a S/N ratio of over 200 for the LLOQ of 5 ng/mL (van Nuland et al., 2017). Taken the variability in account over time and the intended use of TDM methods, we strongly advise increasing the S/N ratio at the LLOQ to increase the robustness of the validated method.

6 | SELECTIVITY

The selectivity of the analytical method is investigated during validation to assess whether the method is able to differentiate the analyte of interest from endogenous and exogenous components within the sample. EMA and FDA guidelines state that selectivity should be proven in at least six independently prepared and analyzed batches of the used biomatrix for PK–TK assays. The interference in these should be ±20 and 5% of the LLOQ for the analyte and the internal standard, respectively. According to the FDA, selectivity should also be ensured at the LLMI. These experiments focus on interference from endogenous source, while it may also be necessary to investigate potential interference from exogenous components, such as metabolites, co-medication, degradation products, excipients of the formulation and other xenobiotics. The FDA specifically adds that "if the method is intended to quantify more than one analyte, each analyte should be tested to ensure that there is no interference".

6.1 | Endogenous selectivity

In 26 of 36 articles for TDM assays, endogenous selectivity was tested in accordance with the guidelines, in at least six different batches of blank matrix. One paper included 10 different batches and two papers investigated endogenous interference in one batch of plasma. Testing for selectivity is important, as it shows that the substance quantified is indeed the analyte. Therefore, selectivity experiments should be performed in different batches of plasma, also for TDM assays. Selectivity at LLOQ was ensured in 14 papers, all analyzing LLOQ samples in six different batches. Although the EMA does not recommend performing

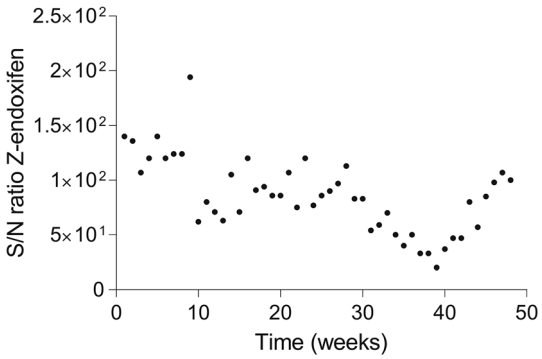

FIGURE 3 *Z*-Endoxifen signal-to-noise (S/N) ratios of the LLOQ (1 ng/mL) as measured by LC–MS/MS for therapeutic drug monitoring for 49 weeks

such an experiment, we believe it is important to investigate the effect of different matrices and endogenous interference on quantification of the analyte. Endogenous components may cause suppression or enhancement of the MS signal and thereby influence quantification of the analyte. This can only be characterized if selectivity has been investigated in biological samples spiked with the analyte of interest. Therefore, we advise assessing endogenous selectivity in control matrix spiked at the LLOQ in six different batches of plasma. One of the papers included in this review investigated the selectivity in hemolyzed, lipidemic and icteric plasma and showed that the assay was not affected. If samples from a special population are included in TDM, such as renally or hepatically impaired populations, it is recommended to study the selectivity in such related samples (European Medicines Agency, 2011).

6.2 | Exogenous selectivity

Potential interferences from nonendogenous sources were investigated in 23 articles for TDM assays; six papers assessed interference of co-medication, 14 papers investigated cross-analyte interferences, two papers studied metabolite interference and one paper analyzed potential interference of photodegradation products of the analyte. Commonly investigated co-medication are over-the-counter drugs, drugs of abuse, immunosuppressants, antibiotics, antiretroviral therapy and antifungal therapy. Testing for potential interference of co-medication and cross-analyte interference is advised if these are structural analogs of the analyte of interest (de Zwart et al., 2016). Otherwise, no further testing is required. For some analytes, such as endoxifen and abiraterone, one needs to be aware of the presence of isomers (Jager, Rosing, Linn, Schellens, & Beijnen, 2012; van Nuland et al., 2017). To preserve selectivity, isomers should be separated at baseline from the analyte. In addition to testing for interferences of other drugs, the EMA recommends to investigate potential interference of excipients in the drug formulation, such as polyethylene glycol or polysorbate. This may be important for intravenously administered drugs; however, TDM will mainly be applied to orally administered drugs, and therefore it is of less relevance.

7 | CARRYOVER

Sample carryover can be a major problem, which influences accuracy and precision of the method. Carryover is caused by residual analyte from a sample analyzed earlier in the analytical run or batch (Hughes, Wong, Fan, & Bajaj, 2007). Both EMA and FDA guidelines recommend monitoring the carryover during validation of PK–TK assays by injecting blank samples after the ULOQ. These blanks should contain ≤20% of the LLOQ response and ≤5% of the internal standard response.

Surprisingly, carryover was investigated in only 20 papers of TDM assays. This simple experiment is important for TDM assays, as high concentrations may occur in patient samples and it should be ensured that these levels do not influence quantification in the next injected sample. In 17 papers, carryover was determined by injecting matrix blanks after the ULOQ. Two papers assessed carryover by injecting QC low samples before and after QC high samples (French et al., 2014; Heideloff et al., 2013). The carryover was defined as the mean difference of QC low samples injected prior to QC high samples and QC low samples injected after QC high samples. Only one article provided acceptance criteria for this experiment, with the mean difference between the low samples before and after the high samples being ≤20% (Heideloff et al., 2013). The carryover in such experiments is difficult to observe, because it involves an additive effect rather than the absence/presence of a peak. Furthermore, the carryover was determined after injection of a QC high sample instead of the ULOQ. One article assessed carryover by injecting an organic solvent sample after the ULOQ instead of a matrix blank (Rezende et al., 2013). Matrix blanks have a similar composition and ion strength as study samples and should therefore be used to determine carryover, while organic blank samples generally do not show carryover. Determining carryover according to EMA and FDA guidelines is rapid and easy to perform and interpret; therefore we recommend using these guidelines for the evaluation of carryover during the validation and for routine assessment in each analytical run because carryover may vary from run to run. It is, however, important to be aware of the difference between carryover and memory effect as these problems may be resolved differently. A memory effect is observed as a downward-drifting baseline in a blank sample that is analyzed after a high-concentration sample and suggests that the analyte was still eluting off the column from the previous injection (Hughes et al., 2007). Both carryover and memory effect may affect quantification of low concentrations by a residual analyte peak or by an increased baseline, respectively, and should therefore be minimized. A procedure to evaluate the carryover during the application phase of the validated assay is described under Section 11.

