QATAR UNIVERSITY

COLLEGE OF PHARMACY

STRENGTHENING THE QUALITY OF CLINICAL PHARMACOKINETIC STUDIES: DEVELOPMENT AND VALIDATION OF A CRITICAL APPRAISAL TOOL FOR CLINICAL PHARMACOKINETIC RESEARCH

BY

ALAA BAHAA ELDEEN FOUAD AHMED SOLIMAN

A Thesis Submitted to

the College of Pharmacy

in Partial Fulfillment of the Requirements for the Degree of

Masters of Science in Pharmacy

June 2020

COMMITTEE PAGE

The members of the Committee approve the Thesis of
Alaa Bahaa Eldeen Fouad Ahmed Soliman defended on 09/04/2020.

Dr. Ousama Rachid
Thesis Supervisor

Dr. Shane Pawluk
Thesis Co-Supervisor

Dr. Kyle Wilby
Thesis Co-Supervisor

Dr. Tony Kiang
External Committee Member

Approved:

Mohammad Diab, Dean, College of Pharmacy

ABSTRACT

SOLIMAN, ALAA, BE., Masters : June : [2020:], Clinical Pharmacy and Practice

Title: Strengthening the Quality of Clinical Pharmacokinetic Studies: Development and Validation of a Critical Appraisal Tool for Clinical Pharmacokinetic Research

Supervisor of Thesis: Ousama M. Rachid.

Background: Critical appraisal process is central to the practice of evidence-based medicine that aids in assessing the quality of the published scientific knowledge. To our knowledge, there is no critical appraisal tool for clinical pharmacokinetics studies. Therefore, this study aimed to develop the first valid and reliable clinical pharmacokinetics critical appraisal tool.

Methodology: A systematic review was conducted through Embase and Pubmed to identify quality markers related to clinical pharmacokinetic studies. Questions that helped in appraising pharmacokinetic studies were formulated from these quality markers. Twenty-five clinical pharmacokinetics experts were involved in a modified Delphi process to achieve their consensus regarding the formulated questions. Percentage of agreement between panelists, median and interquartile range were calculated for each question to determine whether they achieved expert's consensus. Content and face validity of the developed critical appraisal tool were assessed twice through modified Delphi process and by a psychometric expert. Four raters were selected to apply the developed tool on 30 clinical pharmacokinetic articles to evaluate the inter-rater and intra-rater reliability by calculating Kappa values for each of them.

Results: Quality markers of clinical pharmacokinetic studies were identified out of fifteen articles included in the systematic review which encompassed 19 subcategories,

most of them were related to the methods and results subcategories. The modified Delphi process consisted of 3 rounds. Sixty-four quality-related questions were formulated out of these quality markers to appraise clinical pharmacokinetics studies but 42 were sent to round 1. Of 42 items,12 items reached ≥80 % of agreement, median ≥ 4, and interquartile range ≤ 1 consensus from experts. In round 2, of 25 questions, 6 items met ≥80% of agreement, a median ≥ 4, and interquartile range < 1 from experts. In round 3, of 3 questions, 3 items achieved ≥80% of agreement, a median ≥ 4, and interquartile range < 1 from experts. Twenty-one questions achieved expert consensus to be included in the final critical appraisal tool. This tool proved to be valid and reliable to help end-users in appraising retrospective and prospective clinical pharmacokinetics, bioequivalence, and population pharmacokinetics studies.

Conclusion: A clinical pharmacokinetics critical appraisal tool consisting of twenty-one questions was developed.

DEDICATION

This work is dedicated to all my family members for their endless love, support and encouragement to overcome all challenges

ACKNOWLEDGEMENTS

Alhamdulillah (الحمدلله) for his blessing, and guidance, for providing me the patience and strength through my life, without them I could not complete this work successfully after facing several challenges.

It was three years full of many challenges, troubles, and triumphs. Masters is a life challenging experience. I could not get over without Allah's guidance and people who supported me scientifically and emotionally since the start of this journey.

I would like to thank people who supported me to achieve my dream.

I would like to extend my gratitude to my mom and dad, who keep supporting and encouraging me during hard times. To my two sisters, who keep listening to me. My husband, who believed in me and stood by my side.

To Professor Mohamed Izham at Qatar University, I extend a genuine appreciation and an endless thank you for keeping your door open to listen to me and supporting me to finalize this project.

To my main supervisors, thank you Dr. Ousama Rachid for accepting the challenge to be my main supervisor in the last year to help me to finalize my thesis and for presenting different opportunities to me to present my work. To Dr. Shane Pawluk, thank you so much for all of your efforts and for taking the challenge to accept to lead the project instead of Dr. Kyle Wilby to help me to make my dream coming true to do masters in clinical pharmacokinetic. Although you left me before the completion of my thesis, I admit that you helped me and you put a lot of effort during the three phases of this project and you made the road very smooth for me during the first two years.

To my co-supervisor Dr. Kyle Wilby a joint acknowledgment for accepting to give me the project although you were not the main supervisor and for your valued input.

Finally, I must express my profound gratitude to the college of pharmacy at Qatar University for giving me the opportunity to present my work at conferences and providing me with student grant QUST-2-CPH-2018-1.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1. Evidence-based medicine

The father of the evidence-based medicine (EbM), David Sackett and colleagues, defined EbM as "The conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients. The practice of EbM means integrating individual clinical expertise with the best available external clinical evidence from systematic research" (1). Practicing EbM is a key contributor to helping Qatar achieve its national health vision by 2022. Specifically, when clinicians practice EbM through tailoring medical decisions for each patient by integrating their clinical skills and expertise with the best available up-to-date research (2). Consequently, this will improve the provided healthcare to patients in Qatar.

1.2. Critical appraisal

The critical appraisal process is central to the practice of EbM. There are two definitions for critical appraisal process. Firstly, it is defined as "the process of systematically assessing the relevance, quality and the trustworthiness of a published article in terms of its context" (3). Critical appraisal process helps in assessing the validity, reliability, and quality of the published scientific knowledge as it is also defined as "the application of rules of evidence to a published study to assess the validity of the data, completeness of reporting, methods and procedures, conclusions, compliance with ethical standards, etc. The rules of evidence vary with circumstances" (4). Therefore, researchers designed specific critical appraisal tools for each study design in addtion to the generic appraisal tools. Accordingly, the critical appraisal process aids in assessing the quality of the study results and how they are interpreted to decide about the application of published scientific information

in practice settings like policy-making agencies, clinics, and research clinical trials. As such there are three components: validity, importance, and applicability, which are part of critically appraising any published articles (Figure 1) (5). The elements of each previously mentioned component help in determining if the presented results are reliable or not, and to evaluate the feasibility and the value of the displayed results (5).

1.3. Critical appraisal tools

Critical appraisal tools are designed to enhance bias identification methods in the research project by criticizing the quality of the conducted project through analytically evaluating it (6). Therefore, selecting the most appropriate critical appraisal tool is highly essential for the application of evidence-based practice (7). Two types of critical appraisal tools are available: design-specific and generic critical appraisal tools. Design-specific tools contain items that criticize the methodological quality of the study design (8).

Generic critical appraisal tools help in appraising quantitative and qualitative studies (9). There are universal principles in the developed critical appraisal tools which guide the users to assess a published research paper systematically. However, evaluating the validity of the study varies based on the methodological criteria that usually differ based on the study design (6).

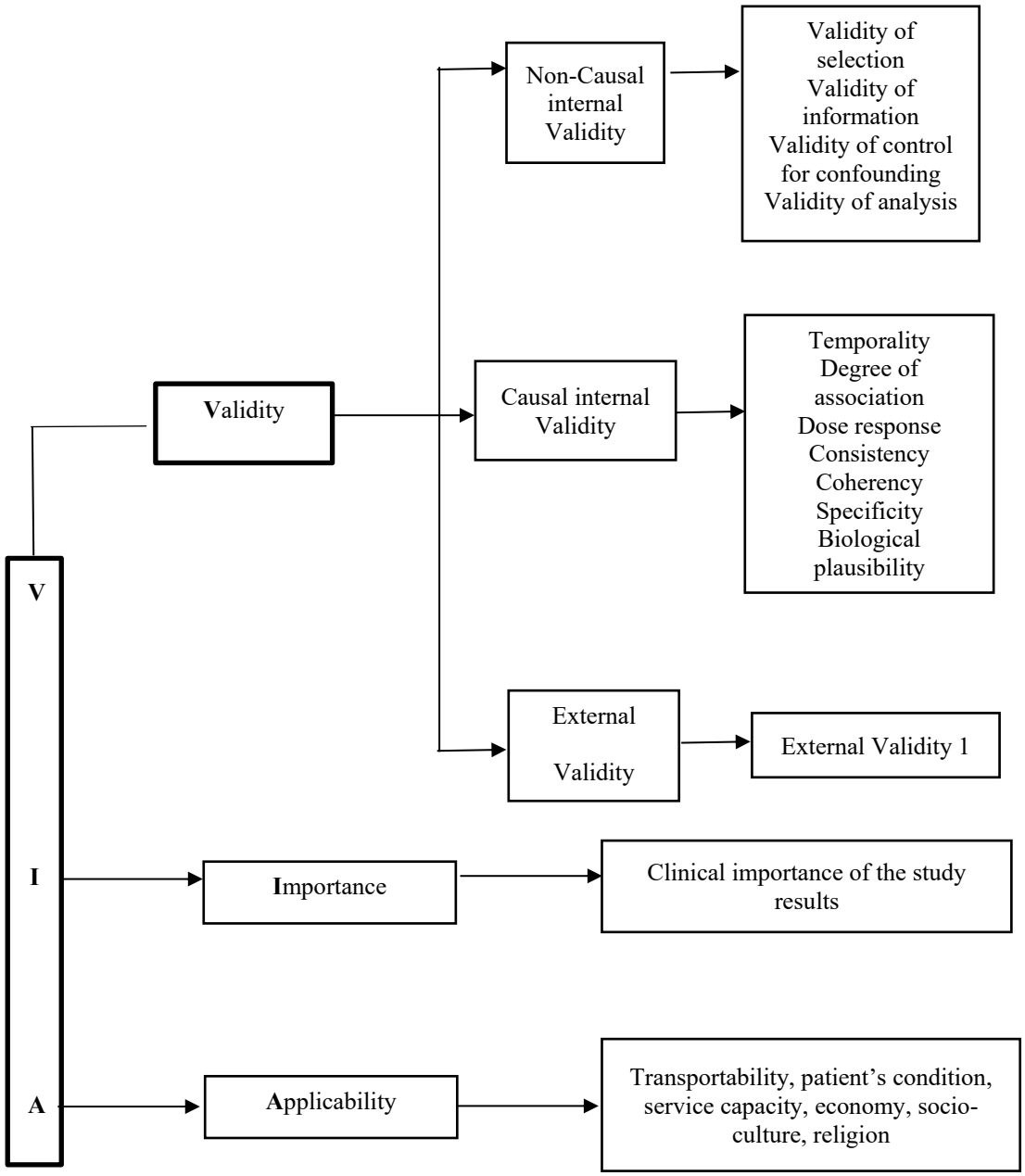

Figure 1. Criticial apprasial components (5)

1.4. Types of critical appraisal tools and reporting checklist

Several critical appraisal tools were developed to help end-users to systematically appraise different kinds of trials. There are two types of appraisal tools: design-specific critical appraisal tools and generic critical appraisal tools.

1.4.1. Critical appraisal tool for systematic reviews

There are numerous critical appraisal tools for appraising systematic reviews. These tools contain items that are specific to systematic reviews and items regarding assessing data analysis and external validity. Items that are related to data analysis are focusing on estimating the methods and evaluating the sensitivity of the results while summarizing them and determining whether the heterogeneity has existed. In these tools, the external validity of the study results is assessed by asking whether the researchers reported and interpreted the main results and if the results are generalizable as well as adherent to the study protocol.

For example, the AMSTAR assessment tool (10) contains 11 questions, all of which guide end-users to appraise systematic reviews. Additionally, the CASP checklists (11) are developed to guide users to evaluate the quality of the published systematic reviews of randomized controlled trials (RCTs). Consequently, the MOOSE (Meta-Analysis Of Observational Studies in Epidemiology) guidelines have been published as guidance to assist users in appraising meta-analyses of non-RCTs (12). After the development of these critical appraisal tools, a lot of researchers started to study the potential source of bias that may occur in systematic reviews like publication language and publication bias (9).

1.4.2. Randomized controlled trials critical appraisal tool

Included items in these appraisal tools evaluate two aspects: data analyses and blinding. Items related to data analysis are focused on assessing the quality of the performed statistical analysis, justification, the power of the calculated sample size and the analysis of the reported side effect of the used intervention (9). Items related to blinding are focused on determining whether clinicaians, participants and assessors were blinded to the used intervention (9).

1.4.3. Observational studies critical appraisal tool

Included items in these appraisal tools evaluate mainly data analyses. Items related to data analysis are focused on assessing the quality of the performed statistical analysis, justification of the calculated sample size and the performed statistical analysis (9).

1.4.4. Qualitative studies critical appraisal tool

Included items in these tools are focused on assessing the rationale behind conducting the study as well as the methods used to analyze the collected data and its external validity. However, items that are not related to the assessment of the qualitative paradigm are excluded from these tools like sample selection, randomization, blinding, intervention or bias (9).

1.4.5. Generic critical appraisal tools

These tools are developed to assess both experimental and observational studies. The items of these tools are focused mainly on evaluating the quality of the selected sample as well as data analyses, precisely the used statistical method and the power of the chosen sample size (9).

1.4.6. Critical appraisal tools for all study designs

These critical appraisal tools are developed to appraise all study designs (quantitative and qualitative). Items in these tools are focused mainly on evaluating the quality of

the performed data analyses, specifically statistical analysis, sample size calculations and identifications of confounders. Furthermore, these items are also focused on assessing external validity of the study results to determine their generalizability on the population (9). Table 1 provides references for different examples of critical appraisal tools based on the study design.

Table 1: Some of the available critical appraisal tools

Research design	References of the critical appraisal tool
Systematic reviews / meta-analyses	(12-35)
Experimental studies	(31, 33-41)
Observational studies	(34, 42)
Qualitative studies	(35, 43, 44)
All study designs	(45, 46)

1.5. Pharmacokinetics

Pharmacokinetics is the cornerstone of understanding a drug's biological fate (47). It is defined as the study of drug movement in the human body over time. The drug passes through four processes simulataneously inside the human body which are absorption, distribution, metabolism, and excretion (48). The absorption process is defined as the movement of the drug molecules from the site of administration to the site of measurement: blood or systemic circulation. The distribution process is defined as the transfer of drug molecules between systemic circulation and the rest of the body. The volume of distribution (V_D) is defined as the volume in which the drug is apparently distributed into. Additionally, the distribution rate constant (K_d) is one of

the important parameters needed to determine the required time to achieve the distribution equilibrium (49). The drug is eliminated through the biotransformation process as well as through the passage of molecules through urine, bile or other routes. The elimination process is expressed by the elimination rate, elimination rate constant (k) and clearance. Clearance is defined as the volume of fluid from which the drug is completely removed per unit of time (50). From k, the half-life ($t_{1/2}$) can be calculated and it is defined as the time required for plasma drug initial concentration to drop to half (50).These pharmacokinetic parameters are used to estimate the drug concentration remaining in the sampling compartment at any time point.

Other pharmacokinetic parameters that are estimated after a single-dose drug administration include the area under the curve (AUC) which describes and quantifies the concentration of the administered drug or its metabolites in the plasma over time, the maximum concentration (C_{max}), and time to reach C_{max} (T_{max}) (50, 51). After multiple dosing, trough (minimum) serum concentration is an important parameter that is commonly used in clinical settings. It is defined as the level of the drug before the administration of the next dose while the drug concentration at the peak of the absorption process is known as the peak (maximum) serum concentration.

1.5.1. Pharmacokinetics models

Mathematical models are used to describe the absorption, distribution, metabolism, and excretion processes. The previously mentioned processes are simplified by using mathematical principles to predict a drug's pharmacokinetic behavior. Mathematical principles are applied by using models that represent the human body to predict drug concentration in fluids and tissues over time (52).

1.5.1.1. Compartmental models

Compartmental models are considered to be the basic type of available

pharmacokinetics models, which are also known as deterministic models. These models are differentiated by a number of compartments like one-compartment, two-compartment, and multi-compartment models. Each compartment is used to describe a group of tissues or fluids that have similar pattern of drug distribution (52).

1.5.1.2. Physiological model

Researchers aim to consider as much data as possible about the absorption, distribution and elimination processes of the drug while using the physiologically-based pharmacokinetic (PBPK) model. Furthermore, this type of model is described as a mathematical model that assists researchers to collect information about the movement and disposition of the drug based on the blood flow to and from the organ in the body and the penetrated spaces in the organ. Both of the physiologic and anatomic information is used to determine the concentration and the distribution of the drug in the tissues by using the PBPK models. Moreover, several factors can be inserted in the PBPK model like drug-protein binding, tissue organ drug partition ratios, and intrinsic hepatic clearance **(53)**.

1.5.2. Clinical pharmacokinetics

Clinical pharmacokinetics is the application of pharmacokinetics principles on individual patients aiming to target safe and effective therapeutic drug management (52). Clinicians aim to improve a patient's response to the dosage regimen and to minimize the chance of the experienced side effects by designing a specific dosage regimen through applying clinical pharmacokinetics concepts (48). Clinicians are able to apply pharmacokinetics concepts in clinical settings through understanding and studying the correlation between the concentration of medications and their pharmacological responses (52).

1.5.2.1. Types of clinical pharmacokinetics studies

There are two types of clinical pharmacokinetics studies: standard pharmacokinetics studies and population pharmacokinetics studies.

1.5.2.1.1. Standard pharmacokinetics study

In standard pharmacokinetics studies, the pharmacokinetics profile of the investigated drug is studied through conducting either single-dose or repeated-dose research. The biological sample (blood, urine or fecal samples, if needed) are collected from participants on a fixed schedule to determine the pharmacokinetics profile of the investigated medication (54).

1.5.2.1.1.1. Single-dose study

Healthy volunteers or patients receive a single dose of the investigated medicine. Then the biological samples (blood, urine, or fecal samples, if needed) are collected from participants to determine the amount of the invistigated drug and its metabolites. The given dose of the investigated medicine is based on the recommendations of toxicity studies, toxicokinetic and nonclinical pharmacokinetics studies, and metabolic studies performed on human tissues (54). A small number of participants are needed in a single dose study, but the number should be appropriate to investigate the existence of inter-individual variability. In dose-esclation studies, clinicians start the process by giving patients the lowest dose and gradually increasing the dose, to include the estimated clinical dose and higher than the estimated dose while monitoring the side effect to study the relationship between the administered dose and pharmacokinetics parameters. Multiple samples are collected from the included participants at fixed time-points to study the relationship between the dosage of the administered medication and the drug concentration in the blood and its pharmacological effect. In the case of collecting urine samples, collection of samples should be continued until the drug and its metabolites are not detectable. The

bioavailability, linearity of pharmacokinetics, binding to plasma proteins, and the effect of meals on the administration of the investigated medication can be assessed in single dose studies. Furthermore, mass balance is evaluated in this type of study by examining the amount of the investigated drug and its metabolites (54).

1.5.2.1.1.2. Repeated dose study

In repeated dose studies, the medication of interest is administered multiple times based on a specific schedule followed in clinical practice. This process helps in measuring the exact concentration, accumulation of the medicine during steady-state, and assessing deviation in the pharmacokinetics parameters. The number of the included participants in the repeated dose study is based on single-dose study results. In the repeated dose study, frequent samples should be collected from participants after the first dose to aid in evaluating the subjects' pharmacokinetics profiles. During the administration of the medication, the samples should be collected corresponding to the trough or peak concentration. Once the steady-state is achieved an adequate number of samples should be obtained to assess the elimination rate, accumulation, and linearity. Collection of samples after the first, middle and final doses helps in understanding drug concentration/response relationships (54).

1.5.2.1.2. Population pharmacokinetics study

Population pharmacokinetics is the study of the variability of the investigated drug concentration among patients who receive specific doses of the investigated medication (55). There are different objectives for conducting population pharmacokinetics studies. One of the objective of the population pharmacokinetics is to assess pharmacokinetics profiles of the investigated drugs through using a repeated dose or single-dose study based on certain protocol that is applied under highly controlled conditions. Another objective of population pharamacokinetics studies is

the determination of the safety and efficacy of the investigated drug through evaluating pharmacokinetics profile by obtaining blood concentration from clinical pharmacokinetics studies. Additionally, population pharmacokinetics guide researchers in determining the influence of pathophysiologic factors (body weight, excretory, and metabolic functions), co-administrated medications, and patient demographics on the drug-concentration. Furthermore, these studies help in identifying the effect of these factors: pathophysiologic factors (body weight, excretory, and metabolic functions), co-administrated medications, and patient demographics on the therapeutic index of the investigated drug. Based on the results of these studies, dosage modification may be recommended in specific patient populations (54, 56).

In population pharmacokinetics studies, a large number of the population are included; therefore, a small number of samples are collected per participant. As a result, this method is considered to be highly suitable for studying the elderly and children population. The advantages of this approach include less inconvenience and stress on the subjects involved. Biological samples are collected at different time-points from different patients who are receiving different dosages and formulations of the same medication. It is important to record the time of administration and sample collection accurately. The number of the selected samples should be sufficient to help in addressing the study objectives to extrapolate the results on the population who have the same characteristics. Moreover, the followed procedures of storing and analyzing the collected samples should be well recorded. Simultaneous measurement of drug concentrations in the blood with efficacy and safety endpoints is useful for understanding drug concentration-response relationships. Furthermore, the Bayesian estimation method can be utilized to estimate pharmacokinetics parameters of

individual subjects from a small amount of blood concentration data. As a result, sparse response data are analyzed using the population pharmacokinetics method (54, 57).

1.5.2.1.2.1. Approaches to designing population pharmacokinetics studies

There are three pharmacokinetics screening methods included in the population pharmacokinetics studies: the single-trough screen, the multiple-trough screen, and the full screen. There are several factors based on which each of the previously mentioned screening method is selected, like dosage form, feasibility, study objectives, and outcomes (54).

A) Single-trough sampling design

The investigators obtain one blood sample from each participant before the administration of the next dose to be collected at the trough of the administered drug concentration. Consequently, the frequency distribution of plasma is calculated for each patient sample (58).

Three criteria should be met to determine the accurate variability in the trough concentrations in the target population by using a single-trough sampling design:

- Having a large sample size
- Minimum assay and sampling errors
- Giving all the patients the same dosing regimen and sampling times (59).

B) Multiple-Trough sampling design

In such a model, investigators obtain more than one blood sample from each participant before the administration of the next dose to be collected at the trough of the administered drug concentration. This design helps in separating both of the interindividual and residual variabilities. Furthermore, participants' characteristics are studied with high precision by using this design (60).

C) Full population pharmacokinetics sampling design

This design is also called experimental population pharmacokinetics design or full pharmacokinetics screen. The investigators obtain many blood samples (1 to 6-time points) from each participant after the administration of the drug of interest (61).

1.5.2.1.3. Bioavailability study

In bioavailability studies, the extent of the absorption and availability of active ingredients in bloodstream from different formulations is evaluated. The area under the blood or plasma concentration-time curve (AUC) is used to measure the extent of drug absorption, and the maximum concentration (C_{max}) is used as an indicator to determine the rate of drug absorption (62).

1.6. Study rationale

There are many developed and published critical appraisal tools and reporting checklists to guide researchers to assess the quality of the published research. However, the published appraisal tools are not highly specific to determine the methodological quality and validity of clinical pharmacokinetics studies. The validity of the study vary based on the methodological criteria that usually differ based on the study design. Furthermore, most of the available critical appraisal tools lack the presence of items that help in analyzing the published articles in depth (63). Additionally, many of the developed critical appraisal tools have not clearly been validated. This may create doubt to end-users about using certain tools to assess the validity and reliability of published articles (64).

The application of pharmacokinetics in the clinical setting is considered to be an integral part of providing pharmaceutical care services that are delivered by the pharmacist (65). Applying the main principles of pharmacokinetics should be based on a good understanding of the pharmacokinetics main principle processes

(absorption, distribution, metabolism, and excretion), which differ based on the specified drug, disease state, and patient population. Patient outcomes are improved by the appropriate application of the clinical pharmacokinetics principles from the published evidence-based information to decrease the following events: mortality, morbidity, length of treatment and hospital stay (LOS), adverse effects, and economic burden (66-70).

To our knowledge, two studies were conducted to assess the quality of reporting clinical pharmacokinetics studies. The first study was a systematic review which evaluated the quality of reporting of pharmacokinetics studies of antibiotics in patients with sepsis receiving continuous renal replacement therapy (CRRT). In this systematic review, researchers found that all of the identified articles during their systematic search did not report the required information that was essential for end-users to interpret the reported results. Furthermore, researchers noticed that 20% of the published pharmacokinetics trials did not contain the fundamental pharmacokinetics parameters (71). Consequently, reporting guidelines for clinical pharmacokinetics studies (The ClinPK Statement) were issued to assess and guide researchers while reporting their clinical pharmacokinetics studies. Researchers developed a Yes/No checklist that was composed of 24 items to guide researchers while writing their research and ensuring the reporting of the minimum required information in the published clinical pharmacokinetics studies (72). While offering a valuable guideline for reporting findings of clinical pharmacokinetics studies, the ClinPK statement guidelines did not cover all the dimensions of quality of trials including design, conduct, analysis, clinical relevance, quality of reporting, and results validity.