8 | DILUTION INTEGRITY

Dilution integrity is evaluated with the purpose of measuring samples above the ULOQ. The dilution of samples should not affect the accuracy and precision of the measurement. According to EMA and FDA, dilution integrity should be demonstrated by spiking a sample at a concentration above the ULOQ and consequently diluting this sample to a concentration within the calibration range. The accuracy and precision of a sample set (n = 5) should be ±15% of the nominal concentration.

Dilution integrity was evaluated in 16 papers for TDM assays. Only 10 papers provided acceptance criteria for the conducted experiments and more than one dilution step was investigated in seven of 15 papers. Dilution of samples is not common practice for TDM as it is time consuming and therefore decreases the throughput. Furthermore, exceeding the ULOQ in general means that the target was attained. Therefore, samples exceeding the calibration range can be reported as above the ULOQ without further analysis. However, we would prefer to minimize the number of samples exceeding the ULOQ by choosing a calibration range that covers clinically observed concentrations. TDM can also be used for monitoring toxicities in which

quantification of high concentrations could be important for the clinical perspective and treatment strategy. When a TDM assay is developed for this purpose, dilution integrity should be demonstrated to cover a larger calibration range and we would recommend incorporating the EMA/FDA experiments.

9 | MATRIX EFFECT AND RECOVERY

Matrix effect can be assessed using a variety of methods described in the literature, for example post-column infusion and post-extraction techniques (Bergeron & Garofolo, 2013). Post-column infusion was first described by Bonfiglio et al.,(1999) and consists of injecting a blank pretreated biological sample during continuous post-column infusion of the analyte of interest. A matrix effect may be observed by comparing changes from baseline across the chromatographic run. For post-extraction techniques a set of samples with and without biomatrix is used. The matrix effect can be calculated by comparison of the analyte response in presence and in absence of biomatrix. The latter method is recommended by the EMA for PK–TK assays, using at least six different batches of blank matrix. For each analyte, the matrix factor (MF) and internal standard (IS)-normalized MF should be calculated and the coefficient of variation (CV) is acceptable when it is ≤15%. A version of the post-extraction spike method has also been described by Matuszewski et al. (2003). Peak areas are compared in three sets of five samples: set one consists of samples in neat solution (mobile phase), set two of matrix blanks spiked with the analyte after sample preparation and set three of processed samples spiked before sample preparation. Sets two and three should be constructed in five different batches of blank matrix. The matrix factor is then calculated by the ratio of sets one and three, while the recovery is calculated by the ratio of sets two and three. FDA guidelines state that matrix effect should be evaluated for PK–TK assays, but do not describe how to achieve this. However, the FDA does describe recovery experiments by comparing the area of extracted samples with unextracted samples at three concentrations (low, mid, high).

Two papers of TDM assays performed matrix effect experiments according the post-column infusion method of Bonfiglio et al. (Bonfiglio et al., 1999; King et al., 2000). The effect of a blank matrix can be monitored with this method; however, the effect on quantification of the analyte is not investigated. Therefore, post-extraction is the preferred method for determination of the matrix effect (Marchi et al., 2010; Matuszewski et al., 2003; Taylor, 2005). This is also reflected by the results of the literature search, as 26 papers investigated the matrix effect with this technique. Whether and how the matrix effect should be examined is still a matter of debate, especially since no acceptance criteria are given by guidelines as to what extent the matrix effect is thought to be acceptable. Poor reproducibility owing to matrix effects will be reflected by a low accuracy and precision. This is already investigated in six different batches at LLOQ in the endogenous selectivity experiments. Furthermore, the use of isotopically labeled internal standards can compensate for matrix effects regarding reproducible quantification of the target analyte, and therefore, it is not necessary to determine the matrix factor in different batches (Viswanathan et al., 2007). In 27 papers, a stable isotopically labeled internal standard was included in the assay; 16 papers used a ^{13}C- or ^{15}N-labeled internal standard, and 11 papers used a deuterated internal standard. Other articles used structural analogs as internal standard. As isotopically labeled internal standards are structurally similar to the analyte of interest, they will have a similar matrix effect and are therefore able to correct for matrix-related variability (Annesley, 2003). However, deuterated internal standards may have a slightly different retention time than the analyte, caused by deuterium isotope effects, which is not observed for ^{13}C- or ^{15}N-labeled internal standard (Wang, Cyronak, & Yang, 2007). Wang et al. demonstrated that a deuterated internal standard had a different degree of ion suppression owing to a slight difference in retention time, causing a significant matrix effect (Wang et al., 2007). Therefore, ^{13}C- or ^{15}N-labeled internal standards should, if available, be first choice rather than deuterated internal standards. Furthermore, since a compound and its internal standard will theoretically co-elute, it is important to have a mass difference between those compounds to be able to separate them in the mass spectrometer to prevent cross talk. For small molecules a mass difference of at least 3 mass units is in most cases sufficient (Stokvis, Rosing, & Beijnen, 2005).

Recovery was studied in 27 TDM assays according to the post-extraction method. These experiments show the degree of analyte that is extracted during sample preparation. Recovery might be of interest if the extraction is low and the sensitivity is not sufficient for the purpose of the method. Such problems should already be addressed during method development to optimize the assay before method validation. Inconsistent and irreproducible recoveries have not been described in the selected TDM assays. Taking the above into account, determining matrix effect and recovery is not mandatory for validation of TDM methods when an isotopically labeled internal standard is used co-eluting with the analyte of interest.

10 | STABILITY

Stability should be evaluated to ensure that storage conditions do not affect the concentration of the analyte. Therefore, stability needs to be established at every step of the analytical method. EMA and FDA guidelines recommend testing stock stability and the EMA advises also testing the stability of working solutions for PK–TK assays. Furthermore, the following stability conditions should be evaluated: freeze–thaw (F/T) stability of at least three cycles, short-term stability at room temperature, long-term stability under the same conditions as study samples are kept and, if applicable, other stability experiments, such as dry extract stability and the stability of processed samples. Stability experiments should be executed at low and high concentrations.

Although stability in stock and working solutions was described in only 14 papers for TDM assays, all papers reported stability in biomatrix, either experimental or from the literature. Stability

experiments were performed at least at low and high concentrations. Most papers did not describe how many replicates were used. Freeze–thaw stability was assessed in 24 papers, of which four tested more than three F/T cycles. For TDM, F/T stability of three cycles should be sufficient, as samples are frozen after withdrawal and generally measured after the first thaw cycle. Additional F/T stability of up to three cycles is important for potential reanalysis of samples. Short-term stability at room temperature was tested in 28 papers, varying from 4 h up to 28 days. Stability at room temperature is pivotal for transporting samples to the laboratory and during sample preparation. TDM assays are not always available in the hospital where blood withdrawal takes place. Therefore, it should be investigated whether samples can be transported at room temperature or should be transported on dry-ice. Furthermore, it is relevant to investigate the stability when exposed to light as blood collection tubes are generally transparent. Nonetheless, stability at room temperature while exposed to light and in the dark was examined in only five papers (de Krou et al., 2017; Escudero-Ortiz et al., 2013; Herbrink et al., 2018; Nijenhuis et al., 2016; van Nuland et al., 2019). The stability of processed samples was examined in 23 papers, either as re-injection reproduciblity or as final extract stability. EMA and FDA guidelines state that stability of processed samples should be measured if applicable, which is certainly the case for TDM to safeguard the possibility of re-analysis after system failure. When investigating the stability in processed samples, final extract stability is recommended instead of re-injection stability to facilitate re-analysis of samples with fresh calibration standards. Long-term stability (>1 month) was described in nine studies. Although long-term stability is not *per se* important for TDM measurements, because results are reported directly for routine clinical care, it might be useful for determining the shelf-life of calibration standards and quality control samples.

11 | ANALYSIS OF STUDY SAMPLES

EMA and FDA guidelines on bioanalytical validation also provide recommendations for application of the validated method for PK–TK assays. Before starting analysis of study samples, the performance of the bioanalytical method should be verified. Similar to validation runs, the analytical run should consist of a blank sample, a zero sample and at least six calibration standards. At least one set of calibration standards should be analyzed and ≥75% of the standards should be within 85–115% of their nominal concentration (80–120% for the LLOQ). If the LLOQ or ULOQ should be rejected in one analytical run, then the second lowest sample will become the LLOQ and the second highest sample the ULOQ. A minimum of three QC concentration levels in duplicate should be interspersed with study samples. The FDA and EMA recommend a minimum number of QC samples of at least 5% of the number of the clinical study samples or a total of six samples, whichever is greater. The accuracy at each concentration level should be ±15% of the nominal concentration and at least two out of three QC samples (one at each level) should be within the acceptance criteria. According to the FDA, the carryover should be assessed and monitored during sample analysis.

Literature reporting analysis of study samples in TDM are sparse. Only one paper described the use of a system suitability test (SST) to prime the system (Rezende et al., 2013) and two papers specified routine sample analysis (Dennison et al., 2008; Lankheet, Hillebrand, et al., 2013 ; Lankheet, Steeghs, et al., 2013). An SST is an integral part of the analytical procedure to ensure the performance of the analytical system. Critical elements of the analytical system should be included in the SST, such as the check for chromatographic separation of isomers. Methods for routine TDM should aim for a rapid turnaround by implementation of a short analytical run. Therefore, a method with fewer calibration standards and QC samples is proposed, consisting of a blank sample, a zero sample and four calibration standards, and three QC mid samples (or at least 5% of the study samples whichever is higher). The calibration standards should be injected at the beginning of each analytical run and at least three out of four (75%) calibration standards should be within 85–115% of their nominal concentration (80–120% for the LLOQ), including the LLOQ and the ULOQ to maintain the anticipated range. QC mid samples are injected after the calibration standards and at the end of the sequence to ensure adequate accuracy and precision for the whole analytical run. A similar strategy with only three calibration standards and one QC mid sample proved accurate and robust for the quantification of sunitinib (Lankheet, Hillebrand, et al., 2013; Lankheet, Steeghs, et al., 2013). The concentration of the QC mid should have a similar concentration to the TDM target, which is the concentration at which dose adjustments are recommended. This concentration is the most critical value to be quantified accurately as results will be reported as being below or above this target. Therefore, we recommend the concentration of QC mid samples to be similar to the target. Furthermore, carryover should be assessed and monitored in each analytical run by injection two blanks samples after the ULOQ. If the carryover exceeds ±20% of the LLOQ response, the integrity of the bioanalytical data should be assessed by calculation the carryover matrix factor for all samples with the following equation:

$$\text{Carry - over factor} = \frac{\left(\frac{\text{Response}_x \cdot \text{Mean response}_{CB}}{\text{Mean response}_{ULOQ}}\right)}{\text{Response}_{x+1}} \times 100\%$$

in which x = sample x, CB = carryover in the blank, ULOQ = upper limit of quantification and $x + 1$ = the sample injected after sample x.

If the CF is >5%, there is a significant carryover effect of sample x on sample $x + 1$ and therefore, sample $x + 1$ should be reanalyzed. However, if the CF is ≤5%, there is no significant effect of sample x on samples $x + 1$ and the result of sample $x + 1$ can be accepted. This carryover protocol gives the opportunity to determine the impact of the carryover on each sample and to accept those results that are not affected by it. In contrast to the analysis of multiple samples of one subject in PK studies, TDM sample concentrations may vary in concentration from sample to sample. It is important to establish that

the carryover has no impact on the quality of the generated data in each study run.