Therefore, there is a gap in the knowledge, as there is no available critical appraisal

tool that aids clinicians in appraising and determining the quality of the published clinical pharmacokinetics studies. Assessment of published clinical pharmacokinetics studies is highly critical as the application of pharmacokinetics principles in clinical practice settings helps in reducing mortality, length of treatment, length of hospital stays (LOS), morbidity, adverse effects of drug therapy, and cost-savings. Therefore, it is essential to develop such a tool to enhance the guidance of applying EbM application in practice.

1.7. Study objectives

1.7.1. General objective

To develop a valid and reliable critical appraisal tool for assessing the quality of clinical pharmacokinetics studies.

1.7.2. Specific objectives for phase I

To determine quality markers for appraisal of clinical pharmacokinetics studies based on available literature.

1.7.3. Specific objectives for phase II

To achieve expert consensus regarding the quality markers of clinical pharmacokinetics studies.

1.7.4. Specific objectives for phase III

To assess the psychometric properties of the developed critical appraisal tool.

1.8. Study significance

In particular, the project is a national health priority and is in total alignment with Qatar National Research Strategy (QNRS) and the National Health Strategy (NHS) 2011 – 2018 (73). The project addresses the specific goals of the QNRS and potentially has an impact on the region and global public health. QNRS strives to:

(a) “Enabling platforms for medical and public health research that can provide an evidence base for public policy and medical practice in Qatar (Goal HE.2)” (73). This goal stresses the need to conduct public health research to inform public policy and guide evidence-based population health programming.

(b) “Conduct public health research to inform public policy and guide evidence-based population health programming (Goal HE.2.2)” (73). This goal is aimed at research projects which will lead to better evaluation of the published clinical pharmacokinetics studies through developing validated critical appraisal tool for clinical pharmacokinetics. Such a tool will help in guiding the clinicians to decide about applying the results of these studies in clinical practice.

(c) Pharmacokinetics principles are applied in clinical practice settings to decrease mortality, reduce the length of treatment, cut the length of hospital stays (LOS), reduce morbidity, minimize adverse effects of drug therapy, and achieve cost-savings. Therefore, performing this project will help in fulfilling the aims of the Qatar National Strategy by 2030: improving the quality of research to enhance the effectiveness and the quality of care provided to patients (74).

Moreover, to our knowledge, this will be the first pharmacokinetics critical appraisal tool to be developed so copyright could be obtained under the name of Qatar. Additionally, this tool will be used by different end-users around the world once it gets published. It will guide researchers about the most important quality markers that they have to consider while conducting their research as well as reporting it in addition to the clinPK statement reporting guidelines. It will guide policymakers like World Health Organizations or Food and Drug Administration to take decisions on approving medications to be used to treat different diseases specifically on critical times like Coronavirus Disease- 2019 (COVID- 19) pandemic.

CHAPTER 2: A SYSTEMATIC REVIEW (PHASE I)

This chapter will provide a systematic review to identify the quality markers of the clinical pharmacokinetics studies. First, the importance of the appraising process is defined and brought into perspective. This is followed by discussions of the systematic review methods. Second, the identified quality markers of the clinical pharmacokinetics, which will be used throughout the subsequent phases, will be described by presenting the results of the systematic review. The chapter ends with discussing the results in terms of the existing literature and the need for additional research.

2.1. Introduction

2.1.1. Importance of the critical appraisal process

Critical appraisal tools are useful in the provision of analytical techniques to evaluate the quality of the study, specifically, the methods that are used to develop the study to reduce the incidence of bias. Quality of trials has several dimensions: study design, conduct, analysis, clinical relevance, quality reporting, and result validity (75). Since these kinds of factors affect the results of the study and the interpretation of the study findings, it is essential for research consumers to ensure if the study results can be generalized into their settings like policy, further research studies, education, or clinical practice. Therefore, selecting an appropriate critical appraisal tool is essential for applying evidence-based practice (9).

2.1.2. Critical appraisal tools

Scientists classified critical appraisal tools broadly into two categories: generic and research design-specific. Tools that are classified as design-specific consist of themes that are related to methodological issues that are unique for the research design (8, 76). Moreover, researchers developed generic critical appraisal tools aiming to

improve the research consumer's ability to appraise quantitative and qualitative studies to be able to come up with reliable evidence (77). There are some studies that were conducted to modify different kinds of tools like AMSTAR study that is developed to assess the quality of the systematic review methods based on previously established tools, other empirical evidence, and expert consensus (10). In addition, different tools like PRISMA for systematic reviews and meta-analysis was developed after noticing the poor reporting of vital information of systematic reviews (78). Despite the acknowledged significant value of the critical appraisal tools (79), there are no consensus regarding the 'gold standard' tool specifically for clinical pharmacokinetics studies.

2.1.3. Importance of clinical pharmacokinetics

Based on the American Society of Health-System Pharmacists, clinical pharmacokinetics is defined as ''the process of applying pharmacokinetics principles to determine the dosage regimens of specific drug products for specific patients to maximize the therapeutic outcomes and minimize toxicity'' (67). Clinical pharmacokinetics studies are rapidly advancing our knowledge pertaining to how drugs affect patients and populations and are essential components of the drug approval process. Data from clinical pharmacokinetics studies are commonly used to make decisions for drug approvals and funding, and also support clinical decision-making in patient care settings (70). In particular, these studies provide guidance on difficult-to-treat situations, such as those patients with organ dysfunction, obesity, comorbid conditions, or those taking other drugs prone to drug-drug interactions. In the era of individualized medicine, clinical pharmacokinetics studies are becoming very important for understanding drug response characteristics that may influence efficacy and/or safety for patients with profiles or characteristics outside of large

phase III studies (65-70). Due to the high utility of these studies for both drug approval/funding and clinical decision-making, it is essential that published studies are of high quality and that results are interpreted in light of actual and potential sources of bias. As such, these studies should be prone to rigorous appraisals, based on quality markers specific for pharmacokinetics studies.

2.1.4. Gap in knowledge

Critical appraisal of scientific literature is a foundation for the evidence-based healthcare movement and important for ensuring clinical decisions are being made using the best data possible. As such, critical appraisal tools are useful to evaluate study quality and to identify actual and potential sources of bias within a published (or submitted) paper. Research to date has focused on reporting guidelines for these studies, which offers excellent criteria to assess for the presence of required elements within a manuscript (72). Reporting checklists, however, are not intended to facilitate the appraisal of a study. A study may contain required aspects for reporting but may not meet expected quality standards. Furthermore, reporting checklists do not assess important dimensions of quality, which relate to the study design, conduct, analysis, clinical relevance, and result validity (75). While much is known about quality of clinical studies in general, these dimensions may consist of many categories or items that may be specific to a study type or even field of research. Quality markers for randomized controlled trials, for example, may not be relevant or all-encompassing for clinical pharmacokinetics studies. Tailored critical appraisal tools may, therefore, assist appraisers to focus their analysis on the most relevant aspects of study design, results, and reporting. Before an appraisal tool aimed at assessing the quality of conduct of a study can be developed, relevant quality markers of the intended study type must be identified. Building on the work of the previously published reporting

checklists, the aims of this systematic review were to create an inventory of quality markers intended for the appraisal of clinical pharmacokinetics studies and to categorize identified markers into associated domains of study quality.

2.2. Methods

2.2.1. Protocol development

A protocol was developed using the principles of the Cochrane Handbook (80). The approach and eligibility criteria of the systematic review to answer the research question was predefined and reported in the protocol. The protocol was registered and published in PROSPERO [registration number CRD42018094571] (81).

2.2.2. Selection criteria

Articles, including primary studies, systematic reviews, reviews, organizational reports, and guidelines, were included in this systematic review. An article was eligible for inclusion if any aspect of the trial's quality relating to study design, conduct, and analysis, clinical relevance, quality of reporting, or result validity were discussed. Articles were limited to those reported in English on human subjects only. Cell-based and animal-based pharmacokinetics studies were excluded.

2.2.3. Data sources and search strategy

A search of MEDLINE (1946 to March 2018), EMBASE (1974 to March 2018), Cochrane database of systematic reviews, Google and Google Scholar was conducted independently by two investigators (AS and SP) to ensure quality and optimization of the results. The following search terms were used and combined using the following Booleans: ("Pharmacokinetics" OR "Pharmacokinet*" OR "Clin*Pharmacokinet*" OR "Population pharmacokinetic") AND ("guidelines as topic" OR "Report* guideline*" OR "Evidence-based practice" OR "Appraisal tool*" OR "Checklist" OR "Scale") AND ("Quality indicators, healthcare" OR "Quality"). In Embase, the

MeSH terms were “Pharmacokinetics”, “population pharmacokinetics”, “practice guideline”, “evidence based practice”, “checklist”, “health care quality”. MeSH terms were exploded where appropriate. Keywords including “Pharmacokinetics”, “Pharmacokinet*”, “clin* pharmacokinet*”, “population pharmacokinetic”, “practice guideline”, “report* guideline*”, “evidence based practice”, “appraisal tool*”, “checklist”, “scale”, “health care quality”, “quality” were also included in the search. In MEDLINE, the MeSH terms were “Pharmacokinetics”, “Guidelines as topic”, “Quality indicators, healthcare”. Keywords including “Pharmacokinetics”, “Pharmacokinet*”, “Clin*Pharmacokinet*”, “Population pharmacokinetic”, “Report* guideline*”, “Evidence-based practice”, “Appraisal tool”, “Checklist”, “Scale”, “Quality” were also included in the search. Reference lists of the included articles were searched manually to include other relevant articles that were not identified while conducting the systematic search.

2.2.4. Selection of studies for inclusion

All identified articles were combined, and duplicates removed using ENDNOTE (Clarivate Analytics, Philadelphia, PA). Two investigators (AS and SP) then independently reviewed the title and the abstract of identified studies against the pre-determined inclusion criteria. Discrepancies for inclusion were resolved through discussion or by consulting a third investigator (KW). Full-text articles were then extracted for assessment of eligibility.

2.2.5. Data extraction

A data extraction tool was developed to extract data from the included studies. The information included: author, journal, title, year, categories and subcategories of quality markers. Categories included different sections of a manuscript including abstract, introduction/background, methodology, results, discussion, and conclusion.

Subcategories included subsections within a category section of the article. An example of this would be the subcategory ‘sampling’ under the category ‘methodology’. Identified quality markers within each subcategory were listed along with the description used to identify the quality marker with each subcategory. Data for analysis were extracted by one investigator (AS) and verified by another (SP or KW).

2.3. Results

2.3.1. Included articles

A systematic electronic search of different databases ended with 607 search results. After seven duplicates were removed, the title and the abstract of 600 articles were reviewed by two independent investigators, and 473 of which were excluded due to their irrelevance to our purpose. Full text of 131 articles was retrieved and reviewed including 4 newly added references identified from manual searching. A total of 15 papers were included for the extraction of quality markers. Articles included one original article about assessing the quality of reporting of clinical pharmacokinetics studies (72), one systematic review (82), one mini-review (83), two organizational reports (84, 85), eight reviews (86-93), and two guidelines (94, 95). The included articles discussed the quality markers pertaining to retrospective and prospective clinical pharmacokinetics studies, bioequivalence studies, as well as population pharmacokinetics studies. Search results, including reasons for exclusion, are provided in Figure 2.

2.3.2. Studies characteristics

The included articles discussed the quality markers of different aspects and types of clinical pharmacokinetics like a retrospective and prospective clinical PK studies, bioequivalence studies, as well as population PK studies. Quality markers

encompassed 19 subcategories. Quality markers were most frequently identified within the subcategories of methods and results (15 and 11 articles, respectively) as shown in Table 2. Quality markers related to title and the abstract were identified in one study (Table 3). One of the most frequently identified quality markers in the background was providing a clear objective identified in 7 studies (Table 4). The most commonly identified quality markers in the methods section were represented in Figure 3. Baseline characteristics, blood and tissue sampling and PK modeling were the most frequently discussed quality markers in 10 studies out of the included 15 in this systematic review. Figure 4. includes the most commonly identified quality markers in the results section. Reporting results through using measure of precision and quantification of missing data and outliers were identified in 6 studies out of the included 15 studies. Table 5 includes the most frequently identified quality markers in the discussion and conclusion sections. The most commonly identified quality marker in the discussion section was the discussion of the generalizability and applicability of the results, which was identified in 5 studies. A description of the identified categories and subcategories of quality markers was provided in Table 6.

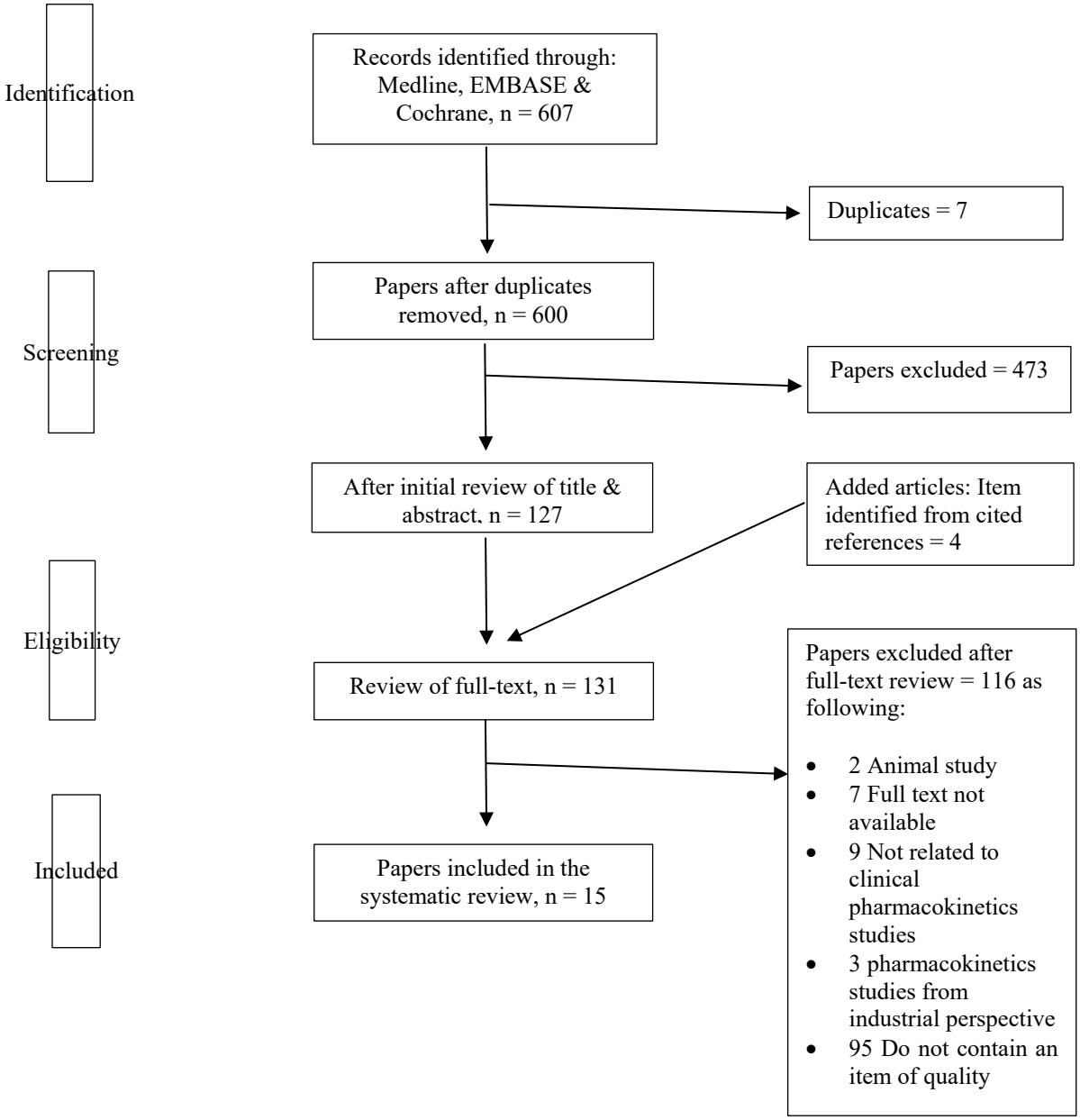

Figure 2. Flowchart showing the literature search and articles selection

Table 2: The frequency of identification of major categories

Quality markers	Frequency
Title/ abstract	1/15
Background	7/15
Methods	15/15
Results	11/15
Discussion/ conclusion	5/15

Table 3: The frequency of the identified quality markers in title and abstract

Quality markers	Frequency
Name of the analyzed medication and patient population	1/15
Brief description of the objectives, methods, results of primary objectives and conclusion	1/15

Table 4: The frequency of the identified quality markers in the introduction section

Quality markers	Frequency
Background about the analyzed drug	4/15
The rationale of the study	2/15
Goals/ Objectives/ hypothesis	7/15

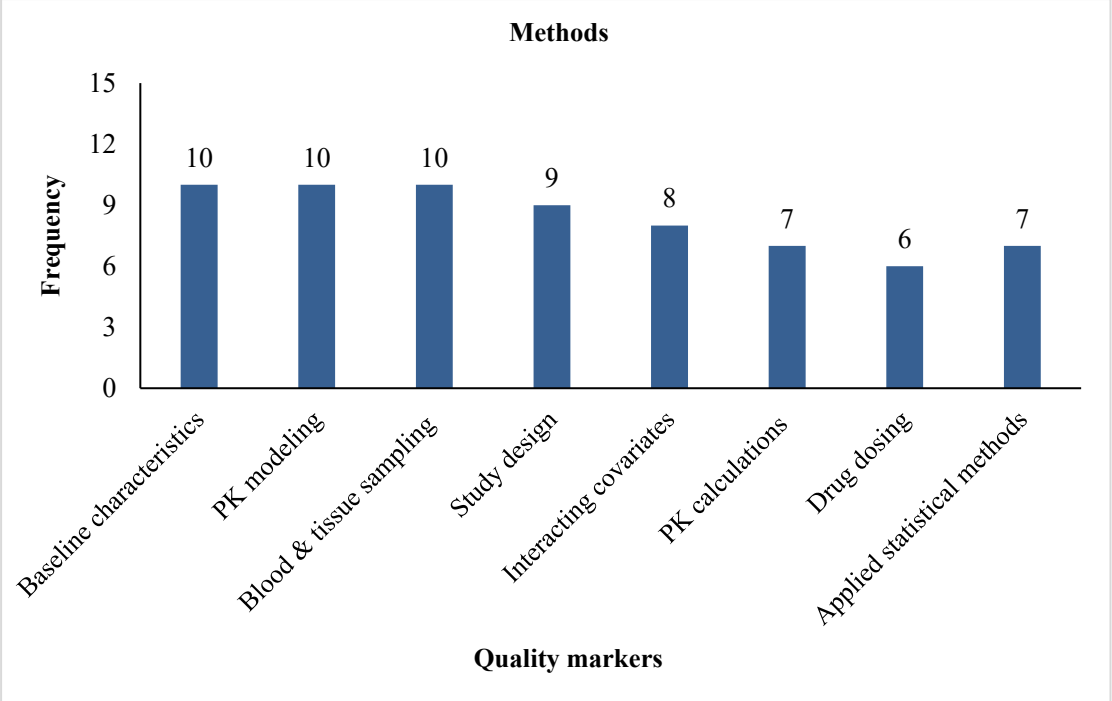

Figure 3. The most frequently identified quality markers in the methods section

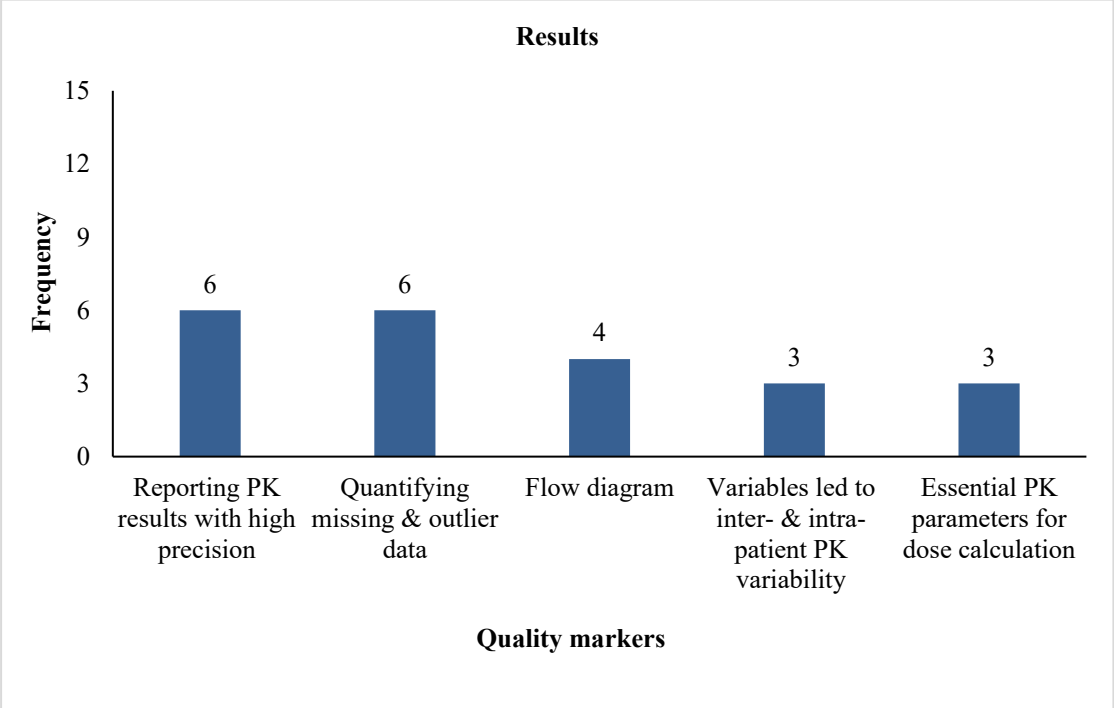

Figure 4. The most frequently identified quality markers in the results section

Table 5: The frequency of the identified quality markers in the discussion and conclusion sections

Quality markers	Frequency
Authors should describe the strength and limitation of the study and any source of bias.	2/15
Authors should discuss the applicability and the external validity of the study findings.	5/15

Table 6: Quality markers identified from the included studies

Category	Description
Title/abstract	
Title	The title should concisely reflect the discussed topic in the paper • The title should reflect the name of the analyzed medication, and comparator (if applicable), the targeted patient population, and the study design.
Abstract	A summary of the article should be provided to the readers within the abstract. • A brief description of the knowledge gap, the objectives, summary of the methods, the results of the primary objectives and the main conclusion should be provided. • The authors should not include any new information that was not mentioned in the article.
Background	
Introduction about the analyzed drug	Researchers should provide a comprehensive information about the analyzed drug. This information could be related to: • Stages of drug development. • Pharmacological aspects of the drug. • Pharmacokinetics aspects which are related to the drug administration, absorption, and elimination.
Rational	Researchers should provide readers with a clear description of: • the generated information out of the study. • what will be added to the existing body of knowledge?
Goals/ objectives/ rational	It should be written in the form of a clear statement to clarify: • the purpose and the scope of the study

Category	Description
Methods	
Subcategories of quality markers	Description
Study design	The chosen study design should be appropriately selected and justified. The following criteria should be considered: • Randomization • Allocation concealment: the used method should be clearly described. • Blinding: single, double, or triple blinded. • Monitoring patients' compliance by using electronic monitoring devices, patient diaries, and pill counts. • The number of subjects who will receive the interventions should be specified. • *Information about the medication:* − Medication name − Dose − Dose units − Schedule or frequency − Route of administration − Starting date of the medication − Stop date of the medication − Reason for use − Frequency − The recent date and time-related to medication administration • Inclusion and exclusion criteria: The inclusion criteria should allow authors to choose to study participants appropriately who are representative of the targeted population to answer the main study question. The exclusion criteria should not be restricted to an extent that interferes with the generalizability of the study results on the targeted population

Category	Description
	Study setting and/or location should be relevant to where the drug would be used**.** This will help practitioners to generalize the results of the research in case of working in the same setting. Sampling*:* *Sampling site:* should be precisely described and justified. • Arterial sampling is preferable during frequent sampling schedule, compared to the venous sampling since it is more representative of the delivered concentration to the effect site in case of peripheral elimination. Arterial sampling is preferable in the case of administering a drug that has a short duration of action or fast onset of action. Furthermore, venous blood is variable as it depends on tissue uptake. *Sampling interval:* the exact times at which samples are obtained should be described precisely. The sampling interval should not exceed the expected half-life of the studied phase. • *Sparse sampling: a* small number of samples collected from patients. 3 to 5 samples are collected in population pharmacokinetics studies. • *Traditional sampling:* (the most commonly used in PK studies), 10 to 20 samples are commonly collected after 1 to 2 minutes from the intake of the injection. • *Early intensive sampling:* It helps in describing and identifying the intravascular mixing phase and characterizing the concentration of medication at the peak. *Sampling schedule:* for the exponential phase (fast distribution, slow distribution and elimination) to be well characterized, 3-4 samples are the minimum number of samples to be collected. The time points of sampling should be translated into time windows which are more applicable in a clinical setting. Study endpoints should be appropriately selected to help in answering the research question. *Storage condition after collection:* The following should be considered: • Anticoagulant: the selected one should not interfere with the analyzed medication or its metabolites. • Stabilizers: should be added to the blood collecting tubes prior to collecting samples. In case if the stabilizers added to the blood, the sample should be cooled in ice prior adding the stabilizer • Centrifugation: proximity of a centrifuge (lab top or microcentrifuge); centrifugation procedure. • Temperature: ice-water bath should be used to keep the samples after collection. Pre-cooled collection tubes can be used for the collection of unstable drugs. • Labeling: should be resistant to freezing • Urine sampling: the volume and the PH of the urine should be measured once the sample is collected.