12 | OVERVIEW OF RECOMMENDATIONS

In this review, method validation of TDM assays is described as compared with method validation of assays for PK–TK studies. For each validation parameter, recommendations of FDA and EMA guidelines for validation of PK–TK studies were described, followed by recommendations that could serve as a guide for future validations and analysis of study samples of TDM assays. A summary of recommendations specifically for validation of TDM assays is included in Table 2. The literature search showed that TDM assays are generally validated based on FDA and EMA guidelines; however, most articles do not fully comply with recommendations given in these guidelines, which is in line with what we recommend in this review. Calibration model and accuracy and precision were investigated in all included assays, which suggests that these parameters are regarded as pivotal for validation of TDM assays. Furthermore, the majority of included articles describe stability and selectivity experiments, stressing the importance of these procedures. Although validation procedures differ among these assays, all articles based their experiments and criteria upon accepted bioanalytical method validation guidelines. In order to harmonize method validation of TDM assays, we aim to provide guidance for future assay validation of TDM methods. Differences with regard to FDA and EMA guidelines are proposed for the calibration model, LLOQ, selectivity, dilution factor, matrix effect and recovery. All proposed adjustments are made considering the importance of high-throughput assays and to simplify validation and implementation of such assays, keeping confidence in the fit-for-use purpose of the bioanalytical method. At least four calibration standards instead of six to eight will be sufficient for TDM methods, as a short calibration range is recommended. Furthermore, a S/N ratio of at least 10 for the LLOQ will increase robustness of the assay regarding large between-assay variability of the MS signal. Endogenous selectivity experiments are of high importance during method validation of TDM methods, in which it is recommended to include blanks and LLOQ samples in six different batches of plasma. Interference of co-medications, degradation products or other xenobiotics only needs to be examined for structural analogous or when there is reason to believe that an interference may occur. Determining the dilution factor is generally not necessary for TDM assays, as concentrations above ULOQ indicate that the target was attained. Matrix effect and recovery experiments are of no additional value in TDM methods, if stable isotopically labeled internal standards are used co-eluting with the analytes of interest. Therefore, we propose to not include matrix effect and recovery in validation procedures. Accuracy and precision, carryover and stability experiments should be assessed according to FDA and EMA guidelines at three levels (Table 2). Analysis of study samples should be focussed on rapid turnaround. This can be achieved by analyzing only four calibration standards and at least three QC mids at TDM target level.

13 | CONCLUSIONS

A wide diversity of assays, for the purpose of TDM in oncology, have been developed and validated. This review presents an overview of publications in which LC–MS/MS assays have been validated for application in TDM. The focus of TDM assays is on developing and validating routine assays in which target attainment is strived for, rather than generating data for PK–TK studies. This is a different type of analytical procedure and, therefore, recommendations from FDA and EMA guidelines on bioanalytical method validation are comprehensive for LC–MS/MS assays specifically designed for TDM purposes. In addition to evaluating current practice, we recommend a minimal validation protocol for TDM assays while preserving a bioanalytical validation approach resulting in reliable bioanalytical results.

14 | FUTURE PERSPECTIVES

The first validation paper of an LC–MS/MS assay specifically designed for TDM in oncology was published in 2003. Since then, many TDM assays have been developed and validated, also in other fields (Figure 1). With individualized drug dosing gaining popularity, implementation of TDM will further increase. Moreover, the class of oral anticancer therapies is rapidly growing and these drugs have a high interpatient variable bioavailability and narrow therapeutic window. The TDM of anticancer drugs is becoming an important tool in treatment of patients with cancer, especially with increased use of oral anticancer drugs with highly variable bioavailability (B. Gao et al., 2012; Herbrink et al., 2018; Lankheet et al., 2014; Paci et al., 2014; Widmer et al., 2014; Yu et al., 2014). TDM has been shown to be a valuable intervention to optimize dosing of some anticancer drugs (Groenland et al., 2019; Paci et al., 2014; Verheijen et al., 2017; Widmer et al., 2014; Yu et al., 2014); however, prospective research is needed to further confirm these TDM targets. With the growing class of oral anticancer therapies, there is an increasing demand for TDM for which new assays have to be developed and validated. Shortened validation protocols could help to provide in this demand, while still offering sufficient confidence in the fit-for-purpose of the bioanalytical method. Simplifying the validation of TDM methods will shorten the time needed for validation and will increase the clinical implementation of such assays. In this review we focussed on TDM assays in oncology; however, recommendations can be applied to other fields and we advise bioanalytical laboratories to consider integrating our recommendations into standard validation of TDM assays.

ORCID

Merel van Nuland https://orcid.org/0000-0003-1763-7334

REFERENCES

Adaway, J. E., & Keevil, B. G. (2012). Therapeutic drug monitoring and LC–MS/MS. *Journal of Chromatography B, Analytical Technologies in the Biomedical and Life Sciences*, *883–884*, 33–49. https://doi.org/10.1016/j.jchromb.2011.09.041

Annesley, T. M. (2003). Ion suppression in mass spectrometry. *Clinical Chemistry*, *49*, 1041–1044. https://doi.org/10.1373/49.7.1041

Arellano, C., Gandia, P., Lafont, T., Jongejan, R., & Chatelut, E. (2012). Determination of unbound fraction of imatinib and *N*-desmethyl imatinib, validation of an UPLC–MS/MS assay and ultrafiltration method. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, *907*, 94–100. https://doi.org/10.1016/j.jchromb.2012.09.007

Bergeron, A., & Garofolo, F. (2013). Importance of matrix effects in LC–MS/MS bioanalysis. *Bioanalysis*, *5*, 2331–2332. https://doi.org/10.4155/bio.13.237

Bobin-Dubigeon, C., Heurgue-Berlot, A., Bouche, O., Amiand, M.-B., Le Guellec, C., & Bard, J.-M. (2011). A new rapid and sensitive LC–MS assay for the determination of sorafenib in plasma: Application to a patient undergoing hemodialysis. *Therapeutic Drug Monitoring*, *33*, 705–710. https://doi.org/10.1097/FTD.0b013e3182357891

Bonfiglio, R., King, R. C., Olah, T. V., & Merkle, K. (1999). The effects of sample preparation methods on the variability of the electrospray ionization response for model drug compounds. *Rapid Communications in Mass Spectrometry*, *13*, 1175–1185. https://doi.org/10.1002/(SICI)1097-0231(19990630)13:12<1175::AID-RCM639>3.0.CO;2-0

Capron, A., Destree, J., Maiter, D., & Wallemacq, P. (2014). Validation of a rapid liquid chromatography–tandem mass spectrometric assay for the determination of octreotide plasma concentrations. *Clinical Biochemistry*, *47*, 139–141. https://doi.org/10.1016/j.clinbiochem.2013.07.009

Cardoso, E., Mercier, T., Wagner, A. D., Homicsko, K., Michielin, O., Ellefsen-Lavoie, K., … Decosterd, L. (2018). Quantification of the next-generation oral anti-tumor drugs dabrafenib, trametinib, vemurafenib, cobimetinib, pazopanib, regorafenib and two metabolites in human plasma by liquid chromatography–tandem mass spectrometry. *Journal of Chromatography B, Analytical Technologies in the Biomedical and Life Sciences*, *1083*, 124–136. https://doi.org/10.1016/j.jchromb.2018.02.008

Clinical and Laboratory Standards Institute (2003). Evaluation of the linearity of quantitative measurement procedures: A statistical approach; Approved Guideline. EP06-A2. https://clsi.org/media/1437/ep06a_sample.pdf

Clinical and Laboratory Standards Institute (2006). Preliminary evaluation of quantitative clinical laboratory measurement procedures. EP10-A3. https://clsi.org/media/1434/ep10a3amd_sample.pdf

Couchman, L., Birch, M., Ireland, R., Corrigan, A., Wickramasinghe, S., Josephs, D., … Flanagan, R. J. (2012). An automated method for the measurement of a range of tyrosine kinase inhibitors in human plasma or serum using turbulent flow liquid chromatography–tandem mass spectrometry. *Analytical and Bioanalytical Chemistry*, *403*, 1685–1695. https://doi.org/10.1007/s00216-012-5970-2

Danso, D., Jannetto, P. J., Enger, R., & Langman, L. J. (2015). High-throughput validated method for the quantitation of busulfan in plasma using ultrafast SPE–MS/MS. *Therapeutic Drug Monitoring*, *37*, 319–324. https://doi.org/10.1097/FTD.0000000000000159

Dasgupta, A. (2016). Immunoassays and issues with interference in therapeutic drug monitoring. In *Clinical challenges in therapeutic drug monitoring* (pp. 17–44). Amsterdam: Elsevier.

Dennison, J. B., Renbarger, J. L., Walterhouse, D. O., Jones, D. R., & Hall, S. D. (2008). Quantification of vincristine and its major metabolite in human plasma by high-performance liquid chromatography/tandem mass spectrometry. *Therapeutic Drug Monitoring*, *30*, 357–364. https://doi.org/10.1097/FTD.0b013e31816b92c9

van Dyk, M., Miners, J. O., Kichenadasse, G., McKinnon, R. A., & Rowland, A. (2016). A novel approach for the simultaneous quantification of 18 small molecule kinase inhibitors in human plasma: A platform for optimised KI dosing. *Journal of Chromatography B, Analytical Technologies in the Biomedical and Life Sciences*, *1033–1034*, 17–26. https://doi.org/10.1016/j.jchromb.2016.07.046

Escudero-Ortiz, V., Perez-Ruixo, J. J., & Valenzuela, B. (2013). Development and validation of a high-performance liquid chromatography ultraviolet method for lapatinib quantification in human plasma. *Therapeutic Drug Monitoring*, *35*, 796–802. https://doi.org/10.1097/FTD.0b013e3182959080

European Medicines Agency. (2011). Guideline on bioanalytical method validation. Committe for Medicinal Produts for Human Use and European Medicines Agency. https://doi.org/EMEA/CHMP/EWP/192217/2009.

French, D., Sujishi, K. K., Long-Boyle, J. R., & Ritchie, J. C. (2014). Development and validation of a liquid chromatography–tandem mass spectrometry assay to quantify plasma busulfan. *Therapeutic Drug Monitoring*, *36*, 169–174. https://doi.org/10.1097/01.ftd.0000443060.22620.cd

Gao, B., Yeap, S., Clements, A., Balakrishnar, B., Wong, M., & Gurney, H. (2012). Evidence for therapeutic drug monitoring of targeted anticancer therapies. *Journal of Clinical Oncology*, *30*, 4017–4025. https://doi.org/10.1200/JCO.2012.43.5362

Gao, S., Zhou, J., Zhang, F., Miao, H., Yun, Y., Feng, J., … Chen, W. (2014). Rapid and sensitive liquid chromatography coupled with electrospray ionization tandem mass spectrometry method for the analysis of paclitaxel, docetaxel, vinblastine, and vinorelbine in human plasma. *Therapeutic Drug Monitoring*, *36*, 394–400. https://doi.org/10.1097/FTD.0000000000000010

Gibbons, J. A., Ouatas, T., Krauwinkel, W., Ohtsu, Y., van der Walt, J.-S., Beddo, V., … Mordenti, J. (2015). Clinical pharmacokinetic studies of enzalutamide. *Clinical Pharmacokinetics*, *54*, 1043–1055. https://doi.org/10.1007/s40262-015-0271-5

Groenland, S. L., van Nuland, M., Verheijen, R. B., Schellens, J. H. M., Beijnen, J. H., Huitema, A. D. R., & Steeghs, N. (2019). Therapeutic drug monitoring of oral anti-hormonal drugs in oncology. *Clinical Pharmacokinetics*, *58*, 299–308. https://doi.org/10.1007/s40262-018-0683-0

Gu, H., Liu, G., Wang, J., Aubry, A.-F., & Arnold, M. E. (2014). Selecting the correct weighting factors for linear and quadratic calibration curves with least-squares regression algorithm in bioanalytical LC–MS/MS assays and impacts of using incorrect weighting factors on curve stability, data quality, and assay perfo. *Analytical Chemistry*, *86*, 8959–8966. https://doi.org/10.1021/ac5018265