Category	Description
PK modeling	There is no one acceptable modeling method as each one has its own assumptions and limitations. Therefore, it is recommended that the authors should describe their choice and justify or reference it. They should describe the selected software package based on which they select the model. It is recommended to provide the following details about the modeling process: • Description of studies from which dataset was driven • Model structure • Used software and fitting algorithm • Methods used to test covariates • Validation and simulation methodology of the methods • The inclusion of uncertainty statement of a lower limit of assay quantitation
Population PK model validation	There is no consensus regarding the recommended validation method that should be used, but they can be described by increasing order of quality: *Basic internal methods* • Goodness-of-fit plots/diagnostic plots. • Uncertainty in parameter estimates • Model sensitivity to outliers *Advanced internal methods* • Data splitting • Bootstrap • Cross-validation • Simulations such as visual or posterior predictive checks (PPCs) *External model evaluation* (validation dataset observations compared with model predictions)

Category	Description
Table of demographic and covariate information	Interacting covariates may include the following: demographic variables, laboratory values, co-medications, environmental factors and disease states. • The covariates should be predetermined based on biological plausibility • Statistical measures like histograms should be used to show the distribution and frequency of covariates values. • In the case of investigating the relationship between different covariates, it is preferable to represent that in the form of graphs. • Researchers should differentiate between statistically significant covariates and clinical relevant ones. • Statistical summary and histograms of the continuous covariates should be documented to enable the regulatory assessor to judge if the range/frequency and the distribution of the values of covariates are sufficient to permit a valid conclusion to be drawn. • It is recommended to document the PK parameters, which are significantly affected by covariates. This will help in determining the subgroup of patients who need specific dosing recommendations. Performing stimulation is recommended to demonstrate the effect of the combination of various covariates on a group of different subjects.
Analytical method description	A clear and detailed description of the used analytical method should be given to an extent that allows its reproducibility For example: *Chromatography type:* (High-Performance Liquid Chromatography (HPLC) or Gas Chromatography (GC)) Detection type: (ultraviolet, fluorescence, electrochemical, mass spectrometry). *Assay characteristics:* mobile phase composition, gradient and flow rate, chromatographic column (packing material, dimensions),analytical run time, operating temperature and detection parameters. *Validation method:* All of the following factors should be discussed: specificity, recovery, linearity and sensitivity, the stability of the assay and its reproducibility

Category	Description
Calculation of pharmacokinetics parameters	Researchers should disclose the used equations to calculate different PK parameters like creatinine clearance, weight metrics used in the pharmacokinetics calculation The method used to estimate Area under the curve (AUC) and Area under the first moment curve (AUMC)
Applied statistical methods	Researchers should define the level of statistical significance. Sample size calculations should be described or referenced. The used software should be documented.
Ethical consideration	Authors supposed to state the approval number that was provided by their Institutional Review Board (IRB)
Results	
Baseline characteristics	All the following participant variables should be clearly defined: sex, race, age, weight, height, concomitant diseases, co-medication, smoking habits, covariates, the severity of illness, residual, renal function, and hepatic function. Acute Dialysis Quality Initiative (ADQI) minimum reporting criteria: a) Operational characteristics • Membrane/dialyzer/filter and area • A measure of time actually spent on therapy • Delivery device • Access and blood flow • Anticoagulation • Replacement fluid composition and administration • Dialysis fluid composition and administration b) Patient characteristics • The measure of time actually spent on therapy • Surgical/trauma/medical/other • The measure of severity of illness • Cointerventions • Integrated hemodynamic status and vasopressor treatment • Outcomes

Category	Description
	Renal failure • Cause • Plasma creatinine concentration/creatinine clearance • Plasma electrolytes • Hemoglobin concentration • Plasma protein level • Time and the nature of the last dialysis • Existence of clinical edema • Existence of peripheral neuropathy Hepatic cirrhosis • Cause • Child's Classification (34) • Prothrombin time, platelet count • Albumin and globulin levels Critically ill patients • Clinical description • Apache II score • Plasma creatinine and electrolyte concentrations • Presence/absence of renal failure • Presence/absence of liver failure Thermal injury • Regular hematocrit Preoperative plasma albumin and globulin level

Category	Description
Quantification of missing data	The method that is used to handle missing data during the analysis of the results (by deletion or imputation) and how this affects the PK parameters should be described and justified.
Quantification of outlier	The method that is used to handle outliers during results analysis should be described. Furthermore, outlying data should be included in the final population PK model and, their effects on different PK parameters should be documented.
Patient flow diagram	The number and reasons for withdrawals and how this is handled during the analysis should be documented and justified.
Variables which cause intra- and inter-patient variability	Should be documented using the appropriate measures of variance
Pharmacokinetics analysis results	Should be documented by using the appropriate measure of precision.
Essential PK parameters	In case of bioequivalence studies: F (bioavailability), AUC, Cmax (maximum concentration), tmax (time to maximum concentration)

Category	Description
Population PK studies	The following items should be present in the results section: • Tables and graphs that show the development of the key model specifically structural models and evaluation of covariate that support the primary objectives should be clearly labeled. • Justification of key models' selection at different stages of the development process through using Goodness-of-fit (GOF) plots: (1) Predicted data versus observed data (PRED versus DV; a line of identity and a trendline should be included) (2) PRED versus weighted residuals (WRES; zero lines and a trend line should be included), (3) Time versus WRES (a zero line and a trend line should be included). Time can be both time after dose and continuous time (time in the study). • A form of the best-selected model • Forest plots density distributions, or histograms which show key model qualifications • Performance of the model • Inter and intra-subject variability should be reported and represented by using graphs.
Discussion	
Discussion	All of the following should be included: • A summary of the most important findings • How well the data are described by using a certain model. • Interpretation and discussion of the results and their clinical relevance to prior knowledge. • The limitations of the study and how they affect the robustness of the results. • Future directions and applications of the generated knowledge.
Conclusion	• The conclusion should be supported by the observed results • Authors should not provide any new information
Others	• Disclosing any funding resources or any conflict of interest.

2.4. Discussion

A major finding from this review is that the published literature supports the appraisal of quality markers specific to clinical pharmacokinetics studies. These include examples such as study design of clinical pharmacokinetics research, pharmacokinetics modeling, appraising the used analytical method, and population pharmacokinetics model validation. Despite also identifying many quality markers that can be extrapolated across research fields and study designs, it is important for any potential tool to consider those markers specific to clinical pharmacokinetics. Sampling strategy including sampling site, sampling interval, and sampling schedule, for example, may greatly influence the representation of the delivered concentration of the medication to the effect site. Additionally, interacting covariates were also identified that could have potential implications for affecting the PK parameters. Moreover, participants' baseline characteristics should be described in detail as it will help in determining the subgroup of patients who need specific dosing recommendations.

The findings of this systematic review must be considered in light of the previously published ClinPK reporting guideline. As discussed previously, this checklist is meant to determine if a manuscript includes information about each included component but does not intend to determine how well or to what extent each component was completed and if there was a bias. Although many identified items crossed over with the ClinPK statement, others were identified that may be more meaningful for appraisal. Some of these include the study design sub-category, which is highly essential to be appropriately selected by researchers to accept the published results of their study. Additionally, appeasing the provided details of the used analytical method such as chromatography system, detection instrument, assay characteristics, and

validation method is highly crucial because this will help end-users to determine if the process is replicable and appropriately was done. Furthermore, the importance of explaining how the sample size was calculated and if this sample size is sufficient is highly essential to be appraised as this will affect the generalizability of the produced data. Moreover, there was detailed information provided in Table 6. and below each subcategory to describe the element that should be present to consider it of high quality.

This review has some limitations that should be addressed. Firstly, the search strategy included terms related to quality, which may not have been stated clearly in the title of the article, the MeSH terms, or keywords when it was indexed in the search engines utilized in this review. Therefore, some studies were likely missed and not included in the final list of studies included in this systematic review. In addition, the method of sample preparation was not identified as a quality marker. The method of sample preparation aids end-users to determine essential information about the integrity of the measured drug or metabolites. Additionally, it also provides insights into other critical pharmacokinetics parameters such as protein binding. Another limitation of this study was that our means of data extraction was not piloted or validated to ensure that quality markers were extracted in a comprehensive manner. While data extraction was performed by one investigator, any uncertainty was discussed openly with the research team. Finally, the results of this systematic review provide a comprehensive inventory of what is thought to represent quality markers from published literature. In the current form, the list should not be used as an appraisal tool but may inform discussion of study quality and future consideration for tool development.

2.5. Conclusion

Critical appraisal is an essential process that aids in evaluating the quality of

published literature. The ability to appraise articles may also foster the successful application of knowledge gained from the literature to practice. In this systematic review, an inventory of quality markers was identified that encompassed both general aspects of study design, as well as specific considerations for clinical pharmacokinetics studies. These quality markers can help readers of clinical pharmacokinetics literature better understand and stratify high-quality research in this area of practice. Furthermore, these quality markers can be used to develop a critical appraisal tool for clinical pharmacokinetics studies.

CHAPTER 3: METHODOLOGY

This chapter will provide a step-by-step description of the used methodology in phase II and Phase III of this project. Phase II describes the modified Delphi process that is used to achieve expert consensus regarding the quality markers identified in phase I which is the systematic review. Refer to Chapter 2 for further details of phase I, the systematic review. This will be followed by phase III which shows the used methods to assess the reliability, validity, and feasibility of the developed tool.

3.1. Phase II

3.1.1. Study design

Phase II was a mixed-method development study design where one method (qualitative or quantitative) was used to help in developing/informing the other method (either the qualitative or the quantitative) (96). Exploratory sequential design was applied in which the qualitative method was used initially to explore the problem (the conducted systematic review in phase I to determine the clinical pharmacokinetics quality markers) as there were not enough studies on the topic, and then the collected data were used to perform the quantitative method (modified Delphi method and reliability testing through calculating Kappa values) (97).

3.1.2. Consensus methods

After conducting a systematic review to identify possible quality markers of clinical pharmacokinetics studies, expert consensus regarding the identified quality markers should be achieved. These methods are helpful when there is a lack of knowledge, evidence, or agreement on a particular topic (98). There are four commonly used types of formal consensus, as shown in Table 7.

The first consensus method is the Nominal Group Technique (NGT), which is a

structured procedure through which qualitative information related to the research question is collected from experts in the field. In the NGT, participants are asked to provide answers to a specific question privately without discussing them. After that, in a round-robin fashion, participants are invited to share their ideas. Then, each provided item is presented and a clarification is provided when necessary. Each participant is asked to rate the formulated ideas separately, and then the views of the group are assessed (99, 100).

The second method is consensus development panels/consensus development conferences, which are a kind of an organized meeting that includes experts in the field of the discussed issue. This method of consensus is usually used to develop policies and strategic plans. There are different methods of conducting consensus development panels. The most commonly used method is the one developed by the National Institute of Health that is used to organize conferences that aim to assess scientific literature surrounding pertinent biomedical issues (101).

The third consensus method is the RAND-UCLA Appropriateness Method (RAM). Two groups are involved in this process: the core panel and the expert panel. The core panel conducts a systematic literature review about the discussed topic to provide the experts' panel with evidence-based information. The expert panel will use this information to come up with a decision regarding the clinical scenarios that will be provided. The experts' panel has to rate the best intervention for each clinical scenario using a 9-point Likert-scale (100, 102).

The Fourth consensus method is Delphi. There are several designs of Delphi as presented in Table 8. Delphi is a process in which judgments are collected and aggregated systematically through controlling feedback by interspersing a series of questionnaires on a group of panelists who are considered to be experts in the field of

the discussed issue (103, 104). The modified Delphi process was chosen to achieve expert consensus on quality markers to be included in a critical appraisal tool for clinical pharmacokinetics studies.

Table 7: Methods of formal consensus

Methods of consensus	Mailed (through a Postal survey, internet, or fax) questionnaires	Private decisions elicited before group discussion	Formal feedback of group choice	Face-to-face contact	Interaction structured
Nominal Group Technique (NGT)	No	Yes	Yes	Yes	Yes
National Institutes of Health (NIH) consensus development conference methodology	No	No	No	Yes	No
RAND/UCLA Appropriateness Method (RAM) RAND	Yes	Yes	Yes	Yes	Yes
Delphi	Yes	Yes	Yes	No	Yes

Note: Contents of this table were derived from Murphy and Black (1998) (105)

Table 8: Different Delphi study design

Delphi study design	Aim	Administration	Round-1 design*
Classical	To determine expert consensus	Postal	The first round is open qualitative to allow participants to record their ideas
Modified		Postal, online	Panelists are asked to rate their agreement regarding predetermined items, that are provided, and select from a previously conducted literature review.
Real-time/consensus conference		Panelists use a computer technology that helps them to achieve consensus in real-time rather than post	Like the design of classical Delphi
e-Delphi		Email or an online web survey	
Online		The questionnaire is applied through using any online instrument like a chat room, or forum.	

Note: Contents of this table were derived from Hasson and Keeney (2011) (106).
*The difference among the Delphi study designs is mainly in round-1.

Modified Delphi design was chosen over the other methods of formal consensus for several reasons:

1- Modified Delphi design was chosen over the other methods because we conducted a systematic review to identify quality markers of clinical pharmacokinetics studies. Then, we formulated meaningful questions from the identified quality markers that can help end-users in appraising clinical pharmacokinetics studies.
2- A large number of participants could be included in the rounds through using modified Delphi design because of the lack of geographical constraints, as researchers could run the process by using SurveyMonkey. Therefore, researchers could consist of participants from different regions around the world. Qatar is a country that has a small population. Thus, few people fulfilled the defined inclusion criteria, so there was a need to look for other experts in different countries. This was possible using this method as there were no face-to-face meetings required in this consensus process compared to others.
3- Cost-effective: This process was carried out through using the available resources like a laptop, internet, and SurveyMonkey.
4- Participants could express their opinions freely, as their answers were anonymous. Therefore, a dominance that might occur during face-to-face interaction when participants meet with people who were highly opinionated in the field, was avoided as well as any bias introduced by moderators.
5- Ease of communication: Participants participated and communicated with the moderators from different countries since the modified Delphi method could be conducted via an online survey platform.
6- Knowledge sharing: Panelists in this study were recruited from different countries and sections of practice like academia, industrial section, clinical section, and

regulatory section. Thus, each participant could share a piece of knowledge regarding the field of clinical pharmacokinetics from their perspective and help in generating new ideas that helped in broadening the knowledge base of other participants (107, 108).

3.1.3. Study setting and mode of interaction

This was a modified Delphi study that was conducted via an online survey platform, which was SurveyMonkey (109). Each participant received an email, which included the customized link of the survey, which was an environment-friendly method of communicating data between participants. This method of communication provided a chance to contact experts in the field from different countries during the same time. Furthermore, this also gave panelists the freedom to answer the questionnaire at the place and time that comfort them, which in turn might affect the quality of their answers and comments. This method of communication helped maintain the anonymity of participants. Moreover, SurveyMonkey (110) helped in collecting data and importing it into SPSS (111, 112).

3.1.4. Inclusion criteria

The inclusion criteria were defined based on the four essential requirements that the expertise should meet (113):

- Selected people should have experience in the discussed issue.
- Participants should agree to participate in the process.
- Participants should be able to communicate and to deliver their point of view.
- Participants should have enough time to participate in the rounds.

Therefore, we defined the people who should have experience in the discussed issue (item #1 in the inclusion criteria) as any one of the following:

- Academicians with a position that reflects their direct involvement in the research field of clinical pharmacokinetics.
- Clinicians who have experience in the application of clinical pharmacokinetics principles in their clinical practice. These individuals should have experience in interpreting the findings of clinical pharmacokinetics studies and applying these to their patients.
- Pharmaceutical industry researchers with experience in clinical pharmacokinetics.
- Individuals in health regulation who assess clinical pharmacokinetics studies when making decisions for their respective health authorities.

3.1.5. Sampling method

A purposeful sampling method was used to identify individuals who are knowledgeable and had experience in the field of the investigated issue (114). Multiple strategies of purposeful sampling were applied during the sampling process. A Criterion-i sampling strategy was applied since all the individuals were selected based on prior explicitly defined inclusion criteria. Furthermore, the maximum variation strategy was used since different pharmacokinetics stakeholders were approached from the academic, industry, health regulation, and the clinical sectors. These strategies were used to make sure that participants represented all end users who might use the developed clinical pharmacokinetics critical appraisal tool. Convenience sampling was also used since team members who worked in this research emailed individuals who knew that they met the inclusion criteria. Different strategies were used to enhance the depth and breadth of the studied issue.

3.1.6. The number of participants

To our knowledge, there was no documented consensus on the exact number of

participants that should be included in the modified Delphi process. Delbecq, Van de Ven, and Gustafson documented that the size of the Delphi panel was variable and recommended that researchers should recruit a sufficient number of participants. They suggested that researchers should recruit ten to fifteen participants in case if they had a homogenous background while the number of participants should be more if they had a different experience (115). Witkin and Altschuld documented that in general, the number of the included participants in the Delphi study was less than 50. In agreement with this, Ludwig recorded that most of the researchers recruited between 15 -20 participants in most of the conducted Delphi studies (116). Therefore, 119 participants were approached through sending emails to people who the supervisory committee members knew that they will fulfill the inclusion criteria. Additionally, WhatsApp messages were sent to groups that include experts in the field of pharmacokinetics. The 119 participants were approached to count for people who may not agree to participate, who may decide not to respond or who is anticipated to drop-out between rounds during the process and, to ensure that the provided judgments by the panelists about the issue were representative. A cover letter that gave the participants a background about the rationale behind conducting the project, the goal and the specific objectives of the project was sent to participants who agreed to participate in a Word and PowerPoint format, Appendix A. The rules of modified Delphi were also disclosed to them, Appendix B.

3.1.7. The number of rounds

The number of rounds in the modified Delphi process is usually based on the agreements and disagreements that occur among participants and reconsideration of certain items based on the received responses from participants. It was recommended by Delbecq, Van de Ven and Gustafson that two to three rounds were enough for most

of the studies for Delphi studies while three rounds or more were required in case of recruiting heterogeneous samples. On the other hand, as the number of rounds increases, the rate of responses might decrease which in turn affects the quality of the process. Therefore, we decided to have 3 to 4 rounds in the modified Delphi process as we have a heterogeneous sample of panelists who represent different sectors and to avoid any decline in the response rate and the quality of the process (117).

3.1.8. Ethical consideration

This study was approved by the Qatar University Institutional Review Board: QU-IRB 970-E/18, Appendix C.

3.1.9. Likert-scales

The most commonly used Likert-scales to measure opinion or attitudes are the ones which have 5 or 7 response categories. 5-point Likert-scale has high reliability (118). In this study, a 5-point Likert-scale is considered understandable and helpful for panelists to express their point of view compared to 3-point Likert-scales (panelists have to entirely agree, entirely disagree, or be neutral regarding the included items). Furthermore, it is less confusing compared to 7-point Likert-scales in this study as it is difficult for the panelists to differentiate between strongly agree, agree, and somewhat agree. 5-point Likert-scale contains a neutral point, which was highly needed in this research so panelists could select it if they did not have the required knowledge to agree or disagree on including any question in the final draft of the clinical pharmacokinetics critical appraisal tool. Furthermore, the selection of a 5-Point Likert-scale also helped in increasing transparency since the inclusion of any item in the final draft of the critical appraisal tool formulated from on the panelists' knowledge and experience (119).

3.1.10. Questionnaire development

The questions were developed based on the pre-identified quality markers from the conducted systematic review in phase I. Sixty-four meaningful questions were formulated from the identified quality markers. These 64 questions were written in a way that helped in assessing the quality of published clinical pharmacokinetics studies. Then, these 64 questions were revised scientifically and linguistically, reworded, and reduced after several discussions with the supervisory committee members of the project. Questions were deleted if they are related to reporting not appraising. Furthermore, some questions were combined together. SurveyMonkey was used to design the questionnaire for each round. The survey was divided into eight main sections (title and abstract, introduction, methodology, results, discussion, conclusion, others, and demographics). A total of 42 questions were formulated, which are related to the previously mentioned sections, including 2 questions serving as a control. Panelists were allowed to suggest modifications and rewording of any item. In addition to that, they were asked to indicate the addition of any question if needed.

3.1.11. Data analysis

There is no agreement between researchers on defining what constitutes consensus (120) . In this study, consensus was used to determine the level of agreement of each individual participant on including each question in the final draft of the clinical pharmacokinetics critical appraisal tool. This was used to determine the opinion of the group and the extent to which participants agree with each other (121). The consensus in this study was based on the percentage of agreement, median and interquartile range. As presented in Table 9, for any question to be included in the final version of

the clinical pharmacokinetics critical appraisal tool the percentage of the intra-rater agreement should be more than or equal 75 %. This percentage was selected because the minimum acceptable percentage of intra-rater agreement was 70 % or more. Furthermore, the panelists should provide a positive result by selecting 4 or 5 on the 5-Point Likert-scale. A median score of more than or equal 4 and interquartile-range of less than or equal 1 were also used to assess agreement. Interquartile range was selected because it is less affected by the presence of the outliers and as the value becomes 1 or smaller this shows that data points that were selected by the panelists were closely packed to the median.

On the other hand, the exclusion criteria of any question from the final draft of the clinical pharmacokinetics critical appraisal tool included a percentage of the intra-rater agreement of less than 75 %, a median score of less than or equal 2 and interquartile-range of more than 1. Moreover, if the change in the distribution of responses was less than 15 % between rounds, the question was excluded. Additionally, every negative comment received from the panelists was considered to take the decision of excluding the item from the tool. If the question did not meet the inclusion or the exclusion criteria, the question was modified based on the panelists' comments, and a summary from the panelists' remarks was disclosed with the question to the next round (122, 123).

Table 9: Consensus thresholds

Inclusion	• More than or equal 75% of participants provide a positive result by selecting 4 or 5 on the 5-Point Likert-scales. • A median score of more than or equal to 4. • Interquartile-range of less than or equal to 1.
Exclusion	• Less than 75% of participants provide a negative result by selecting 1 or 2 on the 5-Point Likert-scales. • A median score of less than or equal to 2. • Interquartile-range of more than to 1. • Negative comments regarding the inclusion of any item.
Non-Consensus	• Questions that did not meet any of the inclusion or exclusion consensus thresholds, researchers modified them based on the panelists' comments and circulated for the second round. Then, the change in the distribution of responses to each question between the rounds was measured. If the change in the distribution of responses was more than 15% between rounds, the question modified again based on the panelists' comments and recirculated to the next round. • On the other hand, if the change in the distribution of responses was 15% or less between rounds, the question was excluded because this shows that there was stability. • Negative comments regarding the inclusion of any item.

Table adapted from Schneider, Evaniew, and Garland (122, 123).

3.2. Phase III: Application of a systematic approach to evaluating psychometric properties of the clinical pharmacokinetics critical appraisal tool

3.2.1. Definitions

Reliability is defined as “the ability to reproduce a consistent result in time and space, or from different observers, presenting aspects on coherence, stability, equivalence, and homogeneity” (124). Reliability is calculated by dividing the true variance by value of the addition of true variance and the error variance. The value of reliability falls between 0 and 1, as the value becomes closer to 1, this shows stronger reliability. Validity refers that the tool precisely assesses what it is designed to measure (125). Validity is a characteristic of an instrument since it should be determined whether a

tool evaluates the specific issue in a defined population (126).

3.2.2. Introduction

This phase describes the methods of our third research question.

3.2.3. Research question

In phases I and II, we identified the quality markers, and out of which we formulated essential questions that help in assessing the quality of published clinical pharmacokinetics literature, respectively. A draft of a clinical pharmacokinetics critical appraisal tool would be the end-product of the modified Delphi process. In phase III, the validity and reliability of the developed clinical pharmacokinetics critical appraisal tool were assessed. The research question was: Is the developed clinical pharmacokinetics critical appraisal tool valid and reliable? We used different statistical tests to answer this question.

3.2.4. Objectives

The purpose of this study was to critically evaluate the psychometric properties of the developed clinical pharmacokinetics critical appraisal tool.

3.2.5. Methodology

3.2.5.1. Study design

3.2.5.1.1. Sampling process

Thirty recently published clinical pharmacokinetics articles were selected from three different journals: Clinical Pharmacokinetics, International Journal of Pharmacokinetics and Journal of Pharmacokinetics and Pharmacodynamics (127). These journals were selected with clinical pharmacokinetics scope, with the highest reported impact factor to ensure that the published articles were of, at minimum, modest quality. Furthermore, these journals were chosen from the subscribed journals by Qatar University library to ensure the availability of the full text of all the selected

articles from these journals. The selected sample of articles contained different types of clinical pharmacokinetics literature including phase I and phase II clinical pharmacokinetics trials, population pharmacokinetics studies, bioequivalence studies, and studies investigating the pharmacokinetics involvement in drug-drug interactions. There is no consensus on the absolute number of papers that should be used to assess the validity and reliability of the developed tool (128). It has been recommended by different guidelines that respondent (in this study it is an article) to item ratio should be a range of 5 articles:1 item, 10 articles:1 item, 15 articles:1 item, or 30 articles:1item (129). It has been recommended to include a minimum number of 30 heterogeneous samples and a minimum of 3 raters in a reliability study (130, 131). The previously mentioned rule was applied by researchers who used 30 articles to assess the validity and reliability of the AMSTAR tool (127). Furthermore, it was reported that the sample size should consist of a minimum of 30 comparisons required to calculate the confidence interval to avoid having confidence interval resulting in no agreement (132). Therefore, the final clinical pharmacokinetics critical appraisal tool was applied to 30 recent clinical pharmacokinetics studies to test the reliability and feasibility of the tool while evaluating the quality of published studies.

3.2.5.1.2. Sample criteria

Articles were selected if they were randomized clinical pharmacokinetics trials, drug-drug clinical pharmacokinetics interaction trials, population pharmacokinetics studies, or bioequivalence studies that were published in the English language and included human beings only. Three different pharmacokinetics journals were selected from the subscribed journals by Qatar University library. The recently published articles were selected based on the previously mentioned inclusion criteria.

3.2.5.1.3. Evaluators selection process

Four evaluators were selected to represent the end-users of the developed clinical pharmacokinetics critical appraisal tool. Two users had experience in academia and clinical practice sectors, the third one had a background in the industrial area, and the fourth evaluator represented students as an end-user of the developed tool.

3.2.5.1.4. Evaluation process

The thirty articles were divided equally amongst raters with each evaluator assessing 15 articles. Therefore, each included study was evaluated by two evaluators. Every two evaluators were assigned randomly by using excel to evaluate the same article. Evaluators were given five weeks to critically appraise the articles that were assigned to them by using the developed clinical pharmacokinetics critical appraisal tool. Evaluators were asked to answer each question by selecting Yes, No, Do Not Know, or Not Applicable. Furthermore, evaluators were asked to write any comments regarding any of the included items in the tool. The exact time they need to evaluate each article was also documented to determine the feasibility of the tool. To assess the reliability of the developed tool, we deemed each Yes response as a score of one and any other selected answer (No, Do Not Know, Not Applicable) a zero score.

3.2.5.2. Reliability testing

We tested two types of reliability in this study: equivalence and stability.

A) Equivalence is defined as the degree of agreement between two or more raters regarding the score of an instrument, which is known as inter-observer reliability. There are several statistical tests to measure the inter-observer reliability like percent agreement, Cohen's Kappa, Fleiss Kappa, and the intra-class correlation coefficient (useful when there are multiple raters and multiple possible ratings).

- Firstly, percent of agreement is obtained through creating a matrix. The columns of the matrix represent the different raters, and the variables of the

collected data are described in the rows of the matrix. Therefore, the scores of each variable should be represented in each cell in the matrix. Among the advantages of using this method is that random errors which are distributed among all the variables are identified. Furthermore, this method allows the identification of variables that may be problematic (the variables that do not meet the cut-off point of agreement, which is 80%) (133).

- Secondly, Cohen's Kappa method is used to assess both inter-rater and intra-rater reliability testing. This approach was developed because previous methods did not consider that agreement could occur due to chance regardless of the systematic process that the rater is using to categorize the subjects. Earlier methods are focused only on determining the proportion of observed agreement. Cohen suggested the use of a technique that helps in correcting the measure of the agreement due to chance, which is called Kappa. The use of Kappa will help in removing the proportion of the observed agreement by the expected level of agreement, considering the observed marginal distributions that show raters' responses while assuming that rates are working independently (134). The value of Kappa ranges from −1 to +1. A value of +1 represents a perfect agreement between raters, while 0 are expected values of agreement that occur due to chance. On the other hand, a value of -1 is a sign that the agreement between the two raters was less than expected and occurred due to chance (132). Kappa values are interpreted as follows based on Cohen's suggestion: values ≤ 0 as indicating no agreement and 0.01– 0.20 as none to slight, 0.21– 0.40 as fair, 0.41– 0.60 as moderate, 0.61– 0.80 as substantial, and 0.81–1.00 as almost perfect agreement. It has been reported by Landis and Koch that Kappa value < 0.00 is unacceptable (135).

The suggested formula (136) of Kappa by Cohen was:

- $\kappa = (Po - Pe)/(1 - Pe)$ *Equation (1)*
- Chance-corrected observed agreement: $Po = (a+d)/N$ *Equation (2)*
- Chance-corrected perfect agreement: $Pe = (f_1g_1+f_2g_2)/N^2$ *Equation (3)*

Table 10 represents a 2x2 table, which shows how each of the previous symbols is derived:

Table 10: 2x2 table

	Rating by observer K		
Rating by observer O	Yes	No	Total
Yes	A	b	g_1
No	C	d	g_2
Total	f_1	f_2	N

Cell A represents the number of "Yes" that rater K and rater O agreed on. Cell d represents the number of "No" that rater K and rater O agreed on. Cells C and b represent the number of questions that they disagree on. Cell f1 represents the total number of "Yes" selected by rater K. Cell f_2 represents the total number of "No" selected by rater K. Cell g1 represents the total number of "Yes" selected by rater O. Cell g_2 represents the total number of "No" selected by rater O. Cell N represents the total number of items that rater O and rater K are evaluating.

- An extended measure of Cohen's Kappa is called Flessi Kappa which is the third method of calculating the inter-observer reliability. Flessi Kappa was generalized to measure agreement and association among a fixed number of raters K (> 2) who are assigned to rate n subjects independently.

- The fourth method of calculating inter-observer reliability is the intra-class correlation coefficient (ICC): Earlier, several methods were used to evaluate reliability including the Pearson correlation coefficient, paired t-test, and Bland-Altman plot. Both paired t-test and Bland-Altman plot methods are used to analyze agreement while the Pearson correlation coefficient is only a measure of correlation. Therefore, the use of only one method will not be ideal for measuring reliability. In 1954, this method was introduced by Fisher as a modification of the Pearson correlation coefficient. Mean squares that are attained from the analysis of variance are used to calculate ICC (137). ICC was defined by McGraw and Wong in 10 forms based on the "Model" (1-way random effects, 2-way random effects, or 2-way fixed effects), the "Type" (single rater/measurement or the mean of k raters/measurements), and the relationship "definition" weather it is consistency or absolute agreement.

  Selection of the model in ICC:

  - 1-Way Random Effects Model is used when a group of raters are selected from a large population of possible raters and are divided into different sets. Each set of raters will rate certain subjects. This is only applicable in a multi-center clinical trial.
  - 2- Way Random Effects Model is used when researchers select raters from a large population who possess similar characteristics. This model is used when raters aim to generalize the reliability results.
  - 2- Way Mixed Effects Model is used when researchers are interested in specific raters. The reliability that comes out through using this model

is not generalizable even on other raters who have similar characteristics.

Selection of the type is based on how the measurement will be conducted: single rater/measurement or the mean of k raters/measurements. The selection of definition is based on what is more important: either absolute agreement or consistency agreement between raters from the researchers' point of view. The absolute agreement is when raters give the same rate for the same subject. Consistency agreement is when the assigned scores by raters to the same subjects are additively correlated (130).

We selected to assess the inter-rater reliability by using Cohen's Kappa because two raters were assigned to evaluate each article independently. Furthermore, the scale that we designed was dichotomous through which each 'Yes' was deemed one score and any other answers (eg. 'No', 'I Don't know', and 'Neither Agree Nor Disagree') were deemed a 0. 'No', 'I Don't Know', and 'Neither Agree nor Disagree' were grouped together as 'No' because the categories of the nominal scale to use Cohen's Kappa should be mutually exclusive (135). Furthermore, as the 'No', 'Neither Agree nor Disagree' and 'I Don't Know' combined together the proportion agreeing, *p*, increases so does the expected proportion agreeing with Pe (chance-corrected perfect agreement). Therefore, Kappa value does not increase since the proportion of agreeing increases. Additionally, this depends on the relationship between the categories. For instance, the value of Kappa gets smaller when there is a probability that an incorrect judgment is placed in a category that is independent of the true category. On the other hand, when the adjacent categories are combined, the incorrect judgment will be in either side of the

truth, thus the value of Kappa increases specifically with ordered categories (138).

Cohen's Kappa method considers the agreement that occurs due to chance regardless of the systematic process that the rater is using to categorize the subjects. We calculated the percentage of the agreement to compare it to the Kappa values in some cases. This was done because the Kappa coefficient was affected by the prevalence of either of the two choices 'Yes' or 'No'. The prevalence effect usually happens due to the skewness of the answers toward either 'Yes' or 'No' and the unbalance totals of the marginals (the value of f1 compared to the value of g1, the value of f2 compared to the value of g2 as shown in Table 10). This is expressed by the prevalence index. When the prevalence index is high, the chance agreement is high and consequently, the Kappa value becomes smaller, therefore, it is difficult to interpret the Kappa values that are known as paradox values in such cases (89, 139-141). Therefore, the prevalence index and the bias index were calculated to determine their effects on the questions that were affected by prevalence (136, 139).

- Prevalence index = |a-d|/N *Equation (4)*
- Bias index = |b-c|/N *Equation (5)*

If the prevalence index and the bias index were not zero, the values of Kappa were corrected based on the following equation (136):

- PABAK = 2P0-1 *Equation (6)*

  (PABAK, prevalence-adjusted-bias-adjusted Kappa)

B) Stability is defined as the consistency of the measurement repetition to determine the extent of the similarity between the results when repeated at

different times. This usually shows how stable the measure is throughout time. The test-retest method is used to perform the stability test by appraising the same paper by the same rater twice at different times. All the factors should be the same in both times of repeating the appraising process. The time span between the two times should not exceed 10 to 14 days. One of the most common measures used to perform test-retest are the ICC and Cohen's Kappa (142). We used Cohen's Kappa to calculate intra-rater reliability for the previously mentioned reasons. Five different articles were randomly selected to be re-evaluated by every 2 raters after 14 days from evaluating them for the first time. However, according to the review of sample size requirements for the design of reliability study, the same number of articles should be re-evaluated by the raters to calculate the intra-rater reliability (143). Additionally, the minimum required sample size is 30 comparisons to avoid having confidence interval resulting in no agreement (132). In this study, however, raters were asked to re-evaluate only 5 articles due to time constraints.

3.2.5.3. Validity testing

Two types of validity were tested: face and content validity. Content validity is defined as the extent to which the test contains all the essential items needed to measure a particular concept (142). Face validity is defined as the degree to which a test appears to measure what it is supposed to measure and this usually does not depend on technical items presented in the tool (131). Content and face validity were tested qualitatively through the modified Delphi process by a group of experts in the field of clinical pharmacokinetics. During the modified Delphi process, panelists were asked to agree or disagree if each of the written items in the questionnaire

appropriately tests the quality of the pharmacokinetics paper. Panelists were asked to evaluate the clarity of each question and were allowed to suggest the addition of any item. Furthermore, they were asked to assess if the language that was used to formulate each item is understandable. By the end of the modified Delphi process, the final draft of the developed tool was sent to an expert in the psychometric field in the College of Pharmacy at Qatar University. We asked him to revise the English and the scientific language of the developed questions and to make sure that there were no double paralleled questions.

CHAPTER 4: RESULTS

This chapter demonstrates the results of the outcomes of the development and testing of a critical appraisal tool for clinical pharmacokinetics articles that were selected for this project. The study was divided into three phases: Phase I - identification of clinical pharmacokinetics studies quality markers through conducting a systematic review (Chapter 2); Phase II – achieve expert consensus regarding the identified quality markers of clinical pharmacokinetics studies and develop a working draft of a clinical pharmacokinetics critical appraisal tool; and Phase III – assess the validity and the reliability of the developed clinical pharmacokinetics critical appraisal tool.

4.1. Phase I results

A search of MEDLINE (1946 to March 2018), EMBASE (1974 to March 2018), Cochrane database of systematic reviews, Google and Google Scholar was conducted. The systematic electronic search of different databases ended with 607 search results. After seven duplicates were removed, the title and the abstract of 600 articles were reviewed by two independent investigators, and 473 of which were excluded due to their irrelevance to our purpose. Full text of 131 articles was retrieved and reviewed including four references identified from manual searching. A total of 15 papers were included for the extraction of quality markers. The included articles discussed the quality markers of different aspects and types of clinical pharmacokinetics like a retrospective and prospective clinical pharmacokinetics studies, bioequivalence studies, as well as population pharmacokinetics studies. Quality markers encompassed 19 subcategories. Quality markers were most frequently identified within the subcategories of methods and results (15 and 11 articles, respectively). Refer to chapter 2 for further details.

4.2. Phase II results

4.2.1. Question formulation

In July 2018, 64 questions were formulated to assess the quality of the published clinical pharmacokinetics studies based on identified quality markers from the systematic review conducted earlier, Appendix D. As demonstrated in Figure 5, the 64 questions were reduced after several discussions with the supervisory committee members of the project to 44 questions summarized in Table 11. The twenty questions were removed from the 64 for several reasons including duplication, irrelevance to clinical pharmacokinetics studies, and some formulated to assess whether the information was reported instead of assessing the quality of the reported information. Then, several discussions were held among the supervisory committee members to modify the scientific and the English language of the formulated 44 questions. The number of the formulated questions was reduced again to 42 items, which were sent to the first round (Table 12). The difference between Tables 11 and 12 was in the used scientific language, as well as the removing, combining and adding new questions as described in the following lines. Questions 24 and 25 in Table 11 were combined with the question that asked about if the authors provided the used pharmacokinetics equation. Questions 6 and 16 were combined because both were assessing if the authors clearly reported the doses of the used medications in the study. The used control questions were modified to be related to the critical appraisal process. Nine additional questions were added to assess the quality of the randomization process, allocation concealment and the appropriateness of the used statistical tests. Participants were asked to rate the best rating scale for end-users. Additionally, the consent form was added as a separate question. Demographic questions were added to ask participants about their years of experience in the clinical

pharmacokinetics field and to evaluate their knowledge in this filed. Finally, participants were asked to add any questions they thought that was missing in the developed tool.

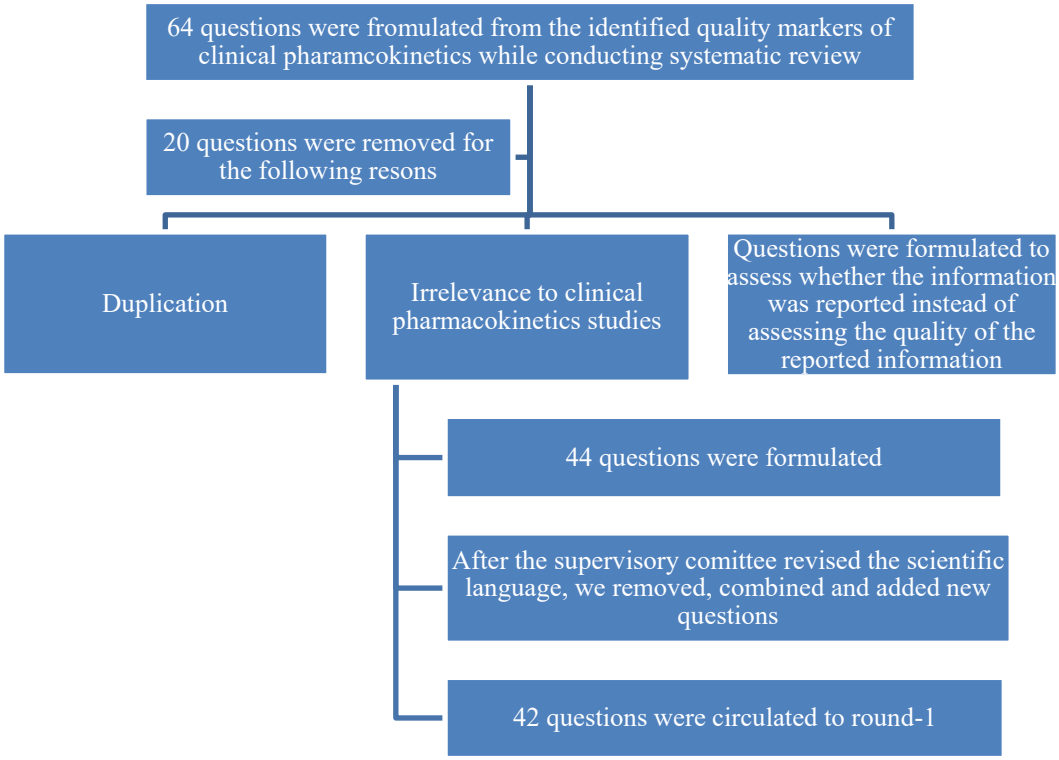

Figure 5. Flowchart of questions formulation before the initiation of the modified Delphi process

Table 11: The formulated questions from the systematic review

Questions	5 Points Likert-scale				
Title					
1- Was the title concisely reflected as the discussed topic in the paper? The title should reflect the name of the analyzed medication, and comparator (if applicable), the targeted patient population, and the study design.	1	2	3	4	5
Abstract					
2- Was a summary of the article provided to the readers within the abstract? A brief description of the knowledge gap, the objectives, summary of the methods, the results of the primary objectives and the main conclusion should be provided.	1	2	3	4	5
Background					
3- Was a comprehensive introduction provided about the analyzed drug that showed the rationale behind the conduction of that study? Authors may provide information related to: • Stages of the analyzed drug development. • Known aspects of the drug's absorption, distribution, metabolism, and elimination. • Previous studies. • What will be added to the existing body of knowledge with their proposed study?	1	2	3	4	5
4- Was a clear description of the objectives of the study clearly provided? Authors should provide a clear statement of the objectives of the research to clarify the purpose and, the scope of the study, so readers will know if the study matches their interest or not	1	2	3	4	5
Methods					
5- Was the chosen study design appropriately selected and justified? Example: • Immediate release formulation, single-dose study design is recommended. • Sustained-release formulation and medication with a long half-life or high intra-patient variability, parallel study design or steady-state design are recommended. • Bioequivalence studies, crossover design is recommended with a washout period between the administered interventions.	1	2	3	4	5

Questions	5 Points Likert-scale
6- Was the method used for drug dosing appropriate and/or referenced?	1 2 3 4 5
7-Were the endpoints of the study clearly stated? The endpoints should be directly related to the objectives.	1 2 3 4 5
8- Was the eligibility criteria of participant inclusion provided? The inclusion criteria should allow authors to choose representative participants appropriately who are representative of the targeted population to answer the main study question	1 2 3 4 5
9- Were the exclusion criteria of participants provided? Exclusion criteria should not be restricted to an extent that interferes with the generalizability of the study results	1 2 3 4 5
10- Were the study setting/location clearly described? The authors should reflect on the setting and/or location, as this will help practitioners utilize the results of the research.	1 2 3 4 5
11- During the last week, how many days did you forget to take your pills?	1 2 3 4 5
12-If applicable, was the used method to generate the random allocation sequence described? Example: Simple randomization Block randomization Stratified randomization Unequal randomization	1 2 3 4 5

Questions	5 Points Likert-scale
13-If applicable, was the used method of allocation concealment described?	1 2 3 4 5
14-Were any of the participants, the investigators or the individuals who analyzed the data blinded while the study was being conducted?	1 2 3 4 5
15-Was the method of data sampling provided? Example: • Extensive data sampling is essential to select the most appropriate structural model number of compartments, first Vs second-order absorption, and lag time. • To determine the linearity of pharmacokinetics sparse data sampling is recommended. Researchers obtain these data from previously conducted studies with completed concentration-time profile e.g. phase I studies.	1 2 3 4 5
16-Was a clear description of both the intervention and if applicable the comparator provided? Medication name, dose, dose units, schedule or frequency, route of administration, starting and stopping date of administering the medication, the reason for use.	1 2 3 4 5
17-Was a clear description of the *sampling site* and the *sampling interval* (the exact times at which samples are obtained) provided and justified? Example: • Arterial sampling is preferable during frequent sampling schedule. • Arterial sampling is more representative of the delivered concentration to the effect site in case of peripheral elimination. • Arterial sampling is preferable when administering a drug that has a short duration of action or a fast onset of action. • The sampling interval should not exceed the expected half-life of the studied exponential phase (fast distribution, slow distribution and elimination). 3-4 samples are the minimum number of samples to be collected. • Bioequivalence studies: researchers should continue to collect samples until 80% of the AUC is covered.	1 2 3 4 5

Questions	5 Points Likert-scale
18-Was a description of participant follow-up clearly described? Example: Monitoring parameters (e.g. signs and symptoms of disease or side effects of the given medication, lab data, etc.) to be collected in each period should be documented.	1 2 3 4 5
19-Was the storage condition of samples clearly described? Example: • Anticoagulant • Stabilizers • Centrifugation • Temperature • Labeling	1 2 3 4 5
20-Was there a clear description of the pharmacokinetics model, its development, and justification for use? It is recommended to provide the following details about the selected modeling process: • Description of studies from which dataset was driven • Model structure Validated software for the PK analysis • Criteria for accepting valid model's parameters • Fitting procedure defined prior to the initiation of the analysis. A reasonable assumption based on which the scheme for weighting is considered to be appropriate and the transformation of data (e.g. logarithmic transformation to achieve the homoscedastic (constant) variance requirements) should be provided.	1 2 3 4 5

Questions	5 Points Likert-scale
21-Was a list of interacting covariates (demographic variables, laboratory values, co-medication, environmental factors, and disease states) provided a priori? • The covariates should be predetermined based on biological plausibility.	1 2 3 4 5
22-Was a detailed description of the used apparatus provided? Example: • Chromatography type • Detection type • Assay characteristics: mobile phase composition, gradient and flow rate, chromatographic column (packing material, dimensions), • Analytical runtime • Operating temperature and detection parameters. • Validation method: specificity, recovery, linearity and sensitivity, the stability of the assay and its reproducibility.	1 2 3 4 5
23-Were the used pharmacokinetics equations to calculate different parameters (e.g. creatinine clearance) disclosed within the article?	1 2 3 4 5
24- Was the method used to estimate the area under the curve (AUC) and the area under the first moment curve (AUMC) provided? Example: • AUC and AUCM can be estimated by using the linear trapezoidal rule in case of increasing or equal concentrations. • AUC and AUCM can be estimated by using the log-linear trapezoidal rule for decreasing concentrations.	1 2 3 4 5
25- Were the used weight metrics in the pharmacokinetics calculation and drug dosing provided?	1 2 3 4 5

Questions	5 Points Likert-scale				
26-Was the used population pharmacokinetics approach and validation method described? Example: *Population PK approach* • Standard two-stage • Naive pooling of data • Mixed-effects modeling *Basic internal methods* • Goodness-of-fit plots/diagnostic plots. • Uncertainty in parameter estimates • Model sensitivity to outliers *Advanced internal methods* • Data splitting • Bootstrap • Cross-validation • Simulations such as visual or posterior predictive checks (PPCs) • *External model evaluation* (validation dataset observations compared with model predictions).	1	2	3	4	5
28-Was the study approved by a regional Research Ethics board?	1	2	3	4	5
29-Was a detailed description or reference of the specific level of statistical significance and the sample size calculations provided before the initiation of the study to ensure adequate power for detecting differences of interest?	1	2	3	4	5

Questions	5 Points Likert-scale
30-Were the chosen statistical tests and software to perform the statistical analysis appropriate to achieve the study objectives?	1 2 3 4 5
Results 31-Was a patient flow diagram provided? Example: Number of patients who enrolled in each arm of the trial Description of withdrawals	1 2 3 4 5
32-Were the baseline characteristics of the included participants provided? All the following variables should be clearly defined for all participants': sex, race, age, weight, height, concomitant diseases, co-medication, smoking habits, covariates, the severity of illness, residual, renal function, and hepatic function.	1 2 3 4 5
33- How did you rate your performance last year in your work?	1 2 3 4 5
34-Was the method used to handle outliers during the analysis provided? • The authors should explain the physiological/study events which result in excluding the data from the analysis. • Outlying data should be included in the final population PK model and their effect on different PK parameters should be documented.	1 2 3 4 5
35-Was an appropriate measure of precision (e.g. descriptive statistics confidence interval, standard deviation, mean, median, range, interquartile range, and trimmed range) used to document the pharmacokinetics results?	1 2 3 4 5

Questions	5 Points Likert-scale				
36-Were the essential pharmacokinetics parameters required to perform dose calculations in practice setting documented? Consider: Total clearance (CL), Fraction of dose excreted unchanged in urine (fe), Volume of distribution at steady state (Vss), Volume of distribution during the terminal phase (VZ), Blood/plasma concentration ratio, Terminal half-life (t1/2 Z), Fraction of unbound drug in plasma (fu), Bioavailable fraction of dose (F), Absorption rate constant (ka), Cmin, Cmax, tmax, EC50, ke0, Hill coefficient, or gamma	1	2	3	4	5
Discussion					
37-Were the study limitations described by the authors consistent with those identified within the study?	1	2	3	4	5
38-Was the provided interpretation consistent with the displayed results?	1	2	3	4	5
39-Did the authors compare their observed results with the results of other relevant studies and if they could be generalized to the targeted population?	1	2	3	4	5
40-Were recommendations of future studies provided?	1	2	3	4	5
Conclusion					
41-Was the provided conclusion supported by the observed results?	1	2	3	4	5
The authors should provide a summary of the observed results. The authors should not provide any new information in conclusion.					
Others					
42-Did the authors disclose any funding resources?	1	2	3	4	5
43- Did the authors disclose any conflict of interest?	1	2	3	4	5

Question	5 Points Likert-scale
44-Were the baseline characteristics of the included participants provided?	1 2 3 4 5

Appendix I:
All the following variables should be clearly defined for all participants sex, race, age, weight, height, concomitant diseases, co-medication, smoking habits, covariates, the severity of illness, residual, renal function, and hepatic function. Authors should describe if `participants are taking any medications that may interact with the analyzed medication.