Heideloff, C., Payto, D., & Wang, S. (2013). Comparison of a stable isotope-labeled and an analog internal standard for the quantification of everolimus by a liquid chromatography–tandem mass spectrometry method. *Therapeutic Drug Monitoring*, *35*, 246–250. https://doi.org/10.1097/FTD.0b013e318283403e

Herbrink, M., de Vries, N., Rosing, H., Huitema, A. D. R., Nuijen, B., Schellens, J. H. M., & Beijnen, J. H. (2016). Quantification of 11 therapeutic kinase inhibitors in human plasma for therapeutic drug monitoring using liquid chromatography coupled with tandem mass spectrometry. *Therapeutic Drug Monitoring*, *38*, 649–656. https://doi.org/10.1097/FTD.0000000000000349

Herbrink, M., de Vries, N., Rosing, H., Huitema, A. D. R., Nuijen, B., Schellens, J. H. M., & Beijnen, J. H. (2018). Development and validation of a liquid chromatography–tandem mass spectrometry analytical method for the therapeutic drug monitoring of eight novel anticancer drugs. *Biomedical Chromatography*, *32*. https://doi.org/10.1002/bmc.4147

Hughes, N. C., Wong, E. Y. K., Fan, J., & Bajaj, N. (2007). Determination of carryover and contamination for mass spectrometry-based

chromatographic assays. *The AAPS Journal*, *9*, E353–E360. https://doi.org/10.1208/aapsj0903042

Huynh, H. H., Pressiat, C., Sauvageon, H., Madelaine, I., Maslanka, P., Lebbe, C., … Mourah, S. (2017). Development and validation of a simultaneous quantification method of 14 tyrosine kinase inhibitors in human plasma using LC–MS/MS. *Therapeutic Drug Monitoring*, *39*, 43–54. https://doi.org/10.1097/FTD.0000000000000357

Jager, N. G. L., Rosing, H., Linn, S. C., Schellens, J. H. M., & Beijnen, J. H. (2012). Importance of highly selective LC–MS/MS analysis for the accurate quantification of tamoxifen and its metabolites: focus on endoxifen and 4-hydroxytamoxifen. *Breast Cancer Research and Treatment*, *133*, 793–798. https://doi.org/10.1007/s10549-012-2000-1

de Jonge, M. E., Huitema, A. D. R., Schellens, J. H. M., Rodenhuis, S., & Beijnen, J. H. (2005). Individualised cancer chemotherapy: strategies and performance of prospective studies on therapeutic drug monitoring with dose adaptation: a review. *Clinical Pharmacokinetics*, *44*, 147–173. https://doi.org/10.2165/00003088-200544020-00002

King, R., Bonfiglio, R., Fernandez-Metzler, C., Miller-Stein, C., & Olah, T. (2000). Mechanistic investigation of ionization suppression in electrospray ionization. *Journal of the American Society for Mass Spectrometry*, *11*, 942–950. https://doi.org/10.1016/S1044-0305(00)00163-X

Kirchherr, H., Shipkova, M., & von Ahsen, N. (2013). Improved method for therapeutic drug monitoring of 6-thioguanine nucleotides and 6-methylmercaptopurine in whole-blood by LC/MSMS using isotope-labeled internal standards. *Therapeutic Drug Monitoring*, *35*, 313–321. https://doi.org/10.1097/FTD.0b013e318283ed5d

Kosovec, J. E., Egorin, M. J., Gjurich, S., & Beumer, J. H. (2008). Quantitation of 5-fluorouracil (5-FU) in human plasma by liquid chromatography/electrospray ionization tandem mass spectrometry. *Rapid Communications in Mass Spectrometry*, *22*, 224–230. https://doi.org/10.1002/rcm.3362

de Krou, S., Rosing, H., Nuijen, B., Schellens, J. H. M., & Beijnen, J. H. (2017). Fast and adequate liquid chromatography–tandem mass spectrometric determination of *Z*-endoxifen serum levels for therapeutic drug monitoring. *Therapeutic Drug Monitoring*, *39*, 132–137. https://doi.org/10.1097/FTD.0000000000000372

Lankheet, N. A. G., Hillebrand, M. J. X., Rosing, H., Schellens, J. H. M., Beijnen, J. H., & Huitema, A. D. R. (2013). Method development and validation for the quantification of dasatinib, erlotinib, gefitinib, imatinib, lapatinib, nilotinib, sorafenib and sunitinib in human plasma by liquid chromatography coupled with tandem mass spectrometry. *Biomedical Chromatography*, *27*, 466–476. https://doi.org/10.1002/bmc.2814

Lankheet, N. A. G., Knapen, L. M., Schellens, J. H. M., Beijnen, J. H., Steeghs, N., & Huitema, A. D. R. (2014). Plasma concentrations of tyrosine kinase inhibitors imatinib, erlotinib, and sunitinib in routine clinical outpatient cancer care. *Therapeutic Drug Monitoring*, *36*, 326–334. https://doi.org/10.1097/FTD.0000000000000004

Lankheet, N. A. G., Steeghs, N., Rosing, H., Schellens, J. H. M., Beijnen, J. H., & Huitema, A. D. R. (2013). Quantification of sunitinib and *N*-desethyl sunitinib in human EDTA plasma by liquid chromatography coupled with electrospray ionization tandem mass spectrometry: Validation and application in routine therapeutic drug monitoring. *Therapeutic Drug Monitoring*, *35*, 168–176. https://doi.org/10.1097/FTD.0b013e31827efd9e

Liu, H., Lam, L., & Dasgupta, P. K. (2011). Expanding the linear dynamic range for multiple reaction monitoring in quantitative liquid chromatography–tandem mass spectrometry utilizing natural isotopologue transitions. *Talanta*, *87*, 307–310. https://doi.org/10.1016/j.talanta.2011.09.063

Lynch, K. L. (2016). CLSI C62-A: A new standard for clinical mass spectrometry. *Clinical Chemistry*, *62*, 24–29. https://doi.org/10.1373/clinchem.2015.238626