The Acute Dialysis Quality Initiative (ADQI) minimum reporting criteria by ADQI should be followed in case of including participants on dialysis.

a) Operational characteristics
- Membrane/dialyser/filter and area
- A measure of time actually spent on therapy
- Delivery device
- Access and blood flow
- Anticoagulation
- Replacement fluid composition and administration
- Dialysis fluid composition and administration

b) Patient characteristics
- Measure of time actually spent on therapy
- Surgical/trauma/medical/other
- Measure of severity of illness
- Cointerventions
- Integrated hemodynamic status and vasopressor treatment
- Outcomes

In case of including participants suffering from renal failure the following information should be provided:
- Cause
- Plasma creatinine concentration/creatinine clearance
- Plasma electrolytes
- Hemoglobin concentration
- Plasma protein level
- Time and the nature of last dialysis
- Existence of clinical edema

Question	5 Points Likert-scale
• Existence of peripheral neuropathy In case of including participants suffering from hepatic cirrhosis the following information should be provided: • Cause • Child's Classification (34) • Prothrombin time, platelet count • Albumin and globulin levels In case of including participants suffering from Critically ill patients the following information should be provided: • Clinical description • Apache II score • Plasma creatinine and electrolyte concentrations • Presence/absence of renal failure • Presence/absence of liver failure In case of including participants suffering from Thermal injury the following information should be provided: • Regular hematocrit • Preoperative plasma albumin and globulin level In the case of Bioequivalence studies, the following criteria should be fulfilled: • Nonsmoker healthy volunteers (males/females) with a body • weight that is ± 20% of the standard and with age between 18 to 55 years old should be enrolled as long as possible.	

Table 12: Questions disclosed to round-1

Questions	5 Points Likert-scale
Q1) Delphi Study Consent Form Consensus Development of Quality Markers for Appraising Clinical Pharmacokinetics Studies: A MODIFIED DELPHI STUDY Name of Lead Researcher: Alaa Soliman Co-Primary Investigator: Shane Pawluk Qatar University Research Ethics Approval No. QU-IRB 970-E/18 I confirm that I have read the information sheet and understand the information provided about this research project. I confirm that I can contact the team if I have any questions or concerns regarding the project. I understand that participation in this research is voluntarily and as a participant, I have the right to withdraw without justifications at any time and there will be no negative consequences. I give permission to the investigators of this research to access and use my anonymous responses during the modified Delphi process. I understand that my name will not be disclosed with any of the materials which are related to this research or identified during any of the modified Delphi processes. If you agree to participate in this research project, please select the "Agree" box below to continue. Further information: If you have any questions or concerns please contact the co-primary investigator (Shane Pawluk). Contact details: Tel:+974 44035619, Fax: +974 44035551, email: shane.pawluk@qu.edu.qa	1- Agree 2- Disagree
Q2) Did the title concisely reflect the topic discussed in the paper? • The title should reflect the name of the analyzed medication, and comparator (if applicable), the targeted patient population, and the study design.	1 2 3 4 5
Q3) Was an adequate summary of the article provided to the readers within the abstract? • A brief description of the knowledge gap, the objectives, summary of the methods, the results of the primary objectives and the main conclusion should be provided.	1 2 3 4 5

Questions	5 Points Likert-scale
Q4) Was a comprehensive introduction provided that explained the rationale behind the conduction of the study? Authors may provide information related to: • Stages of the analyzed drug development. • Known aspects of the drug's absorption, distribution, metabolism and elimination. • Previous studies. • What will be added to the existing body of knowledge with their proposed study.	1 2 3 4 5
Q5) Was a clear description of the objectives of the study provided? Authors should provide a clear statement of the objectives of the research to clarify the purpose and the scope of the study.	1 2 3 4 5
Q6) Was the chosen study design appropriately selected and justified? Example: • Immediate release formulation, single dose study design is often recommended. • Sustained release formulation and medication with a long half-life or high intra-patient variability, parallel study design or steady-state design are often recommended. • Bioequivalence studies, crossover design is often recommended with a washout period between the administered interventions.	1 2 3 4 5
Q7) Was the dosing of the drug in the study justified and/or referenced for the intended study?	1 2 3 4 5
Q8) Were the endpoints of the study clearly relevant for the intended use of the drug? • The endpoints should be directly related to the objectives.	1 2 3 4 5
Q9) Did the eligibility criteria of participant inclusion reflect the population of interest for which the drug is intended for use? • The inclusion criteria should allow authors to choose study participants appropriately who are representative of the targeted population to answer the main study question.	1 2 3 4 5
Q10) Were the exclusion criteria of participants appropriate for the intended outcomes of the study? • Exclusion criteria should not be restricted to an extent that interferes with the generalizability of the study results.	1 2 3 4 5

Questions	5 Points Likert-scale
Q11) Was the study setting/location relevant to where the drug would be used? • Authors should reflect on the setting and/or location of the study.	1 2 3 4 5
Q12) Was the blinding of the study participants, the investigators and/or those analyzing the data appropriate while the study was being conducted?	1 2 3 4 5
Q13) Was the method of data sampling appropriate for the study? Example: • Extensive data sampling is essential to select the most appropriate structural model number of compartments, first vs second order absorption, and lag time. • To determine the linearity of pharmacokinetics, sparse data sampling is recommended. Researchers obtain these data from previously conducted studies with completed concentration-time profile e.g. phase I studies.	1 2 3 4 5
Q14) Was a clear description of the sampling site and the sampling interval (the exact times at which samples are obtained) provided and justified? Example: • Arterial sampling is preferable during frequent sampling schedule. • Arterial sampling is more representative of the delivered concentration to the effect site in the case of peripheral elimination. • Arterial sampling is preferable when administering a drug that has a short duration of action or fast onset of action. • Sampling interval should not exceed the expected half-life of the studied exponential phase (fast distribution, slow distribution and elimination)	1 2 3 4 5
Q15) Accurate participant follow-up was clearly described and rationalized? Example: • Monitoring parameters (e.g. signs and symptoms of disease or side effects of the given medication, lab data, etc.) to be collected in each period should be documented.	1 2 3 4 5

Questions	5 Points Likert-scale
Q16) Were sample storage conditions described in a manner that could be accurately replicated? Example, use of: • Anticoagulants • Stabilizers • Centrifugation • Temperature	1 2 3 4 5
Q17) Was there a clear description of the pharmacokinetics model, its development, and justification for use? • It is recommended to provide the following details about the selected modeling process: • Description of studies from which dataset was driven • Model structure • Validated software for the pharmacokinetics analysis • Criteria for accepting valid model's parameters • Fitting procedure defined prior to the initiation of the analysis. A reasonable assumption based on which the scheme for weighting is considered to be appropriate and the transformation of data (e.g. logarithmic transformation to achieve the homoscedastic (constant) variance requirements) should be provided.	1 2 3 4 5
Q18) Were plausible interacting covariates (demographic variables, laboratory values, co-medication, environmental factors and disease states) described a priori?	1 2 3 4 5
Q19) Was the description of the used apparatus for analysis adequate? Example: • Chromatography type. • Detection type. • Assay characteristics: mobile phase composition, gradient and flow rate, chromatographic column (packing material, dimensions). • Analytical runtime. • Operating temperature and detection parameters. • Validation method: specificity, recovery, linearity and sensitivity, the stability of the assay and its reproducibility.	1 2 3 4 5

Questions	5 Points Likert-scale
Q20) Were the pharmacokinetics equations used to calculate different patient parameters (e.g. creatinine clearance) disclosed within the article?	1 2 3 4 5
Q21) Was the described population pharmacokinetics approach and validation method appropriate for the analysis? Example: • Population pharmacokinetics approach • Standard two-stage • Naive pooling of data • Mixed-effects modeling • Basic internal methods • Goodness-of-fit plots/diagnostic plots • Uncertainty in parameter estimates • Model sensitivity to outliers • Advanced internal methods • Data splitting • Bootstrap • Cross validation • Simulations such as visual or posterior predictive checks (PPCs) • External model evaluation (validation dataset observations compared with model predictions).	1 2 3 4 5

Questions	5 Points Likert-scale
Q22) Did the authors justify the selection of the key models at different stages of the development process? • Justification of key models' selection at different stages of the development process through using Goodness-of-fit (GOF) plots: • Predicted data versus observed data (PRED versus DV; a line of identity and a trendline should be included). • PRED versus weighted residuals (WRES; zero line and a trend line should be included). • Time versus WRES (a zero line and a trend line should be included). Time can be both time after dose and continuous time (time in the study).	1 2 3 4 5
Q23) Was the approval number provided by a regional Research Ethics Board stated?	1 2 3 4 5
Q24) The time duration from study submission to publication was less than 1 year.	1 2 3 4 5
Q25) Was a detailed description or reference of the specific level of statistical significance and the sample size calculations provided before the initiation of the study to ensure adequate power for detecting differences of interest?	1 2 3 4 5
Q26) Were the chosen statistical tests and software to perform the statistical analysis appropriate to achieve the study objectives?	1 2 3 4 5
Q27) Was a patient flow diagram detailed to fully understand patient logistics? Example: • Number of patients who enrolled in each arm of the trial • Description of withdrawals	1 2 3 4 5
Q28) Were the baseline characteristics of the included participants representative of the population of interest? • The following variables should be clearly defined for all participants': sex, race, age, weight, height, concomitant diseases, co-medication, smoking habits, severity of illness, renal function, and hepatic function.	1 2 3 4 5

Questions	5 Points Likert-scale
Q29) In the event of missing data or outliers, was the process for analysis clearly justified?	1 2 3 4 5
Q30) Was an appropriate measure of precision (e.g. descriptive statistics confidence interval, standard deviation, mean, median, range, interquartile range, and trimmed range) used to document the pharmacokinetics results?	1 2 3 4 5
Q31) Were the study limitations described by the authors consistent with those identified within the study?	1 2 3 4 5
Q32) Were the author interpretations of the data consistent with the reported results?	1 2 3 4 5
Q33) Did the authors compare their observed results with the results of other relevant studies?	1 2 3 4 5
Q34) Were recommendations of future studies justified based on the results of this study?	1 2 3 4 5
Q35) Were the provided conclusions supported by the observed results? Authors should not provide any new information in the conclusion.	1 2 3 4 5
Q36) The authors referenced at least 10 other studies in order to defend their conclusions?	1 2 3 4 5
Q37) Were reported funding resources likely to influence the results of the study?	1 2 3 4 5
Q38) Were disclosed conflicts of interest likely to influence the results of the study?	1 2 3 4 5
Q39) In the final version of this proposed tool, which rating system do you feel would be the best method to help potential users appraise a clinical pharmacokinetics study?	1 2 3 4 5

Questions	5 Points Likert-scale
Q40) Do you think any other questions should be added to this appraisal tool that would help users appraise?	1 2 3 4 5
Demographics	
41) How many years of experience do you have in the field of clinical pharmacokinetics as a researcher or clinician?	1 2 3 4 5
42) In your personal opinion, please rate your overall knowledge in the field of clinical pharmacokinetics?	1 2 3 4 5

4.2.2. Sociodemographic characteristics

Twenty-five participants were identified and invited to participate in the modified Delphi process. Of these, 25 participants agreed to participate through sending emails replying to our invitation to participate in the modified Delphi process. The average level of experience of the participants was 8.5 years in the field of clinical pharmacokinetics. There was a high representation from the clinical sector and an even representation from academic and industrial areas, while few participants were from the regulatory sector (Table 13). The majority of panelists were practicing in Canada 52% (13/25) while there was even distribution from participants practicing in the United States of America and Qatar. Most of the participants (only 21 out of 25 participants rate their knowledge in clinical pharmacokinetics) had average (14/21) to advance (6/21) knowledge in the field of clinical pharmacokinetics.

Table 13. Sociodemographic characteristics of modified Delphi panelists

Variables	Percentage (Actual number)
Filed of experience	
Clinicians	56% (14/25)
Academic sector	20% (5/25)
Industrial sector	16% (4/25)
Regulatory sector	4% (1/25)
Project director	4% (1/25)
Geographical distribution	
Canada	52% (13/25)
USA	24% (6/25)
Qatar	24% (6/25)

4.2.3. Consensus through modified Delphi rounds

Over a period of two months (October 8, 2018, to December 1, 2018), the modified Delphi process was conducted to determine experts' consensus regarding the identified clinical pharmacokinetics quality markers.

4.2.3.1. Consensus through round-1

In the modified Delphi process, 25 surveys were received but a duplication occurred so 24/25 (96%) experts responded to round 1 survey, as presented in Figures 6 and 9. It was found that 3 participants agreed on including the control questions in the final tool. Thus, we conclude that these three participants agreed with every question without reading them. Therefore, 3 participants were removed from the total number of participants who agreed with the inclusion of the questions in the final draft of the developed tool. The percentage of agreement was calculated twice for questions that met consensus after removing 4 participants (4 participants include 3 participants who agreed with the inclusion of each question + one participant who answered the survey twice). Of 42 potential clinical pharmacokinetics critical appraisal tool items, 12 items reached ≥80 % agreement, median ≥4, and interquartile range ≤1 consensus from experts for inclusion and were retained and simply modified based on the received recommendations (Table 14). These 12 questions were not affected by the duplication or even the 3 participants who agreed on every question because the percentage of agreement was calculated again after removing 4 participants from the total number of people who agreed on including the question and percentage of the agreement were not below 75%. None of the items reached < 80%, median ≤2, and interquartile range >1 consensus from experts for exclusion. Fourteen items reached <80% consensus from experts to be added to the final tool and were reformulated based on the panelists' comments and resent for round 2. Another 9 items were reformulated and

sent to round 2 despite achieving ≥80%. We chose to resend these questions due to the polarizing comments relating to the questions' importance in assessing quality (Table 15). Although the two control questions did not reach a consensus from experts for inclusion or exclusion, both were excluded, as they did not have any relevance in assessing the quality of a clinical pharmacokinetics study (Table 16). Furthermore, the two demographic questions were also removed. Two items were added based on the received recommendations/comments from the panelists. An appendix was added to clarify baseline characteristics in one of the questions, so the whole question recirculated to round-2 although it met consensus, Appendix E. Although question 21 achieved expert consensus, one of the panelists commented this question was doubled barreled and recommended to split the question, thus it was recirculated to round-2. Furthermore, panelists were asked to select the best rating scale to help end-users use the developed clinical pharmacokinetics critical appraisal tool. The consent form was sent again in round-2. We asked the panelists if they want to recommend the addition of any question, they thought that it was missing in the developed tool.

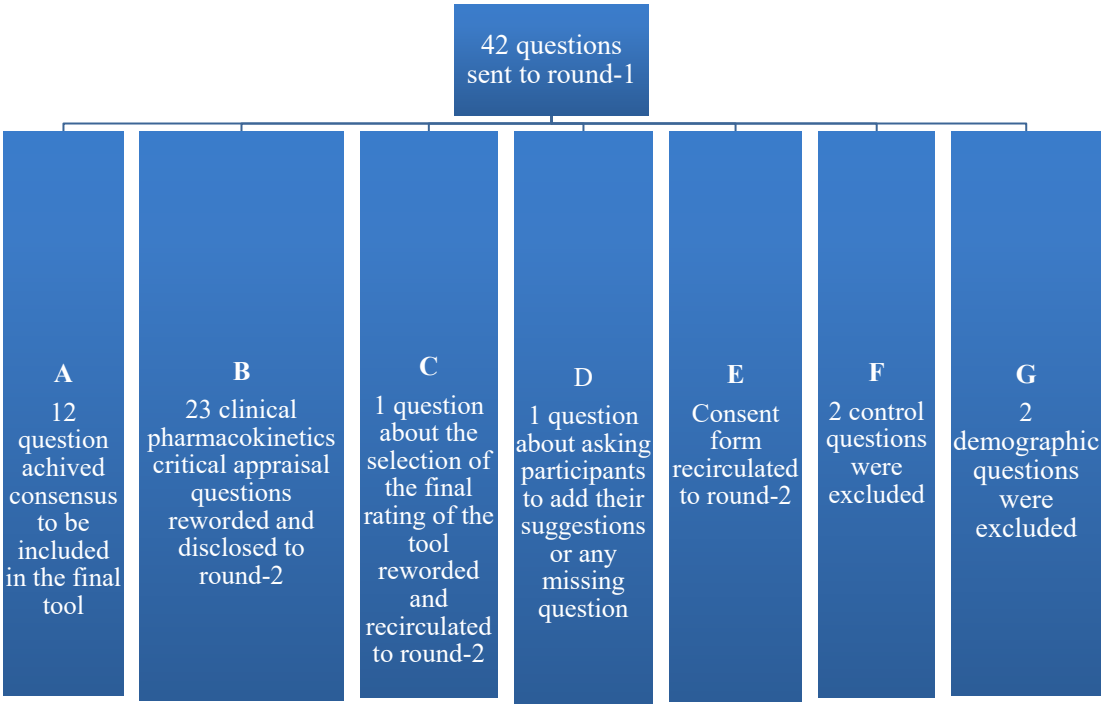

Figure 6. Flowchart of round-1 questions

Table 14: Agreement through round-1

Question	Modified question	Consensus			
		Percentage of agreement	Median	Interquartile range	Percentage of an agreement after removing 4 participants
Q5) Was a clear description of the objectives of the study provided? Authors should provide a clear statement of the objectives of the research to clarify the purpose and the scope of the study.		Agree: (21) 100% Neither Agree Nor Disagree: (0) 0% Disagree: (0) 0%	5	0	Agree: (17) 100% Neither Agree Nor Disagree: (0) 0% Disagree: (0) 0%
Q6) Was the chosen study design appropriately selected and justified? Example: Immediate release formulation, single dose study design is often recommended. Sustained release formulation and medication with a long half-life or high intra-patient variability, parallel study design or steady-state design are often recommended. Bioequivalence studies, crossover design is often recommended with a washout period between the administered interventions	Q6) Was the chosen study design appropriately selected and justified?	Agree: (20) 95.24% Neither Agree Nor Disagree: (0) 0% Disagree: (1) 4.76%	5	1	Agree: (16) 94.11% Neither Agree Nor Disagree: (0) 0% Disagree: (1) 5.88%
Q7) Was the dosing of the drug in the study justified and/or referenced for the intended study?	Q7) Was the dosing (dose, route of administration, dosing interval) of the drug in the study justified for the intended study? Example: Authors should justify the use of single-dose versus steady-state analysis	Agree: (19) 90.48% Neither Agree Nor Disagree: (1) 4.76% Disagree: (1) 4.76%	4	1	Agree: (15) 88.23% Neither Agree Nor Disagree: (1) 5.88% Disagree: (1) 5.88%

Question	Modified question	Consensus			
		Percentage of agreement	Median	Interquartile range	Percentage of an agreement after removing 4 participants
Q8) Were the endpoints of the study clearly relevant for the intended use of the drug? The endpoints should be directly related to the objectives.	Q8) If applicable, 'Were the study endpoints reflective of what the stated objectives were?' OR "Were the endpoints of the study appropriate to answer the objectives of the study	Agree: (17) 80.95% Neither Agree Nor Disagree: (2) 9.52% Disagree: (2) 9.52%	4	1	Agree: (13) 76.47% Neither Agree Nor Disagree: (2) 11.76% Disagree: (2) 11.76%
Q10) Were the exclusion criteria of participants appropriate for the intended outcomes of the study? Exclusion criteria should not be restricted to an extent that interferes with the generalizability of the study results.	Q10) Were the exclusion criteria of participants appropriate for the intended outcomes of the study? The exclusion criteria should be relevant to assist with decreasing significant confounders (e.g. co-administration of drugs, organ impairment, special populations) that may impact the ability to achieve the study objectives	Agree: (18) 85.71% Neither Agree Nor Disagree: (0) 0% Disagree: (3) 14.28%	4	1	Agree: (14) 82.35% Neither Agree Nor Disagree: (0) 0% Disagree: (3) 17.64%

Question	Modified question	Consensus			
		Percentage of agreement	Median	Interquartile range	Percentage of an agreement after removing 4 participants
Q13) Was the method of data sampling appropriate for the study? Example: Extensive data sampling is essential to select the most appropriate structural model number of compartments, first vs second order absorption, and lag time. To determine the linearity of pharmacokinetics, sparse data sampling is recommended. Researchers obtain these data from previously conducted studies with completed concentration-time profile e.g. phase I studies.	Q13) Was the method of data sampling appropriate for the study? Examples: first vs second order absorption, and lag time. Evaluating for nonlinearity requires multiple dose levels and a complete profile is recommended. Researchers obtain these data from previously conducted studies with completed concentration-time profile e.g. phase I studies. The method of data sampling should reference previously validated quantitative bioanalytical methods and if those are not available then the full description or defense of data sampling should be included	Agree: (19) 90.47% Neither Agree Nor Disagree: (1) 4.76% Disagree: (1) 4.76%	5	1	Agree: (15) 88.23% Neither Agree Nor Disagree: (1) 5.88% Disagree: (1) 5.88%

Question	Modified question	Consensus			
		Percentage of agreement	Median	Interquartile range	Percentage of an agreement after removing 4 participants
Q14) Was a clear description of the sampling site and the sampling interval (the exact times at which samples are obtained) provided and justified? Example: Arterial sampling is preferable during frequent sampling schedule. Arterial sampling is more representative of the delivered concentration to the effect site in the case of peripheral elimination. Arterial sampling is preferable when administering a drug that has a short duration of action or fast onset of action. Sampling interval should not exceed the expected half-life of the studied exponential phase (fast distribution, slow distribution and elimination).	Q14) Was a clear description of the sampling site and the sampling interval (the exact times at which samples are obtained) provided and justified? Example: Sampling site should be consistent for all subjects in the study. Arterial sampling is preferable during frequent sampling schedule. Arterial sampling is more representative of the delivered concentration to the effect site in the case of peripheral elimination. Arterial sampling is preferable when administering a drug that has a short duration of action or fast onset of action. Sampling interval should not exceed the expected half-life of the studied exponential phase (fast distribution, slow distribution and elimination).	Agree: (19) 90.47% Neither Agree Nor Disagree: (1) 4.76% Disagree: (1) 4.76%	5	1	Agree: (15) 88.23% Neither Agree Nor Disagree: (1) 5.88% Disagree: (1) 5.88%

Question	Modified question	Consensus			
		Percentage of agreement	Median	Interquartile range	Percentage of an agreement after removing 4 participants
Q16) Were sample storage conditions described in a manner that could be accurately replicated? Example, use of: Anticoagulants Stabilizers Centrifugation Temperature	Q16) Were sample storage conditions appropriate and described in a manner that could be accurately replicated? Examples: sample storage temperature, use and description of anticoagulants, stabilizers, centrifugation	Agree: (20) 95.24% Neither Agree Nor Disagree: (1) 4.76% Disagree: (1) 0%	5	1	Agree: (16) 88.88% Neither Agree Nor Disagree: (1) 5.56% Disagree: (1) 5.56%
Q17) Was there a clear description of the pharmacokinetics model, its development, and justification for use? It is recommended to provide the following details about the selected modeling process: Description of studies from which dataset was driven Model structure Validated software for the pharmacokinetics analysis Criteria for accepting valid model's parameters Fitting procedure defined prior to the initiation of the analysis. A reasonable assumption based on which the scheme for weighting is considered to be appropriate and the transformation of data (e.g. logarithmic transformation to achieve the	Q17) If applicable, was there a clear description of the pharmacokinetics model, its development, validation and justification for use? It is recommended to provide the following details about the selected modeling process: Description of studies from which dataset was driven	Agree: (20) 95.23% Neither Agree Nor Disagree: (0) 0% Disagree: (1) 4.76%	5	1	Agree: (16) 94.11% Neither Agree Nor Disagree: (0) 0% Disagree: (1) 5.88%

Question	Modified question	Consensus			
		Percentage of agreement	Median	Interquartile range	Percentage of an agreement after removing 4 participants
homoscedastic (constant) variance requirements) should be provided.	Model structure Validated software for the pharmacokinetics analysis Criteria for accepting valid model's parameters Fitting procedure defined prior to the initiation of the analysis. A reasonable assumption based on which the scheme for weighting is considered to be appropriate and the transformation of data (e.g. logarithmic transformation to achieve the homoscedastic (constant) variance requirements) should be provided.				
Q26) Were the chosen statistical tests and software to perform the statistical analysis appropriate to achieve the study objectives?	Q26) Were the statistical tests and software to perform the PK analysis appropriate based on the study objectives	Agree: (19) 90.47% Neither Agree Nor Disagree: (1) 4.76% Disagree: (1) 4.76%	5	1	Agree: (15) 88.23% Neither Agree Nor Disagree: (1) 5.88% Disagree: (1) 5.88%
Q29) In the event of missing data or outliers, was the process for analysis clearly justified?	Q29) In the event of missing data or outliers, was the process for analysis justified and appropriate?	Agree: (19) 90.47% Neither Agree Nor Disagree: (2) 9.52% Disagree: (0) 0%	4	1	Agree: (15) 88.23% Neither Agree Nor Disagree: (2) 11.76% Disagree: (0) 0%

Question	Modified question	Consensus			
		Percentage of agreement	Median	Interquartile range	Percentage of an agreement after removing 4 participants
Q30) Was an appropriate measure of precision (e.g. descriptive statistics confidence interval, standard deviation, mean, median, range, interquartile range, and trimmed range) used to document the pharmacokinetics results?	Q30) Were appropriate summary statistics to describe centrality and variance used to document the pharmacokinetics results? Example: Descriptive statistics confidence interval, standard deviation, mean, median, range, interquartile range, standard error and trimmed rang	Agree: (19) 90.47% Neither Agree Nor Disagree: (1) 4.76% Disagree: (1) 4.76%	5	1	Agree: (15) 88.23% Neither Agree Nor Disagree: (1) 5.88% Disagree: (1) 5.88%

Table 15: Questions that did not meet consensus through round-1

Question	Modified question	Consensus before removing 3 participants			Consensus after removing 3 participants	Panelist's Comments
		Percentage of agreement	Median	Interquartile range	Percentage of agreement	
Q2)* Did the title concisely reflect the topic discussed in the paper? The title should reflect the name of the analyzed medication, and comparator (if applicable), the targeted patient population, and the study design.	Q2) Did the title concisely reflect the topic discussed in the paper? The title should reflect the name of the analyzed medication, and comparator (if applicable), the targeted patient population, and the study rationale or design.	Agree: (21) 87.5 % Neither Agree Nor Disagree: (0) 0% Disagree: (3) 12.5%	5	1		• I am unclear what the title is. Is it "PK Appraisal Tool" or "Appraisal of Clinical Pharmacokinetics Studies." Is it clinical only or more basic pharmacokinetics studies too? If Clinical only, I would suggest something like "Checklist for Appraising Clinical Pharmacokinetics Studies." • I strongly agree that the experimental arm and comparator need to be disclosed along with the patient population. However, I do not always find that the study design is necessary to include in the title. • Suggest changing to The title should reflect the name of the analyzed medication, and comparator (if applicable), the targeted patient population, and the study rationale or design. • The title should reflect the main conclusion of the paper, not necessarily the study design. • This element is important for reporting checklists/guidelines; however, it cannot be used to

Question	Modified question	Consensus before removing 3 participants			Consensus after removing 3 participants	Panelist's Comments
		Percentage of agreement	Median	Interquartile range	Percentage of agreement	
						• evaluate the methodological validity of the study (ie I do not believe assessing this formally is useful for critical appraisal). • This is an important element for a reporting checklist; however, it is not important for critical appraisal.
Q3) Was an adequate summary of the article provided to the readers within the abstract? A brief description of the knowledge gap, the objectives, summary of the methods, the results of the primary objectives and the main conclusion should be provided.	Q3) Was an adequate summary of the article provided to the readers within the abstract? A brief description of the knowledge gap, the objectives, summary of the methods (patient population studied, the route of administration of the drug(s) studied), the results of the primary objectives, as well as, secondary or exploratory outcomes and the main conclusion should be provided. Summary of comments from round-1 a.This item is essential to	Agree: (16) 69.56% Neither Agree Nor Disagree: (2) 8.70% Disagree: (5) 21.74%	5	2		• Strongly disagree with this point. The abstract does not impact the quality of a PK study. It is important in the reporting of a PK study (e.g. like CONSORT) but again does not influence if it is at high risk of bias or not. • I believe you're referring to the email sent October 8th? I am unclear if there are already other checklists in existence and the feeling is they are inadequate? Or is there a lack of checklists in this area and you are developing a novel checklist? That would be helpful to include in the summary of the article. • I am not sure how this will help to evaluate the article in full. It is useful to have a good abstract, but I am not sure if this is necessary.

Question	Modified question	Consensus before removing 3 participants			Consensus after removing 3 participants	Panelist's Comments
		Percentage of agreement	Median	Interquartile range	Percentage of agreement	
	assess the quality of reporting of clinical pharmacokinetics studies but does not influence the quality of a study and is therefore not necessary. b. It is essential to have an excellent abstract, but there is a doubt of how this will help in assessing the quality of the whole study.					• Another abstract inclusion which is often overlooked is the results generated from secondary or exploratory outcomes. Not only do these provide some insights into the overall design of the study (by way of outcome variables), but also an idea of how the study was powered. • I believe the abstract should also include the patient population studied, the route of administration of the drug(s) • studied, the main clinical pharmacokinetics findings (which I would assume is the primary objective results - but perhaps we should be explicit). • This element is important for reporting checklists/guidelines; however, it cannot be used to evaluate the methodological validity of the study (ie I do not believe assessing this formally is useful for critical appraisal). • As with the title, this is an important element for a reporting checklist, but

Question	Modified question	Consensus before removing 3 participants			Consensus after removing 3 participants	Panelist's Comments
		Percentage of agreement	Median	Interquartile range	Percentage of agreement	
						not for an appraisal checklist
Q4) Was a comprehensive introduction provided that explained the rationale behind the conduction of the study? Authors may provide information related to: Stages of the analyzed drug development. Known aspects of the drug's absorption, distribution, metabolism	Q4) Was a comprehensive introduction provided that explained the background and rationale behind the conduction of the study? Authors may provide information related to: the analyzed drug indication (proposed or labelled) including dosing and patient population. Stages of the analyzed drug development. Known aspects of the drug's absorption, distribution, metabolism and elimination. Previous studies knowledge gaps, and addressing how	Agree: (16) 76.19% Neither Agree Nor Disagree: (1) 4.76% Disagree: (4) 19.05%	4	1	Agree: (13) 72.22% Neither Agree Nor Disagree: (1) 5.56% Disagree: (4) %	• Again, does not influence bias but is part of a well written research paper. • Please see previous response. I did not feel there was adequate background provided to justify this study based on what I received. • May consider explicitly stating "knowledge gaps" • It is also important to describe the clinical questions or why the investigation is required in the first place but I suppose this is part of the last section of "what • will be added to the existing body of knowledge" • Stages of drug development are not always necessary, but rather a dialogue on the drugs indication (proposed or labelled) including dosing and patient population. What is most important is identifying knowledge gaps, and addressing how those knowledge gaps will be

Question	Modified question	Consensus before removing 3 participants			Consensus after removing 3 participants	Panelist's Comments
		Percentage of agreement	Median	Interquartile range	Percentage of agreement	
and elimination. Previous studies. What will be added to the existing body of knowledge with their proposed study.	those knowledge gaps will be closed, or brought closer through the experimental design presented in the current manuscript. What will be added to the existing body of knowledge with their proposed study. Summary of comments from round -1: a. This item is used to assess how a research paper is well written but not to evaluate the quality of the study. b. This item is not appropriate for an appraisal tool that aims to assess the quality of the trial; it is only essential for reporting checklist. c. This item is used to assess how a					closed, or brought closer through the experimental design presented in the current manuscript. • Suggest changing to: 4. Was a comprehensive introduction provided that explained the background and rationale behind the conduction of the study? Previous in vitro and in vivo studies. Gaps in current understanding of the area. • Often more detailed information about previous studies goes in the discussion and the qualifier "previous studies" is too vague. Unclear what is meant by "stages of the analyzed drug development" • Could also include the main hypothesis or aim of the current study.

Question	Modified question	Consensus before removing 3 participants			Consensus after removing 3 participants	Panelist's Comments
		Percentage of agreement	Median	Interquartile range	Percentage of agreement	
	research paper is well written but not to evaluate the quality of the study. d. This item is not appropriate for an appraisal tool that aims to assess the quality of the trial; it is only essential for reporting checklist.					• This element is important for reporting checklists/guidelines; however, it cannot be used to evaluate the methodological validity of the study. Although a poorly-written introduction generally provides a clue that the rest of the paper will be unclearly reported, it is not in-and-of-itself useful to appraise the methodology of the present study? Previous in vitro and in vivo studies. Gaps in current understanding of the area. • Often more detailed information about previous studies goes in the discussion and the qualifier "previous studies" is too vague. Unclear what is meant by "stages of the analyzed drug development" • Could also include the main hypothesis or aim of the current study. • This element is important for reporting checklists/guidelines; however, it cannot be used to

Question	Modified question	Consensus before removing 3 participants			Consensus after removing 3 participants	Panelist's Comments
		Percentage of agreement	Median	Interquartile range	Percentage of agreement	
						evaluate the methodological validity of the study. Although a poorly-written introduction
Q9) Did the eligibility criteria of participant inclusion reflect the population of interest for which the drug is intended for use? The inclusion criteria should allow authors to choose study participants appropriately who are representative of the targeted population to answer the main study question.	Q9) If applicable, did the eligibility criteria of participant inclusion reflect the population of interest for which the drug is intended for use? The inclusion criteria should represent the targeted population to answer the main study question. If renal or hepatic function impacts the PK, patients who have varying degrees of renal insufficiency should be included.	Agree: (16) 76.19% Neither Agree Nor Disagree: (2) 9.52% Disagree: (3) 14.28%	4	1	Agree: (13) 72.22% Neither Agree Nor Disagree: (2) 11.11% Disagree: (3) 16.67%	• Generally provides a clue that the rest of the paper will be unclearly reported, it is not in-and-of-itself useful to appraise the methodology of the present study. • This item is not important for every clinical pharmacokinetics study. For example, in a drug-drug interaction study, the medication is given to a patient who might not be receiving it in the clinical setting. However, in cases in which renal or hepatic function might affect the pharmacokinetics parameters, patients suffering from hepatic or renal insufficiency should be included. If the pharmacokinetics parameters are not affected by the diseases, the study could be conducted on healthy volunteers then the results extrapolated to patients in clinical setting

Question	Modified question	Consensus before removing 3 participants			Consensus after removing 3 participants	Panelist's Comments
		Percentage of agreement	Median	Interquartile range	Percentage of agreement	
Q11) Was the study setting/location relevant to where the drug would be used? Authors should reflect on the setting and/or location of the study.	Q11) Was the study setting/location relevant to where the drug would be used? Authors should describe the setting and/or location of the study. They should also follow practice guideline recommendations to ensure that there is an understanding of where/how the drug is used. Example: critically ill patients, as changes in physiology may alter drug disposition.	Agree: (11) 52.38% Neither Agree Nor Disagree: (4) 19.05% Disagree: (6) 28.57%	4	2		• This may not be relevant to all clinical pharmacokinetics studies. • This is relevant when the drug disposition is altered due to changes in the physiology, such as in critically ill patients. • The location is important for generalizability of the results, but the quality of the research should not be affected if its methodology was comprehensive and robust
Q12) Was the blinding of the study participants, the investigators and/or those analyzing the data appropriate while the study was being conducted?	Q12) If applicable, was the blinding of the study participants, the investigators and/or those analyzing the data appropriate while the study was being conducted?	Agree: (14) 66.67% Neither Agree Nor Disagree: (2) 9.52% Disagree: (5) 23.80%	4	1		• In most pharmacokinetics studies, blinding is not essential for participants, investigators and data analyzers because they measure objective endpoints like drug concentration. The outcome of interest is the only factor that defines the importance of blinding. For example: when dose-response is investigated, blinding is important to avoid subjective influence.

Question	Modified question	Consensus before removing 3 participants			Consensus after removing 3 participants	Panelist's comments
		Percentage of agreement	Median	Interquartile range	Percetage of agreement	
Q15) *Accurate participant follow-up was clearly described and rationalized? Example: Monitoring parameters (e.g. signs and symptoms of disease or side effects of the given medication, lab data, etc.) to be collected in each period should be documented	Q15) Was participant follow-up clearly described and rationalized? Example: Monitoring parameters (e.g. signs and symptoms of disease or side effects of the given medication, lab data, etc.) to be collected in each period should be documented	Agree: (17) 80.95% Neither Agree Nor Disagree: (0) 0% Disagree: (2) 9.52% Comments: (2) 9.52%	5	1		• Participant follow up is not required in many clinical pharmacokinetics studies.
Q18) *Were plausible interacting covariates (demographic variables, laboratory values, co-medication, environmental factors and disease states) described a priori?	Q18) Were plausible interacting covariates described a priori or in Post hoc evaluation? Examples: Demographic variables, laboratory values, concomitant medications, and relevant disease states to the drug being studied.	Agree: (18) 85.71% Neither Agree Nor Disagree: (2) 9.52% Disagree: (1) 4.76%	4	1		• This information should be mentioned in the inclusion and exclusion criteria. • This only applies if a covariate analysis is included in the clinical pharmacokinetics analysis. • It would be helpful if some of these factors were acknowledged as it is difficult to describe all of these covariates a priori.

Question	Modified question	Consensus before removing 3 participants			Consensus after removing 3 participants	Panelist's comments
		Percentage of agreement	Median	Interquartile range	Percetage of agreement	
Q19) *Was the description of the used apparatus for analysis adequate? Example: Chromatography type. Detection type. Assay characteristics: mobile phase composition, gradient and flow rate, chromatographic column (packing material, dimensions). Analytical runtime. Operating temperature and detection parameters. Validation method: specificity, recovery, linearity and sensitivity, the stability of the assay and its reproducibility	Q19) Was the description of the used sample analysis methods or citations of prior validation studies provided in the publication or affiliated appendix? Example: Chromatography type. Detection type. Assay characteristics: mobile phase composition, gradient and flow rate, chromatographic column (packing material, dimensions). Analytical runtime. Operating temperature and detection parameters. Validation method: specificity, recovery, linearity and sensitivity, the stability of the assay and its reproducibility Received comments from round-1: a. This will help readers to compare the results of	Agree: (18) 85.71% Neither Agree Nor Disagree: (1) 4.76% Disagree: (2) 9.52%	5	1	Agree: (15) 83.33% Neither Agree Nor Disagree: (1) 5.56% Disagree: (2) 11.11%	• This is essential - many articles do not mention this and the reader is left to wonder how they reached a certain number • Move the examples to below the question like the other categories. Examples: creatinine clearance, body weight calculations, Michaelis Menten, Volume of distribution • I am not sure how valuable this would actually be for an appraisal. I see how being comprehensive/reproducibility is important but if the equation wasn't disclosed, it may not reduce the value of the article. • Should clarify what is meant by "different patient parameters" - PK values, estimations of GFR?

Question	Modified question	Consensus before removing 3 participants			Consensus after removing 3 participants	Panelist's comments
		Percentage of agreement	Median	Interquartile range	Percetage of agreement	
	different studies. b. This will be beyond the level of the reader's knowledge to judge if the used sample analysis methods are appropriate or not. c. A brief description of the assay characteristics and validation method is essential to be included in a paper. The details should be available in an online appendix or contacting the corresponding author					• You don't need to disclose the equation, you can cite it if applicable (CG, MDRD, CKD, etc...) • Along with the specified patient weight used for these calculations - total body weight vs. ideal body weight etc
Q20) *Were the pharmacokinetics equations used to calculate different patient parameters (e.g. creatinine clearance) disclosed within the article?	Q20) Were the pharmacokinetics equations used to calculate patient pharmacokinetics parameters disclosed or cited within the article? Examples: creatinine clearance, body weight calculations, Michaelis Menten, volume of distribution, patient weight: total body weight vs. ideal body weight	Agree: (18) 85.71% Neither Agree Nor Disagree: (1) 4.76% Disagree: (2) 9.52%	5	1		• Received comments from round-1: This is important item since pharmacokinetics equations are not disclosed in a lot of articles. • Disclosing equations is very important for reproducibility but will not reduce the quality of an article if they are not disclosed.

Question	Modified question	Consensus before removing 3 participants			Consensus after removing 3 participants	Panelist's comments
		Percentage of agreement	Median	Interquartile range	Percetage of agreement	
Q21) *Were the described population pharmacokinetics approach and validation method appropriate for the analysis? Example: Population pharmacokinetics approach Standard two-stage Naive pooling of data Mixed-effects modeling Basic internal methods Goodness-of-fit plots/diagnostic plots Uncertainty in parameter estimates Model sensitivity to outliers Advanced internal methods Data splitting Bootstrap Cross validation Simulations such as visual or posterior	Q21) Was the described population pharmacokinetics approach validation method appropriate for the analysis? Basic internal methods • Goodness-of-fit plots/diagnostic plots. • Uncertainty in parameter estimates Model sensitivity to outliers. • Advanced internal methods • Data splitting • Bootstrap • Cross validation • Simulations such as visual or posterior • predictive checks (PPCs)	Agree: (20) 95.24% Neither Agree Nor Disagree: (1) 4.76% Disagree: (0) 0%	5	1		• Recommend an appendix to help guide users of the checklist. If we are assuming this is a general clinician using this tool, they may not know what all these terms mean. This question is double-barreled (approach AND validation; what if one was appropriate and the other was not)? Could this be split into two questions? If not, change grammar to "Were the described..." not "Was" • Not an expert on these methods myself • Recommend an appendix to help guide users of the checklist. If we are assuming this is a general clinician using this tool, they may not know what all these terms mean. This question is double-barreled (approach AND validation; what if one was appropriate and the other was not)? Could this be split into two questions? If not, change grammar to "Were the described..." not "Was" • Not an expert on these methods myself

Question	Modified question	Consensus before removing 3 participants			Consensus after removing 3 participants	Panelist's comments
		Percentage of agreement	Median	Interquartile range	Percetage of agreement	
predictive checks PPCs) External model evaluation (validation dataset observations compared with model predictions).	External model evaluation (validation dataset observations compared with model predictions). Received Comments from round 1: a. This will be beyond the level of the reader's knowledge to judge if the used population pharmacokinetics approach and validation are appropriate or not.					I think the advanced internal methods are beyond what an average clinician would be able to assess if doing an appraisal of a clinical PK study. The basic population model and internal methods may be appropriate to include.
Q22) Did the authors justify the selection of the key models at different stages of the development process? Justification of key models' selection at different stages of the development process through using Goodness-of-fit (GOF) plots: Predicted data versus		Agree: (14) 66.67% Neither Agree Nor Disagree: (6) 28.57% Disagree: (1) 4.76%	4	2		• What do you mean by "key models"? If you're referencing GOF plots only, change question to say that explicitly. Are there other "key models" you want to include? Recommend adding "Examples:" like the other questions. • Typically, on the final model should be presented and not all of the iterations evaluated. Otherwise you would present each model separately.

Question	Modified question	Consensus before removing 3 participants			Consensus after removing 3 participants	Panelist's comments
		Percentage of agreement	Median	Interquartile range	Percetage of agreement	
observed data (PRED versus DV; a line of identity and a trendline should be included).PRED versus weighted residuals (WRES;zero line and a trend line should be included).Time versus WRES (a zero line and a trend line should be included). Time can be both time after dose and continuous time (time in the study).	-					• In RCTs and observational studies, for example, the authors don't always explicitly write why it is that they chose a given model. As long as it is appropriate for the stated outcomes, it can be up to the reader to decide if it's justified. • Depending on the purpose of the paper, detailed description of all model selection steps might not be needed. • Again, this may be beyond the average clinicians ability to assess if doing a critical appraisal of clinical PK study. If having a Goodness-of-fit (GOF) plot is the key component that should be present in the paper - then perhaps that is the appraisal parameter that you want the clinician to assess.

Question	Modified question	Consensus before removing 3 participants			Consensus after removing 3 participants	Panelist's comments
		Percentage of agreement	Median	Interquartile range	Percetage of agreement	
Q23) Was the approval number provided by a regional Research Ethics Board stated?	Q23) Was appropriate Research Ethics Board approval received Received comments from round-1: a. This element is important to assess the quality of reporting but not the quality of the overall study. b. Authors should state that REB has approved the study, but there is no need explicitly mention the approval number.	Agree: (14) 66.67% Neither Agree Nor Disagree: (4) 19.04% Disagree: (3) 14.28%	4	2		• What is the approval number? • Are we including only clinical PK studies? If not clinical, they may not have needed ethics approval. Change wording to "For clinical PK studies, was information regarding ethics approval provided?" • Not sure I understand the approval number. Certainly, an expectation for approval should be included, however this could be a sub-mention in the background • Why is the approval number relevant? Would just state Was appropriate Research Ethics Board approval received? • I believe the fact that a REB has approved the study should be included - not sure that the approval number needs to be explicitly stated. • Again, this element is important for reporting checklists/guidelines, but not necessarily to critical appraisal. • Approval should be included, however this could be a sub-

Question	Modified question	Consensus before removing 3 participants			Consensus after removing 3 participants	Panelist's comments
		Percentage of agreement	Median	Interquartile range	Percetage of agreement	
						• mention in the background • Why is the approval number relevant? Would just state Was appropriate Research Ethics Board approval received? • I believe the fact that a REB has approved the study should be included - not sure that the approval number needs to be explicitly stated. • checklists/guidelines, but not necessarily to critical appraisal. • approval should be included, however this could be a sub-mention in the background • Why is the approval number relevant? Would just state Was appropriate Research Ethics Board approval received? • I believe the fact that a REB has approved the study should be included - not sure that the approval number needs to be explicitly stated. • Again, this element is important for reporting checklists/guidelines, but not necessarily to critical appraisal.

Question	Modified question	Consensus before removing 3 participants			Consensus after removing 3 participants	Panelist's comments
		Percentage of agreement	Median	Interquartile range	Percetage of agreement	
Q25) Was a detailed description or reference of the specific level of statistical significance and the sample size calculations provided before the initiation of the study to ensure adequate power for detecting differences of interest?	Q25) "Was the level of statistical significance appropriate for the intended outcomes of the study?" Received comments from round-1: a. This item is not applicable for all clinical pharmacokinetics studies as the primary endpoints in some of the studies are to describe pharmacokinetics parameters. b.This item is relevant to some studies, but not all clinical pharmacokinetics studies are powered enough to identify any difference. If the goal of the study was to detect a difference, authors should clearly state how did they calculate the sample size	Agree: (16) 76.19% Neither Agree Nor Disagree: (5) 23.81% Disagree: (0) 0%	4	1	Agree: (13) 72.22% Neither Agree Nor Disagree: (5) 27.78% Disagree: (0) 0%	• Double-barrelled question; split into two questions. Also, re-word to be more succinct, for example: "Was the level of statistical significant specified?" "Was the sample size determined a priori"? • I would also suggest adding in do you agree with the numbers used, not just was there a description • Not sure a reviewer/reader will know if this was done before initiation of the study. For some clinical PK studies which are describing the PK as a primary endpoint, it may be challenging to calculate a sample size and often sample sizes of convenience are used. Suggest rephrasing this question • This is relevant but keep in mind that not all PK studies are necessarily powered to detect a specific difference. I would just add "if appropriate" or "if applicable" • I agree if they are trying to detect a difference then they should indicate how they determined the sample size and their power. Not sure this is often the case in clinical PK studies.

Question	Modified question	Consensus before removing 3 participants			Consensus after removing 3 participants	Panelist's comments
		Percentage of agreement	Median	Interquartile range	Percetage of agreement	
						• This is very context-sensitive, as "identifying differences" may not be the goal of all PK studies, and precision rather than power may be the driving factor in selecting a sample size.
Q27) Was a patient flow diagram detailed to fully understand patient logistics? Example: Number of patients who enrolled in each arm of the trial Description of withdrawals	Q27) "Were all patients enrolled in the study accounted for?" Example: Description of patient screening, enrolment, run-in or wash-out phases, study period and follow-up periods are adequately described. Any loss to follow-up or withdrawals are described. Received comments from round-1: a.This information is important to be described either in the form of a flow diagram or within the text. b.This information should be moved to the appendix.	Agree: (15) 71.43% Neither Agree Nor Disagree: (4) 19.05% Disagree: (2) 9.52%	4	2		• I think this concept is important, but a flow diagram isn't necessarily required. Reword, something like "Were all patients enrolled in the study accounted for?" • In certain cases this is necessary, but dependent on the complexity of the design. Typically I would prefer to see a simplistic design scheme in the publication, with these specifics in the supplemental data • Note: If applicable • I agree that this information is important, but it does not necessarily need to be in a flow diagram. As long as it is described somewhere, be it in the text or depicted in a flow diagram, that would be sufficient. • Can go in appendix.