Marchi, I., Viette, V., Badoud, F., Fathi, M., Saugy, M., Rudaz, S., & Veuthey, J.-L. (2010). Characterization and classification of matrix effects in biological samples analyses. *Journal of Chromatography. A*, *1217*, 4071–4078. https://doi.org/10.1016/j.chroma.2009.08.061

Matuszewski, B. K., Constanzer, M. L., & Chavez-Eng, C. M. (2003). Strategies for the assessment of matrix effect in quantitative bioanalytical methods based on HPLC–MS/MS. *Analytical Chemistry*, *75*, 3019–3030. https://doi.org/10.1021/ac020361s

Meesters, R. J. W., Cornelissen, R., van Klaveren, R. J., de Jonge, R., den Boer, E., Lindemans, J., & Luider, T. M. (2010). A new ultrafast and high-throughput mass spectrometric approach for the therapeutic drug monitoring of the multi-targeted anti-folate pemetrexed in plasma from lung cancer patients. *Analytical and Bioanalytical Chemistry*, *398*, 2943–2948. https://doi.org/10.1007/s00216-010-4192-8

Mei, S., Shi, X., Du, Y., Cui, Y., Zeng, C., Ren, X., … Lin, S. (2018). Simultaneous determination of plasma methotrexate and 7-hydroxy methotrexate by UHPLC–MS/MS in patients receiving high-dose methotrexate therapy. *Journal of Pharmaceutical and Biomedical Analysis*, *158*, 300–306. https://doi.org/10.1016/j.jpba.2018.06.011

Miura, M., & Takahashi, N. (2016). Routine therapeutic drug monitoring of tyrosine kinase inhibitors by HPLC–UV or LC–MS/MS methods. *Drug Metabolism and Pharmacokinetics*, *31*, 12–20. https://doi.org/10.1016/j.dmpk.2015.09.002

Morgan, P. E., Brown, N. W., & Tredger, J. M. (2014). A direct method for the measurement of everolimus and sirolimus in whole blood by LC–MS/MS using an isotopic everolimus internal standard. *Therapeutic Drug Monitoring*, *36*, 358–365. https://doi.org/10.1097/FTD.0000000000000006

Nijenhuis, C. M., Haverkate, H., Rosing, H., Schellens, J. H. M., & Beijnen, J. H. (2016). Simultaneous quantification of dabrafenib and trametinib in human plasma using high-performance liquid chromatography–tandem mass spectrometry. *Journal of Pharmaceutical and Biomedical Analysis*, *125*, 270–279. https://doi.org/10.1016/j.jpba.2016.03.049

van Nuland, M., Hillebrand, M. J., Rosing, H., Schellens, J. H. M., & Beijnen, J. H. (2017). Development and validation of an LC–MS/MS method for the simultaneous quantification of abiraterone, enzalutamide, and their major metabolites in human plasma. *Therapeutic Drug Monitoring*, *39*, 243–251. https://doi.org/10.1097/FTD.0000000000000387

van Nuland, M., Venekamp, N., de Vries, N., de Jong, K. A. M., Rosing, H., & Beijnen, J. H. (2019). Development and validation of an UPLC–MS/MS method for the therapeutic drug monitoring of oral anti-hormonal drugs in oncology. *Journal of Chromatography B*, *1106–1107*, 26–34. https://doi.org/10.1016/j.jchromb.2019.01.001

Paci, A., Veal, G., Bardin, C., Leveque, D., Widmer, N., Beijnen, J., … Chatelut, E. (2014). Review of therapeutic drug monitoring of anticancer drugs part 1--cytotoxics. *European Journal of Cancer*, *50*, 2010–2019. https://doi.org/10.1016/j.ejca.2014.04.014

Peters, F. T., & Maurer, H. H. (2007). Systematic comparison of bias and precision data obtained with multiple-point and one-point calibration in six validated multianalyte assays for quantification of drugs in human plasma. *Analytical Chemistry*, *79*, 4967–4976. https://doi.org/10.1021/ac070054s

Pressiat, C., Huynh, H.-H., Plé, A., Sauvageon, H., Madelaine, I., Chougnet, C., … Goldwirt, L. (2018). Development and validation of a simultaneous quantification method of ruxolitinib, vismodegib, olaparib, and pazopanib in human plasma using liquid chromatography coupled with tandem mass spectrometry. *Therapeutic Drug Monitoring*, *40*, 337–343. https://doi.org/10.1097/FTD.0000000000000497

Rezende, V. M., Rivellis, A. J., Gomes, M. M., Dorr, F. A., Novaes, M. M. Y., Nardinelli, L., … Bendit, I. (2013). Determination of serum levels of imatinib mesylate in patients with chronic myeloid leukemia: Validation and application of a new analytical method to monitor treatment compliance. *Revista Brasileira de Hematologia e Hemoterapia*, *35*, 103–108. https://doi.org/10.5581/1516-8484.20130030

Rood, J. J. M., van Bussel, M. T. J., Schellens, J. H. M., Beijnen, J. H., & Sparidans, R. W. (2016). Liquid chromatography–tandem mass spectrometric assay for the T790M mutant EGFR inhibitor osimertinib (AZD9291) in human plasma. *Journal of Chromatography B, Analytical Technologies in the Biomedical and Life Sciences*, *1031*, 80–85. https://doi.org/10.1016/j.jchromb.2016.07.037

Rosing, H., Man, W. Y., Doyle, E., Bult, A., & Beijnen, J. H. (2000). Bioanalytical liquid chromatographic method validation. A review of current practices and procedures. *Journal of Liquid Chromatography & Related Technologies*, *23*, 329–354. https://doi.org/10.1081/JLC-100101455

Rousset, M., Titier, K., Bouchet, S., Dutriaux, C., Pham-Ledard, A., Prey, S., … Molimard, M. (2017). An UPLC–MS/MS method for the quantification of BRAF inhibitors (vemurafenib, dabrafenib) and MEK inhibitors (cobimetinib, trametinib, binimetinib) in human plasma. Application to treated melanoma patients. *Clinica Chimica Acta; International Journal of Clinical Chemistry*, *470*, 8–13. https://doi.org/10.1016/j.cca.2017.04.009

Schellens, J. H. M., McLeod, H. L., & Newell, D. R. (2005). *Cancer clinical pharmacology*. Oxford: Oxford University Press.