Question	Modified question	Consensus before removing 3 participants			Consensus after removing 3 participants	Panelist's comments
		Percentage of agreement	Median	Interquartile range	Percetage of agreement	
						• I don't think a patient flow diagram is necessary but a description of patient screening, enrolment, run-in or wash-out phases, study period and follow-up periods are adequately described. Any loss to follow-up or withdrawals are described. • I agree that this information is important, but it does not necessarily need to be in a flow diagram. As long as it is described somewhere, be it in the text or depicted in a flow • diagram, that would be sufficient. • Can go in appendix • I don't think a patient flow diagram is necessary but a description of patient screening, enrolment, run-in or wash-out phases, study period and follow-up periods are adequately described. Any loss to follow-up or withdrawals are described.

Question	Modified question	Consensus before removing 3 participants			Consensus after removing 3 participants	Panelist's comments
		Percentage of agreement	Median	Interquartile range	Percetage of agreement	
Q28) Were the baseline characteristics of the included participants representative of the population of interest? The following variables should be clearly defined for all participants': sex, race, age, weight, height, concomitant diseases, co-medication, smoking habits, severity of. Illness, renal function, and hepatic function.	Q28) Were the relevant baseline characteristics of the included participants reflective of the inclusion/exclusion criteria? Examples of important participant characteristics: sex, race, age, weight, height, concomitant diseases, administered medications, smoking habits, severity of illness that may affect pharmacokinetics parameters, renal function, and hepatic function Received comments from round-1: a. This information is important to be documented. However, this is not important for every pharmacokinetics study, such as bioequivalence studies that are conducted in a population who are not	Agree: (14) 66.67% Neither Agree Nor Disagree: (3) 14.28% Disagree: (4) 19.04%		4	2	• Should include all relevant known prognostic factors for the outcome of interest. • Consider changing "The following variables should be clearly defined" to "Examples of important participant characteristics:" • I would add all **relevant** baselines characteristics included • I disagree with some of the variables listed as depend on the medication. Please do not use the term "co-medication" as it means a secondary medication used to treat the side effects of another medication. Smoking habits may or may not be relevant depending on the drug, ditto for race. Severity of what illness? Not sure why this question includes "representative of the population of interest" - should be reflective of the inclusion/exclusion criteria.

Question	Modified question	Consensus before removing 3 participants			Consensus after removing 3 participants	Panelist's comments
		Percentage of agreement	Median	Interquartile range	Percetage of agreement	
	going to receive the drug					I agree that it is important for this information to all be described in the study when possible. However, as mentioned in a previous comment, some PK studies do not necessarily need to be in the population of patients intended to receive the drug. E.g. if looking at PK interactions via cytochrome P450 enzymes - regardless of whether you're the intended patient population, the interactions should be captured in an appropriately designed study if an interaction truly does exist. If looking at things such as impact of renal impairment, hepatic impairment, critical illness, etc. then it would be prudent to be studied in the actual population of interest. • Again, BE or clinpharm studies do not necessarily have to be done in the same population of interest, they address specific questions that should be population agnostic. • Please see my response to previous questions on population. • Some of these factors are generally important (race, sex, age, etc); however, others are condition-

Question	Modified question	Consensus before removing 3 participants			Consensus after removing 3 participants	Panelist's comments
		Percentage of agreement	Median	Interquartile range	Percetage of agreement	
						• specific (severity, relevant co-interventions and comorbidities, smoking habit) and may not be necessary to collect for all studies.
Q31) Were the study limitations described by the authors consistent with those identified within the study?	Q31) "Were the study limitations acknowledged by the authors"?	Agree (15) 71.43% Neither Agree Nor Disagree: (3) 14.28%Disagree: (3) 14.28%	4	1		• This item is not appropriate for an appraisal tool that aims to assess the quality of the trial, it is only essential for reporting checklist. • This information is relevant to be discussed as this will help in determining sources of bias and any imprecision in the displayed results.
Q32) *Were the author interpretations of the data consistent with the reported results?	Q32) Were the author interpretations of the data consistent with the reported results?	Agree: (18) 85.71% Neither Agree Nor Disagree: (0) 0% Disagree: (3) 14.28%	5	1	Agree: (15) 83.33% Neither Agree Nor Disagree: (0) 0% Disagree: (3) 16.67%	• • Authors interpretations should be comparable to the results that they reported. • This item is not appropriate for an appraisal tool that aims to assess the quality of the trial; it is only important for reporting checklist.
Q33) Did the authors compare their	Q33) Did the authors compare their observed	Agree: (16) 76.19%	4	1	Agree: (13) 72.22%	• This item is not appropriate for

Question	Modified question	Consensus before removing 3 participants			Consensus after removing 3 participants	Panelist's comments
		Percentage of agreement	Median	Interquartile range	Percetage of agreement	
observed results with the results of other relevant studies?	results with the results of other relevant studies?	Neither Agree Nor Disagree: (2) 9.52% Disagree: (3) 14.28%			Neither Agree Nor Disagree: (2) 11.11% Disagree: (3) 16.67%	an appraisal tool that aims to assess the quality of the trial; it is only essential for reporting checklist. • This will not affect the methodological quality of a study.
Q34) Were recommendations of future studies justified based on the results of this study?	Q34) Were recommendations for future studies justified based on the results of this study?	Agree: (12) 57.14% Neither Agree Nor Disagree: (5) 23.80% Disagree: (4) 19.04%	4	1		• It is important to discuss future studies in the article, but the quality of the study will be affected if the author does consider that. • This information is relevant to be discussed as this will help in determining sources of bias and any imprecision in the displayed results.
Q35) *Were the provided conclusions supported by the observed results? Authors should not provide any new information in the	-	Agree: (19) 90.47% Neither Agree Nor Disagree: (0) 0%	5	1	-	

Question	Modified question	Consensus before removing 3 participants			Consensus after removing 3 participants	Panelist's comments
		Percentage of agreement	Median	Interquartile range	Percetage of agreement	
conclusion.		Disagree: 2) 9.52%				•
Q37) Were reported funding resources likely to influence the results of the study?	Q37) "Could reported funding sources have possibly influenced the results of the study?"	Agree: (15) 71.42% Neither Agree Nor Disagree: (3) 14.28% Disagree: (3) 14.28%	4	2		• This is essential information that should be disclosed when the funding source is involved in in the data collection, data analysis and reporting of results. • This will not affect the quality of the study because the rigor of the design is what mainly has an impact on the quality of the results.
Q38) Were disclosed conflicts of interest likely to influence the results of the study?	Q38) "Could disclosed conflicts of interests have possibly influenced the study results"?	Agree: (13) 61.90% Neither Agree Nor Disagree: (4) 19.04% Disagree: (4) 19.04%	4	2		• Conflict of interest is essential to be documented. • If the study is well designed, conflict of interest will not affect its quality.

*Despite of achieving consensus, questions were recirculated because of receiving negative comments.

Table 16: The excluded control questions from round-1

Questions	Percentage of agreement	Median	Interquartile range
The time duration from study submission to publication was less than 1 year.	Agree: (3)14.28% Neither Agree Nor Disagree: (7) 33.33% Disagree: (11) 52.38%	2	1
The authors referenced at least 10 other studies in order to defend their conclusions?	Agree: (3)14.28% Neither Agree Nor Disagree: (6) 28.57% Disagree: (12) 57.14%	2	1

4.2.3.2. Consensus through round 2

In round 2 of the modified Delphi process, 23/25 (92%) of participants responded to the survey. Round 2 survey composed of 28 questions: the consent form, 25 potential questions related to apprising clinical pharmacokinetics studies and another question asking them about the best rating scale to be used by the end-users of the tool and again to suggest further questions or modifications (Figures 7 and 9). Of 25 potential clinical pharmacokinetics critical appraisal tool items, 6 items reached ≥80%, a median ≥ 4, and interquartile range < 1 consensus from experts for inclusion and were retained and simply modified from the linguistic side based on the panelist's comments (Table 17). Ten items did not achieve ≥80% consensus from experts for exclusion, but the change in the distribution of responses was less than 15% between round 1 and round 2 (Table 18), therefore, these items were removed. Although there was more than a 15% change in the distribution of responses in 6 items, these were excluded as more participants suggested to exclude the items and many comments were negative from panelists regarding the inclusion of these items in the final tool (Table 19). Three items were reformulated based on the panelists' comments and sent to round 3 (Table 20). Furthermore, 2 appendices: Appendix E and Appendix F, were attached to clarify the following two questions respectively "Where applicable, were the relevant baseline characteristics of the participants adequately described?", "Was the described population pharmacokinetics approach validation method appropriate for the analysis?" (Appendix E, and Appendix F). Additionally, the consent form was sent to ask participants about their agreement to participate in the process again in round 2.

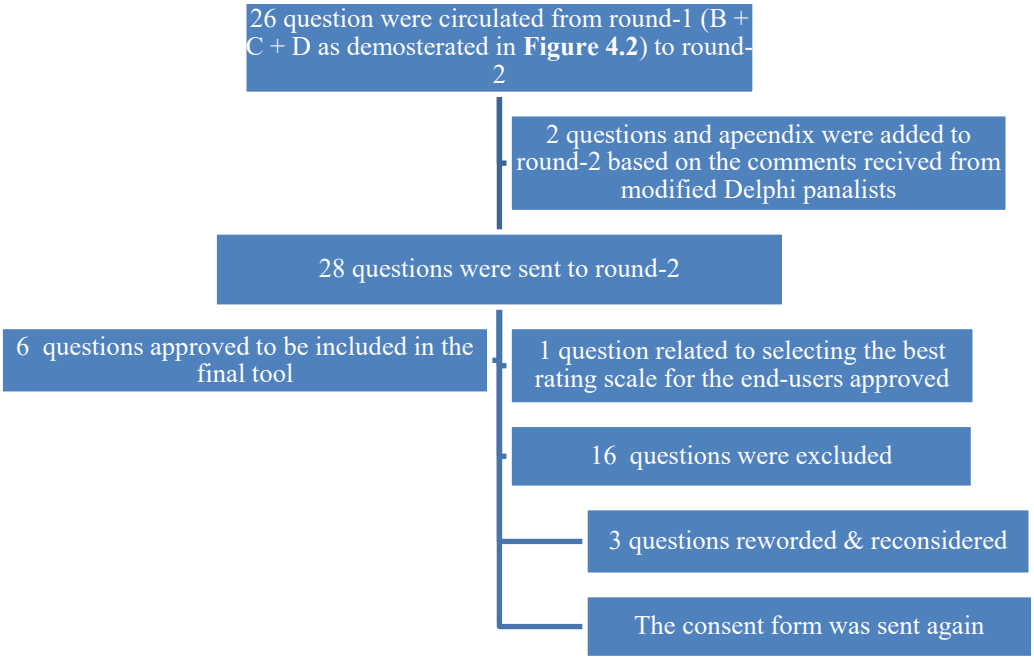

Figure 7. Flowchart of round-2 questions

Table 17: Questions that met consensus for inclusion through round-2

Question	Modified question	Consensus				Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Median	Interquartile range	
Appraising Design						
Q9) Were plausible interacting covariates described a priori or in post hoc evaluation? Examples: demographic variables, laboratory values, concomitant medications, and relevant disease states to the drug being studied. Summary of comments from round-1: 1- This information should be mentioned in the inclusion and exclusion criteria. 2- This only applies if a covariate analysis is included in the clinical pharmacokinetics analysis. 3- It would be helpful if some of these factors were acknowledged as		Agree: (15) 83.33% Neither Agree Nor Disagree: (2) 11.11% Disagree: (1) 5.56%	Agree: (18) 94.73% Neither Agree Nor Disagree: (1) 5.26% Disagree: (0) 0%	4	0	• Agree with comment #3. Not sure that a post-hoc evaluation is truly necessary - but that any plausible interacting covariates were described a priori or in the discussion of the findings. • I agree this should be in the inclusion/exclusion question, it should be mentioned and ideally addressed a priori and analyzed post hoc if possible • Covariate analysis should be conducted after the development of the structural PK

Question	Modified question	Consensus				Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Median	Interquartile range	
it is difficult to describe all of						model. Covariate analysis is not necessary for all clinical PK trials.
Q10) Was the description of the used sample analysis methods or citations of prior validation studies provided in the publication or affiliated appendix? Example: Chromatography type. Detection type. Assay characteristics: mobile phase composition, gradient and flow rate, chromatographic column (packing material, dimensions). Analytical runtime. Operating temperature and detection parameters. Validation method: specificity, recovery, linearity and sensitivity, the stability of the assay and its reproducibility.		Agree: (15) 83.33% Neither Agree Nor Disagree: (3) 16.67% Disagree: (0) 0%	Agree: (16) 84.21% Neither Agree Nor Disagree: (2) 10.52% Disagree: (1) 5.26%	4	1	• Honestly, I don't think that level of details is needed to interpret the results of the study. If comparison to other studies that used different assays are conducted, the authors should highlight that difference. • Should be described in the paper and therefore a part of the appraisal consideration. The person doing the appraisal should not have to judge

Question	Modified question	Consensus				Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Median	Interquartile range	
Summary of Comments from round-1: 1- This will help readers to compare the results of different studies. 2- This will be beyond the level of the reader's knowledge to judge if the used sample analysis methods are appropriate or not. 3- A brief description of the assay characteristics and validation method is essential to be included in a paper. The details should available in an online appendix or contacting the corresponding author.						appropriateness of these analyses. • I believe this to be essential for the precision of the study. It speaks to the reproducibility of the results and while it may be beyond many clinicians understanding, it will allow context to error rates and compare between studies • The details are needed, but I agree with comment 2 that the details of those assays may be too complex.

Question	Modified question	Consensus				Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Median	Interquartile range	
						. • Though agreed that not all readers would be able to understand/judge the quality of the analysis, I also agree with comment #3, in that this information should be available to readers. • 6-As long as the method is referenced in the article then it is good.
Q11) *Were the pharmacokinetics equations used to calculate patient pharmacokinetics parameters disclosed or cited within the article? Example: creatinine clearance, body weight calculations, Michaelis Menten, volume of distribution, patient weight: total body weight vs. ideal body weight.		Agree: (15) 83.33% Neither Agree Nor Disagree: (1) 5.56% Disagree: (2) 11.11%	Agree: (15) 78.94% Neither Agree Nor Disagree: (3) 15.78% Disagree: (1) 5.26%	4	0.5	• A studies quality is, in part, dependent on the transparency and reproducibility. Having equations (or names of them) in text is beneficial for this purpose.

Question	Modified question	Consensus				Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Median	Interquartile range	
Summary of comments from round-1: 1- This is important item since pharmacokinetics equations are not disclosed in a lot of articles. 2- Disclosing equations is very important for reproducibility but will not reduce the quality of an article if they are not disclosed						• I disagree with comment 2. If inappropriate or incorrect equations are used it will affect the quality of the article, and the interpretation of the results. These equations should be included, if not in the body of the article, in supplemental data. • This can help when assessing quality, as it is possible for inappropriate models/equations to be applied. This can affect results.

Question	Modified question	Consensus				Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Median	Interquartile range	
Q15) **Was the number of half-lives elapsed within the sampling period appropriate for the analyzed drug?			Agree: (17) 89.47% Neither Agree Nor Disagree: (2) 10.52% Disagree: (0) 0%	4	1	• Must be 3-5 half lives unless the steady is estimating steady state PK, in which case, time to achieve steady state needs to be taken into consideration in the design of the study. • I understand this to be a part of the follow up period and should be in that question. It is an important question to ask . • which speaks to the quality of the methodology and accuracy of the outcome. I agree it should be here but not as a stand alone question. • This will depend on the study design. It is critical to obtain

Question	Modified question	Consensus				Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Median	Interquartile range	
						the true elimination phase of the drug for PK analysis. Additionally, if studies are looking at steady- state concentrations, an understanding of half-life is required to know if you have obtained true steady-state concentrations. • This question seems out of place in isolation. Essentially, the follow-up timing (sampling frequency) and duration should match the question. For example, half-life may be irrelevant in a study evaluating achieved Cmax with a loading dose.

Question	Modified question	Consensus				Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Median	Interquartile range	
Q16) **Were the essential pharmacokinetics parameters required to make the results applicable in clinical settings? Consider: Total clearance (CL), Fraction of dose excreted unchanged in urine (fe), Volume of distribution at steady state (Vss), Volume of distribution during the terminal phase (VZ), Blood/plasma concentration ratio, Terminal half-life (t1/2 Z), Fraction of the unbound drug in plasma (fu), Bioavailable fraction of dose (F), Absorption rate constant (Ka),Cmin, Cmax, tmax, EC50, Ke0, Hill coefficient, or gamma, AUC and bioavailability for the two drug formulations.	Were the essential pharmacokinetics parameters required to make the results applicable in clinical settings addressed? Consider: Total clearance (CL), Fraction of dose excreted unchanged in urine (fe), Volume of distribution at steady state (Vss), Volume of distribution during the terminal phase (VZ), Blood/plasma concentration ratio, Terminal half-life (t1/2 Z), Fraction of the unbound drug in plasma (fu), Bioavailable fraction of dose (F), Absorption rate constant (Ka),Cmin, Cmax, tmax, EC50, Ke0, Hill coefficient, or gamma, AUC and		Agree: (16) 84.21% Neither Agree Nor Disagree: (3) 15.78% Disagree: (0) 0%	5	1	• Not sure I understand this question • I think the wording of this question can be improved. "Were the essential pharmacokinetics parameters required to make the results applicable in clinical settings addressed?" • I am not clear on what this criteria is asking the appraiser to assess ... this needs clarity. • These are the essential pieces of a PK study and to not have the description of appropriate PK parameters would significantly hinder the quality of the outcome.

Question	Modified question	Consensus				Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Median	Interquartile range	
	bioavailability for the two drug formulations					• The essential PK parameters necessary to adequately describe the data will be dependent on the specific study design. This may be difficult to generalize across multiple study designs. • I think the question language is incomplete. Does not read well.
Q17) Were all patients enrolled in the study accounted for? Example: Description of patient screening, enrollment, run-in or wash out phases, study period and follow-up periods are adequality described. Any loss to follow-up or withdrawals are described. Summary of comments from round-1: 1- This information is important to be	Were all patients enrolled in the study accounted for? Example: Description of patient screening, enrollment, run-in or wash out phases, study period and follow-up periods are "adequately" described. Any loss to follow-up or withdrawals are described.	Agree: (15) 71.43% Neither Agree Nor Disagree: (4) 19.05% Disagree: (2) 9.52%	Agree: (17) 89.47% Neither Agree Nor Disagree: (2) 10.52% Disagree: (0) 0%	4	1	• typo "adequality" to "adequately" • 2-Should be included as a study flow chart in the appendix • 3-This information may be moved to the appendix (this depends on the journal format and regulations) • 4-Should be presented in a flow diagram if the

Question	Modified question	Consensus				Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Median	Interquartile range	
	described either in the form of a flow diagram or within the text. 2- This information should be moved to the appendix.					numbers are significant. Must be commented on to ensure bias can be identified For PK model development and analysis this isn't critical because analysis should be based only on available data.

*The percentage of agreement about including this question was reduced because a summary of the panelists' comments in round-1 was populated in round-2 survey for the panelists to think again about the question.
**This question was newly added to round-2 based on the received comments from panelists in round-1.

Table 18: Questions that met consensus for exclusion through round-2

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Difference in %*	Median	Interquartile range	
Q5) If applicable, did the eligibility criteria of participants inclusion reflect the population of interest for which the drug is intended for use? The inclusion criteria should represent the targeted population to answer the main study question. If renal or hepatic function impacts the pharmacokinetics parameters, patients who have varying degrees of renal insufficiency should be included . Summary of comments from round-1: 1-This item is not important for every clinical pharmacokinetics study. For example, in the drug-drug interaction study, the medication is		Agree: (13) 72.22% Neither Agree Nor Disagree: (2) 11.11% Disagree: (3) 16.67%	Agree: (14) 73.68% Neither Agree Nor Disagree: (2) 10.52% Disagree: (3) 15.78%	Agree: 1.46% Neither Agree Nor Disagree: -0.59% Disagree: -0.89	4	1	• As long as, the results can be extrapolated from HV to patients, the population doesn't matter. • Agree but suggest changing wording to "If applicable, did the inclusion criteria of participants reflect the population of interest for which the drug is intended for use? For example, if renal or hepatic function impacts the pharmacokinetics parameters, patients who have varying degrees of renal insufficiency should be included."

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Difference in %	Median	Interquartile range	
given to a patient who might not be receiving it in the clinical setting. However, in case in which renal or hepatic function might affect the pharmacokinetics parameters, patients suffering from hepatic or renal insufficiency should be included. 2-If the pharmacokinetics parameters are not affected by the diseases, the study could be conducted on healthy volunteers then the results extrapolated to patients in clinical setting.							• The patient population should reflect the population of interest. However depending on the type of study healthy volunteers may or may not be necessary. • I agree with comment #2 above - you could re-word this criteria to have the reader consider whether the PK parameters are affected by disease(s) target or co-morbid, and then consider whether the study population, be it healthy volunteers or clinical patients, was appropriate or not. • I still believe this is an important

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Difference in %	Median	Interquartile range	
							question to ask for the reasons mentioned above. The question which comes out of this is basically asking the summary of the comments from round 1 in my opinion: if the PK is not altered by the disease, then a healthy patient is does reflect the population of interest. This question would not be asked if you did not assess the included patients (although there may be a better way to word the question?) • This should be included but with the parameters"if applicable" to

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Difference in %	Median	Interquartile range	
Q6) Was the study setting/location relevant to where the drug would be used? Authors should describe the setting and/or location of the study. They should also follow local practice guideline recommendations to ensure that there is an understanding of where/how the drug is used. Summary of comments from round-1 1- This may not be relevant to all clinical pharmacokinetics studies. 2- This is relevant when the drug disposition is altered due to changes in the physiology, such as in a critically ill		Agree: (11) 52.38% Neither Agree Nor Disagree: (4) 19.05% Disagree: (6) 28.57%	Agree: (10) 52.63% Neither Agree Nor Disagree: (5) 26.31% Disagree: (4) 21.05%	Agree: 0.25% Neither Agree Nor Disagree: 7.26% Disagree: -7.52%	4	1	• address the comment in #1. • There are instances where this would be critical; however, it does not broadly a pply across all clinical trial types. In some cases this would be a crucial measure, eg. if women are excluded from the study, or individuals with varying genotypes or disease progression were excluded. • Not important for every clinical PK study Per the above comments, this would be a 'Strongly Agree' if the PK parameters could be affected by the diseases. Therefore, for an

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Difference in %	Median	Interquartile range	
	patients.						• all-inclusive tool, this should be included in the checklist, and an option could be 'N/A' for the studies where it is irrelevant.
3-	The location is important for generalizability of the results, but the quality of the research should not be affected if its methodology was comprehensive and robust.						• Location to me implies geographic location which I believe is less important. Clinical setting, however, is relevant (pregnant women, obese subjects, critically ill, etc), therefore I recommend changing wording to "Was the study's clinical setting relevant to the patient population in which the drug would be used? If relevant or

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Difference in %	Median	Interquartile range	
							applicable, study participants should reflect the intended patients with altered pharmacokinetics, eg. pregnant women, obese subjects, critically ill." • Agree with all the comments; speaks to generalizability of the results and could be one component of an appraisal tool as it is with an appraisal tool for randomized clinical trials. If results are not generalizable then the readers should take this into consideration when using/applying the findings. • While not

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Difference in %	Median	Interquartile range	
							important to all PK studies, it is still an important question to ask for the accuracy of the results to a population you are extrapolating to • I agree with the comments, and would extend that to include the ethnic populations included in the study. Studies with narrow population characteristics are not applicable to more diverse populations. • I agree with the above comments that it could go either way. If the local practice guidelines are quite variable from the general standard of

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Difference in %	Median	Interquartile range	
							• practice, this may be more of a quality-related piece.
Q7) If applicable, was the blinding of the study participants, the investigators and/or those analyzing the data appropriate while the study was being conducted? Summary of comments from round-1: 1- In most pharmacokinetics studies, blinding is not essential for participants, investigators and data analyzers because they measure objective endpoints like drug concentration. 2- The outcome of interest is the only factor that defines		Agree: (14) 66.67% Neither Agree Nor Disagree: (2) 9.52% Disagree: (5) 23.80%	Agree: (14) 73.68% Neither Agree Nor Disagree: (2) 10.52% Disagree: (3) 15.78%	Agree: 7.01% Neither Agree Nor Disagree: 1% Disagree: -8.02%	4	0.5	• As appropriate, the blinding should take place. For example, in biosimilars BE studies, it's important to maintain a blind. However, most of the clinpharm studies are unblinded. • I agree with previous comments that blinding may not always be necessary if the outcome is objective. • The question is worded "if

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Difference in %	Median	Interquartile range	
the importance of blinding. For example: when dose-response is investigated, blinding is important to avoid subjective influence.							• applicable" which can account for these studies but it may be helpful to include the phrase "if an objective outcome isn't used, was the blinding of the study participants.. etc." or something along those lines • This item could be re-worded to be clearer. "If applicable, aside from objective laboratory measurements, was the blinding of study participants, the investigators and/or those analyzing the data pertaining to Outcome assessment appropriate while

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Difference in %	Median	Interquartile range	
							the study was being conducted?" • While not important for every PK study, blinding is important for clinical PK studies especially if there is an objective laboratory measurements, was the blinding of study participants, the investigators and/or those analyzing the data • May not be applicable for all studies but should be for all studies but should be included for Dose-response studies. • I agree with the

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Difference in %	Median	Interquartile range	
							comments listed. • As per the above, this would be important more so for subjective outcomes (e.g. dose-response). Perhaps this could pertaining to outcome critical to be included in the checklist, and an 'N/A' option be provided for studies this is not relevant (e.g. for dose-concentration or time-concentration).
Q8) Was participant follow-up clearly described and rationalized? Example: Monitoring parameters (e.g. signs and symptoms of disease or side effects of the given medication, lab data, etc.) to	Q8) Was participant follow-up clearly described and adequate time frame allowed for	Agree: (14) 77.78% Neither Agree Nor Disagree: (0) 0% Disagree: (2) 11.11%	Agree: (14) 73.68% Neither Agree Nor Disagree: (4) 21.05% Disagree: (1) 5.26%	Agree: -4.1% Neither Agree Nor Disagree: 21.05%	4	1	• I agree that in general follow-up is not required, however, adequate and appropriate sampling should be collected based on the known

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Difference in %	Median	Interquartile range	
be collected in each period should be documented. Summary of comments from round-1: 1- Participants follow up is not required in many clinical pharmacokinetics studies.	the outcomes of interest? (e.g. serum concentrations vs clinical effect - was there enough time for a clinical effect to be present) OR Was the follow up appropriate for the information trying to gather is another way of putting it. Summary of comments from round-2:	Comments: (2) 11.11%		Disagree: -5.85%			pharmacokinetics (eg. half-life, distribution models). Is that captured in another question? • Would add "If applicable" to the beginning of this criteria. And if applicable, it should be considered in an appraisal of the quality of the study. • I am not sure I see this question in the same light as others. Follow up is absolutely required for adequate monitoring and assessment of a PK trial. Was the follow up appropriate for the information trying

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Difference in %	Median	Interquartile range	
	1-Participants follow up is essential specifically for laboratory parameters. 2-8Participants follow up is not required in many clinical pharmacokinetics studies						to gather is another way of putting it. If we are looking at drug-drug interactions with enzyme induction, is looking at the drug levels 1 hr post dose 1 dose even appropriate? Or should we follow daily x1 week? or x2 weeks? • Important to know follow up for laboratory parameters. • I agree that this is Not need In clinical PK studies, only pharmacodynamic studies. • Important to know follow up for laboratory parameters.