Schofield, R. C., Ramanathan, L. V., Murata, K., Grace, M., Fleisher, M., Pessin, M. S., & Carlow, D. C. (2015). Development and validation of a turbulent flow chromatography and tandem mass spectrometry method for the quantitation of methotrexate and its metabolites 7-hydroxy methotrexate and DAMPA in serum. *Journal of Chromatography B, Analytical Technologies in the Biomedical and Life Sciences*, *1002*, 169–175. https://doi.org/10.1016/j.jchromb.2015.08.025

Shao, R., Yu, L., Lou, H., Ruan, Z., Jiang, B., & Chen, J. (2016). Development and validation of a rapid LC–MS/MS method to quantify letrozole in human plasma and its application to therapeutic drug monitoring. *Biomedical Chromatography*, *30*, 632–637. https://doi.org/10.1002/bmc.3607

Stokvis, E., Rosing, H., & Beijnen, J. H. (2005). Stable isotopically labeled internal standards in quantitative bioanalysis using liquid chromatography/mass spectrometry: Necessity or not? *Rapid Communications in Mass Spectrometry*, *19*, 401–407. https://doi.org/10.1002/rcm.1790

Taylor, P. J. (2005). Matrix effects: the Achilles heel of quantitative high-performance liquid chromatography–electrospray–tandem mass spectrometry. *Clinical Biochemistry*, *38*, 328–334. https://doi.org/10.1016/j.clinbiochem.2004.11.007

Taylor, P. J., Franklin, M. E., Graham, K. S., & Pillans, P. I. (2007). A HPLC–mass spectrometric method suitable for the therapeutic drug monitoring of everolimus. *Journal of Chromatography B, Analytical Technologies in the Biomedical and Life Sciences*, *848*, 208–214. https://doi.org/10.1016/j.jchromb.2006.10.029

US Food and Drug Administration (FDA). (2018). FDA guidance for industry: Bioanalytical method validation. https://doi.org/http://www.labcompliance.de/documents/FDA/FDA-Others/Laboratory/f-507-bioanalytical-4252fnl.pdf.

Verheijen, R. B., Thijssen, B., Rosing, H., Schellens, J. H. M., Nan, L., Venekamp, N., … Huitema, A. D. R. (2018). Fast and straightforward method for the quantification of pazopanib in human plasma using LC–MS/MS. *Therapeutic Drug Monitoring*, *40*, 230–236. https://doi.org/10.1097/FTD.0000000000000479

Verheijen, R. B., Yu, H., Schellens, J. H. M., Beijnen, J. H., Steeghs, N., & Huitema, A. D. R. (2017). Practical recommendations for therapeutic drug monitoring of kinase inhibitors in oncology. *Clinical Pharmacology and Therapeutics*, *102*, 765–776. https://doi.org/10.1002/cpt.787

Viswanathan, C. T., Bansal, S., Booth, B., DeStefano, A. J., Rose, M. J., Sailstad, J., … Weiner, R. (2007). Workshop/conference report—Quantitative bioanalytical methods validation and implementation: Best practices for chromatographic and ligand binding assays. *The AAPS Journal*, *9*, E30–E42. https://doi.org/10.1208/aapsj0901004

Wang, S., Cyronak, M., & Yang, E. (2007). Does a stable isotopically labeled internal standard always correct analyte response? A matrix effect study on a LC/MS/MS method for the determination of carvedilol enantiomers in human plasma. *Journal of Pharmaceutical and Biomedical Analysis*, *43*, 701–707. https://doi.org/10.1016/j.jpba.2006.08.010

Widmer, N., Bardin, C., Chatelut, E., Paci, A., Beijnen, J., Leveque, D., … Astier, A. (2014). Review of therapeutic drug monitoring of anticancer drugs part two—Targeted therapies. *European Journal of Cancer*, *50*, 2020–2036. https://doi.org/10.1016/j.ejca.2014.04.015

Wojnicz, A., Colom-Fernandez, B., Steegmann, J. L., Munoz-Calleja, C., Abad-Santos, F., & Ruiz-Nuno, A. (2017). Simultaneous determination of imatinib, dasatinib, and nilotinib by liquid chromatography–tandem mass spectrometry and its application to therapeutic drug monitoring. *Therapeutic Drug Monitoring*, *39*, 252–262. https://doi.org/10.1097/FTD.0000000000000406

Wu, D., Wang, Y., Sun, Y., Ouyang, N., & Qian, J. (2015). A simple, rapid and reliable liquid chromatography–mass spectrometry method for determination of methotrexate in human plasma and its application to therapeutic drug monitoring. *Biomedical Chromatography*, *29*, 1197–1202. https://doi.org/10.1002/bmc.3408

Yu, H., Steeghs, N., Nijenhuis, C. M., Schellens, J. H. M., Beijnen, J. H., & Huitema, A. D. R. (2014). Practical guidelines for therapeutic drug monitoring of anticancer tyrosine kinase inhibitors: focus on the pharmacokinetic targets. *Clinical Pharmacokinetics*, *53*, 305–325. https://doi.org/10.1007/s40262-014-0137-2

Zhang, Y., & Zhang, R. (2018). Recent advances in analytical methods for the therapeutic drug monitoring of immunosuppressive drugs. *Drug Testing and Analysis*, *10*, 81–94. https://doi.org/10.1002/dta.2290

de Zwart, M., Lausecker, B., Globig, S., Neddermann, D., Le Bras, B., Guenzi, A., … Timmerman, P. (2016). Co-medication and interference testing in bioanalysis: A European Bioanalysis Forum recommendation. *Bioanalysis*, *8*, 2065–2070. https://doi.org/10.4155/bio-2016-0179

How to cite this article: van Nuland M, Rosing H, Schellens JHM, Beijnen JH. Bioanalytical LC–MS/MS validation of therapeutic drug monitoring assays in oncology. *Biomedical Chromatography*. 2020;34:e4623. https://doi.org/10.1002/bmc.4623