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Difference in %	Median	Interquartile range	
							• I agree that this is not need in clinical PK studies, only pharmacodynamic studies. • Suggest rewording: 8. Was participant follow-up clearly described and adequate time frame allowed for the outcomes of interest? (e.g. serum concentrations vs clinical effect - was there enough time for a clinical effect to be present)
Q13) Was appropriate Research Ethics Board approval received? Summary of comments from round-1: 1- This element is		Agree: (14) 66.67% Neither Agree Nor Disagree: (4) 19.04%	Agree: (12) 63.15% Neither Agree Nor Disagree: (3) 15.78% Disagree: (4)	Agree: -3.52% Neither Agree Nor Disagree: -3.26%	4	1.5	• This is an ethics question, all studies should be ethical, yes all studies should indicate IRB approvals were granted.

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Difference in %	Median	Interquartile range	
1- important to assess the quality of reporting but not the quality of the overall study. 2- Authors should state that REB has approved the study, but there is no need to explicitly mention the approval number.		Disagree: (3) 14.28%	21.05%	Disagree: 6.77%			• Change wording to "Was appropriate Research Ethics Board approval received and documented?" • I think it does speak to the quality of a study if the methodology has been reviewed and approved by an external ethics board. Agree with comment #2 • I agree with the comments listed. To be published Ethics approval must be included.
Q14) Was the level of statistical significance appropriate for the intended outcomes of the study?		Agree: (13) 72.22% Neither Agree Nor	Agree: (12) 63.15% Neither Agree Nor Disagree:	Agree: -9.07% Neither Agree Nor	4	1	• comments from Round 1. • The question needs to be

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Difference in %	Median	Interquartile range	
Summary of comments from round-1: 1- This item is not applicable for all clinical pharmacokinetics studies as the primary endpoints in some of the studies are to describe pharmacokinetics parameters. 2- If the goal of the study was to detect a difference, authors should clearly state how did they calculate the sample size.		Disagree: (5) 27.78% Disagree: (0) 0%	(5) 26.31% Disagree: (2) 10.52%	Disagree: -1.47% Disagree: 10.52%			• modified to ask about the statistical rigor of the study. Most clinpharm studies are interested in the central tendency unless a claim for BE is warranted. In such case, the sample size calculation and the study especially to generalize findings to a greater population. • Agree with comment three from round 1: when it is applicable to the study outcomes, the authors should describe how the • sample size was calculated and if they had enough

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Difference in %	Median	Interquartile range	
							• power to detect a statistically significant difference • If applicable" should be added to the item. • An important question to ask and understand the results. The authors must describe the outcome context to properly understand the level of statistical significance (a good background should do this). And as noted in the round 1 comments, may not be important to every study. May want to add "if applicable, ..." • Agree with above comments that

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Difference in %	Median	Interquartile range	
							this may not apply to all studies. • This should not be included in the evaluation of quality of trial. • Agreed that this may not be applicable for all study types (e.g. qualitative analysis) • Agree with comments from round 1. Essentially, the statistical analysis and thresholds for statistical significance should make sense in the context of the study objectives. • Suggest wording change: If statistical tests were conducted, Was the level of

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Difference in %	Median	Interquartile range	
							statistical significance appropriate for the intended outcomes of the study? - do you just mean
Q18) Were the relevant baseline characteristics of the included participants reflective of the inclusion/exclusion criteria? Example of important participants characteristics: sex, race, age, weight, height, concomitant diseases, administrated medications, smoking habits, severity of illness that may affect pharmacokinetics parameters, renal function, and hepatic function. Summary of comments from round-1: 1- This information is important to be		Agree: (14) 66.67% Neither Agree Nor Disagree: (3) 14.28% Disagree: (4) 19.04%	Agree: (14) 73.68% Neither Agree Nor Disagree: (3) 15.78% Disagree: (2) 10.52%	Agree: 7.01% Neither Agree Nor Disagree: 1.5% Disagree: -8.52%	4	1.5	p<0.05 or that the statistical test was appropriate given the data type/number • Yes, subject demographics needs to be reported and complying with I/E critter is is an indication of the proper conduct of the study • Change wording to "Where applicable, were the relevant baseline. • characteristics of the participants

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Difference in %	Median	Interquartile range	
documented. However, this is not important for every pharmacokinetics study, such as bioequivalence studies that are conducted in a population who are not going to receive the drug.							• included? Examples: sex, race, age, weight, height, concomitant diseases, administrated medications, smoking status, pregnancy, severity of illness that may affect pharmacokinetics parameters, renal function, and hepatic function." • I think having the word "relevant" in the criteria above addresses comment #1.I agree with the first comment.
Q19) Were the study limitations acknowledged by the authors? Summary of comments from round-1:	Q19) "are the limitations in the study that hinder your ability	Agree (15) 71.43% Neither Agree Nor Disagree:	Agree: (13) 68.42% Neither Agree Nor Disagree: (0) 0%	Agree: -3.01% Neither Agree Nor Disagree: -	4	2.5	• This is important because it forces the reader/user to consider what the limitations are and then see if the

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Difference in %	Median	Interquartile range	
1-This item is not appropriate for an appraisal tool that aims to assess the quality of the trial, it is only essential for reporting checklist. 2-This information is relevant to be discussed as this will help in determining sources of bias and any imprecision in the displayed results.	to apply the results to a specific patient population? Were these limitations discussed/rationalized"? Summary of comments from round-2: 1-This question is important because it will help readers to identify what the limitations are and determine if the	14.28% Disagree: 14.28%	Disagree: (6) 31.57%	14.28% Disagree: 17.29%			authors acknowledged those same things. • Agree - this is not needed for an appraisal tool but rather for a reporting checklist. • agree with comment #2 This is important information however a clinician should be able to identify study limitations on their own. A better question could be "are the limitations in the study that hinder your ability to apply the results to a specific patient population? Were these limitations discussed/

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Difference in %	Median	Interquartile range	
	authors acknowledged them or not. 2-This item is not appropriate for an appraisal tool that aims to assess the quality of the trial, it is only essential for reporting checklist.						rationalized"? • This should not be included in an appraisal tool, but it is important for interpretation and application of the results. For example, the trial may be in a limited patient population, so the results cannot be extended to the general the clinical trial and report were of high quality for the specific patient population. • Limitations listed by the authors may be useful in understanding other types of bias or confounders in the study • Neither funding or

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Difference in %	Median	Interquartile range	
							COI should be used in judging the quality of the study.
Q24) Could reported funding sources have possibly influenced the results of the study? Summary of comments from round-1: 1- This is essential information that should be disclosed when the funding source is involved in the data collection, data analysis, and reporting of results. 2- This will not affect the quality of the study because the rigor of the design is what mainly has an impact on the quality of the results.		Agree: (15) 71.42% Neither Agree Nor Disagree: (3) 14.28% Disagree: (3) 14.28%	Agree: (12) 63.15% Neither Agree Nor Disagree: (4) 21.05% Disagree: (3) 15.78%	Agree: -8.27% Neither Agree Nor Disagree: 6.77% Disagree: 1.5%	4	1.5	• This question forces the reader/user to seek out this information in the manuscript. It may not have influence on the results though. Consider wording: "Were reported funding sources involved in data collection, analysis, and results reporting?" If yes, were measures taken to mitigate potential influence from these sources?" • Include in checklist not critical appraisal tool

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Difference in %	Median	Interquartile range	
							agree with comment #1 • I think the wording of this question is important to addressing the quality of the study and if the reporting could have been altered. (which assesses the quality of not just the design but the outcomes)This must be disclosed and is required by all journals. This may affect study design and the rigor of analysis. • Agree that funding • should be disclosed, but that quality of results is dependent on methodology.

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Difference in %	Median	Interquartile range	
Q25) Could disclosed conflicts of interests have possibly influenced the study results? Summary of comments from round-1: 1- Conflict of interest is essential to be documented. 2- If the study is well designed, conflict of interest will not affect its quality.		Agree: (13) 61.90% Neither Agree Nor Disagree: (4) 19.04% Disagree: (4) 19.04%	Agree: (13) 68.42% Neither Agree Nor Disagree: (2) 10.52% Disagree: (4) 21.05%	Agree: 6.52% Neither Agree Nor Disagree: -8.52% Disagree: 2.01%	4	2	• Neither funding or COI should be used in judging the quality of the study Include in checklist not critical appraisal tool • Agree that a well designed study could overcome potential influence from funding or conflicts of interest, but either way this should be clearly reported by the authors • agree with comment #1 • This is important to address as is the sources of funding question. • To imply that a well designed x

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Difference in %	Median	Interquartile range	
							study could not be impacted by the conflicts is narrow-minded • I agree with comment 1. Conflict of interest may influence the quality of the study design and analysis. • Agree that conflicts of interest should be disclosed, but that quality of results is dependent on the methodology. • Comment 2 is flawed. CoI is essential to assess.Funding/sponsorship bias is well known phenomenon.

* Difference in % was calculated starting from round-2 only to compare the difference in agreement between round-1 and round-2.

Table 19: Questions that were excluded through round-2 despite a change in the distribution of responses of more than 15%

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Difference in %	Median	Interquartile range	
Q2) Did the title concisely reflect the topic discussed in the paper? The title should reflect the name of the analyzed medication, and comparator (if applicable), the targeted patient population, and the study rational or design. Summary of comments from round-1:	Did the title concisely reflect the topic discussed in the paper? The title should reflect the name of the analyzed medication, and comparator (if applicable), and the targeted population, Summary of comments from round-2: 1- This item is not appropriate for an appraisal tool that aims to assess the quality of the trial; it is only important for reporting checklist.	Agree: (21) 87.5 % Neither Agree Nor Disagree: (0) 0% Disagree: (3) 12.5%	Agree: (10) 47.61 % Neither Agree Nor Disagree: (3) 14.28% Disagree: (8) 38.09%	Agree: -39.89 Neither Agree Nor Disagree: 14.28% Disagree: 25.59%	3	2	• While the title of the publication is not a metric of the robustness of the study or the data, it's an important aspect in reporting the data. • Agree with the comments from round 1. • Suggest removing the word "patient" because population is sufficient (the study could be done in healthy volunteers) Will the checklist have binary yes/no option to answer this question? This can be challenging for checklist users to say yes, if one aspect is not included (eg. The

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Difference in %	Median	Interquartile range	
1- This item is not appropriate for an appraisal tool that aims to assess the quality of the trial; it is only important for reporting checklist. 2- It is essential to mention the experimental arm, the comparator and, the patient population, however, the study design is not considered as critical element.	2- There is no need to include the study design or the rational since some of the journals may apply character limitation for the title. Therefore, it will be impossible to add all of the listed items						title mentions the patient population and analyzed medication, but not the comparator). How will users reconcile this situation? I agree that "study design" could be removed from the item • Agree with the second comment from round 1 that the study rational or design may not need to be included, but the other elements are essential. • It is nice to have but would not tell me anything about the quality of the trial. Agree that it would be beneficial for a checklist but not

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Difference in %	Median	Interquartile range	
							for appraisal of the actual evidence. • Would agree with both comments - item isn't necessary to appraise quality of trial. Ideally a study title should reflect the target drug comparator and study population. This could be a checklist type item in a tool rather then impacting the overall quality assessment. • Agree with comments above. Would remove "design". • I agree with comment 1. The title should not be

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Difference in %	Median	Interquartile range	
							an assessment measure for the quality of the study. Some journals have character. • limitations on titles, so all of the information listed may not fit. Can be included but as with the comment #2, type of study design is not vital in here • I agree that it is not necessary for a quality assessment, per se, but would be important for a reporting checklist. • Title is not relevant to quality of study or the interpretation of

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Difference in %	Median	Interquartile range	
Q3) Was an adequate summary of the article provided to the readers within the abstract? A brief description of the knowledge gap, the objectives, summary of the methods (patient population studied, the route of administration of the drug(s) studied), the results of the primary objectives, as well as, secondary or exploratory outcomes and the main conclusion should be provided. Summary of comments from round-1: 1- This item is essential to assess the quality of reporting of		Agree: (16) 69.56% Neither Agree Nor Disagree: (2) 8.70% Disagree: (5) 21.74%	Agree: (10) 47.61% Neither Agree Nor Disagree: (2) 8.70% Disagree: (10) 47.61%	Agree: -21.95% Disagree: -3.94% Disagree: -25.87	3	2	the study. • I agree that this is important for quality of reporting so it is helpful for the checklist, but if something absolutely needs to be cut, this could be cut. • Agree with the comments from round 1 - as a peer reviewer is essential to review but for critical appraisal not needed. • Agree with comments made about the abstract not necessarily linking to the quality of the study and therefore not necessary for the question tool

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Difference in %	Median	Interquartile range	
	clinical pharmacokinetics studies but does not influence the quality of a study and is therefore not necessary.						• Similar to question 1, it is helpful for clarity of quick reference and utilization for bedside use but it does not assess the quality of the actual study
2-	It is essential to have an excellent abstract, but there is doubt of how this will help in assessing the quality of the whole study.						• Agree with both comments from round 1. Similar to the title – a well written abstract is desirable but not necessarily reflective of the quality of the actual study being reported. Perhaps more of a checklist • item rather than impacting quality assessment. • Agree with comment 2. • I agree with the

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Difference in %	Median	Interquartile range	
							• comments from round 1. • I agree with the comments from round 1 re: that this is not a necessity for quality assessment but is important for a reporting checklist. • Agree with comments from round 1.Agree with comment
Q20) Were the author interpretations of the data consistent with the reported results? Summary of comments from round-1: 1- Authors interpretations should be comparable	Q20) Were the author interpretations of the data consistent with the reported results? Summary of comments from round-2: 1- This item is not appropriate for an appraisal tool that aims to	Agree: (15) 83.33% Neither Agree Nor Disagree: (0) 0% Disagree: (3) 16.67%	Agree: (12) 63.15% Neither Agree Nor Disagree: (2) 10.52% Disagree: (5) 26.31%	Agree: -20.18% Neither Agree Nor Disagree: 10.52% Disagree: 9.64%	4	2.5	• agree this should be in an appraisal tool as it is for clinical trials. • This is a key question that I review clinical trials for however, I do agree it does not speak to the quality of the actual trial. • This is not

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Difference in %	Median	Interquartile range	
to the results that they reported. 1- This item is not appropriate for an appraisal tool that aims to assess the quality of the trial; it is only important for reporting checklist	assess the quality of the trial; it is only important for reporting checklist. 2- This item is related to the quality of the author's appraisal.						important for an appraisal tool. • I agree with comment #2, in that it would not necessary affect quality per se. More so the quality of the authors' appraisal.
Q21) Did the authors compare their observed results with the results of other relevant studies? Summary of comments from round-1: 1- This item is not appropriate	Q21) Did the authors compare their observed results with the results of other relevant studies? Summary of comments from round-2: 1- This item is not appropriate	Agree: (13) 72.22% Neither Agree Nor Disagree: (2) 11.11% Disagree: (3) 16.67%	Agree: (6) 31.57% Neither Agree Nor Disagree: (2) 10.52% Disagree: (11) 57.89%	Agree: -40.65% Neither Agree Nor Disagree: -0.59% Disagree: 41.22	2	2	• It's best practice to put the results of the study within context of other similar studies. Not doing so doesn't mean the study results should be discredited • Not necessary

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Difference in %	Median	Interquartile range	
for an appraisal tool that aims to assess the quality of the trial; it is only essential for reporting checklist. 1- This will not affect the methodological quality of the study.	for an appraisal tool that aims to assess the quality of the trial; it is only essential for reporting checklist. 2- This will not affect the methodological quality of the study.						• Agree - checklist only • agree with comments - not • This does give context to the results as they have found them but does not speak to the quality of study. It does allow a well rounded understanding to help interpret the results and may help with identifying bias. • This item will not influence that methodological quality of the study • This is not important for an appraisal tool.

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Difference in %	Median	Interquartile range	
							• Agree with above comments. •
Q22) Were recommendations for future studies justified based on the results of this study? Summary of comments from round-1: 1- It is important to discuss future studies in the article, but the quality of the study will not be affected if the author does not consider that. 2- This information is relevant to be	Q22) Were recommendations for future studies justified based on the results of this study? Summary of comments from round-2: 1- This item is important to assess the quality of a written paper but not the quality of the conducted study	Agree: (12) 57.14% Neither Agree Nor Disagree: (5) 23.80% Disagree: (4) 19.04%	Agree: (6) 36.84% Neither Agree Nor Disagree: (2) 10.52% Disagree: (10) 52.63%	Agree: -20.3% Neither Agree Nor Disagree: -13.28% Disagree: 33.59%	2	2	• Not necessary • Agree that recommendations for future studies is not a requirement of a well done study. Rather it reflects the quality of the writing of the paper. • not sure this needs to be in the appraisal tool. • Agree with comment 2 but do not feel strongly that this would be required for determination of the overall quality of the study. • Not important for

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Difference in %	Median	Interquartile range	
discussed as this will help in determining sources of bias and any imprecision in the displayed results.							evaluation of the clinical trial
Q23) Were the provided conclusions supported by the observed results? Authors should not provide any new information in the conclusion. Summary of comments from round-1: 1- This item is not appropriate for an appraisal tool that aims to	Q23) Were the authors interpretation of results based on observed data and results? Authors should not provide any new information in the conclusion. Summary of comments from round-2: 1- This item is important to be included in an appraisal tool. 2- This item is not appropriate for an	Agree: (19) 90.47% Neither Agree Nor Disagree: (0) 0% Disagree: 2) 9.52%	Agree: (12) 66.67% Neither Agree Nor Disagree: (1) 5.55% Disagree: (5) 27.78%	Agree: -23.8% Neither Agree Nor Disagree: 5.55% Disagree: 18.26%	4	2.5	• This question is similar to if the results the authors reported matched the data they presented • Could just have one question in regards to authors interpretation of results based on observed data and results for ease of using the tool. • I respectfully disagree with comment #1 and think that this should be in an

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	Difference in %	Median	Interquartile range	
assess the quality of the trial; it is only important for reporting checklist.	appraisal tool that aims to assess the quality of the trial.						• appraisal tool. • While this is helpful, it is not valuable for assessing the quality of the study. • Not important for an appraisal tool

Table 20: The recirculated questions to round-3

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	The difference in %	Median	Interquartile range	
Q4) Was a comprehensive introduction provided that explained the background and rational behind the conduction of the study? Authors may provide information related to: the analyzed drug indication (proposed or labelled) including dosing and patient population; stages of the analyzed drug development; known aspects of the drug's absorption, distribution, metabolism and elimination;	Was the rationale behind the purpose of the study reasonable and comprehensible? Summary of comments from round-2: 1- Some readers might not be knowledgeable about the discussed drug, and they need a well-written background and rationale to appraise the quality of the study.	Agree: (16) 76.19% Neither Agree Nor Disagree: (1) 4.76% Disagree: (4) 19.05%	Agree: (8) 42.10% Neither Agree Nor Disagree: (2) 10.52% Disagree: (9) 47.36%	Agree: -34.09 Neither Agree Nor Disagree: 5.76 Disagree: 28.31	3	2	• I agree that this is important for quality of reporting so it is helpful for the checklist, but if something absolutely needs to be cut, this could be cut. • I agree that this item is not quite important as an appraisal tool for clinpharm studies. • Similar to previous, I believe quality of

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	The difference in %	Median	Interquartile range	
previous studies, knowledge gaps, and addressing how those knowledge gaps will be closed, or brought closer through the experimental design presented in the current manuscript; what will be added to the existing body of knowledge with their proposed study. Summary of comments from round-1:							could be cut reporting is important so this should be included, but if needed to shorten. The text could be made to be more concise, for example: "Relevant background and rationale information includes the analyzed drug indication and dosing, stages of the drug development, known aspects of the drug's pharmacokinetics,

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	The difference in %	Median	Interquartile range	
1- This item is used to assess how a research paper is well written but not to evaluate the quality of the study.							previous relevant studies, knowledge gaps, and potential added value of the proposed study." • Agree with comments from round 1
2- This item is not appropriate for an appraisal tool that aims to assess the quality of the trail; it is only essential for reporting checklist.							• Agree with comments about the introduction not necessarily linking to the quality of the study done • Rather the introduction correlates with how

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	The difference in %	Median	Interquartile range	
							well the paper was written. • I do believe having an clear and concise background is important to a well rationalized research study. It should identify the gap in literature and setup the context for what the research question SHOULD be doing and then allow the reader to assess the accuracy of the

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	The difference in %	Median	Interquartile range	
							design/outcomes. If you care only about the precision of the study, then the background is of no benefit. Or if you have considerable knowledge in the topic at hand, it probably isn't necessary. • I would respectfully disagree with the comments from round 1. If a reader is not familiar with the PK

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	The difference in %	Median	Interquartile range	
							aspects of the drugs in question or in the particular type of analysis - this background information would then help inform their remaining appraisal of the study • I agree with both comments; however, itis helpful in understanding the rationale for the trial. • I agree with the above comments

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	The difference in %	Median	Interquartile range	
							from round 1 • Agree with comments from round 1.
Q12) Was the described population pharmacokinetics approach validation method appropriate for the analysis? Basic internal methods Goodness-of-fit plots/diagnostic plots Uncertainty in parameter estimates Model sensitivity to outliers Advanced internal methods Data splitting Bootstrap Cross validation Stimulation such as visual or posterior predictive checks (PPCs) External model evaluation Validation dataset	Q12) Was the described population pharmacokinetics approach validation method appropriate for the analysis? 1- Basic internal method 2- Advanced internal method 3- External model evaluation Note: Please refer to Appendix-1 Model Evaluation attached in the invitation email for further clarification. This appendix will be provided in the final version of the appraisal tool.	Agree: (20) 95.24% Neither Agree Nor Disagree: (1) 4.76% Disagree: (0) 0%	Agree: (11) 57.89% Neither Agree Nor Disagree: (5) 26.31% Disagree: (3) 15.78%	Agree: -37.35% Neither Agree Nor Disagree: 21.55% Disagree: 15.78%	4	2	• I disagree with the comment here, this is an important point. • The model diagnostics need to be included at least in an appendix to judge where the model could be failing. • I believe this is important, however

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	The difference in %	Median	Interquartile range	
observations compared with model predictions Summary of comments from round-1: 1- This will be beyond the level of the reader's knowledge to judge if the used population pharmacokinetics approach and validation are appropriate or not.	Summary of comments from round-2: 1- Although this information might be beyond the level of reader's knowledge to judge if the used population pharmacokinetics approach and validation are appropriate or not, it is important to be included.						am concerned about the average user's knowledge to assess this. • Would reference to an appendix or glossary help? • Beyond the capability of the average reader to be able to assess this for a quality appraisal tool. • This will likely be beyond the clinicians

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	The difference in %	Median	Interquartile range	
							understanding to judge if it was done appropriately. However if we are critically appraising the study to a high standard, the clinician should be familiar and agree/disagree with the method of analysis. Even reading the statistics in a • clinical study is difficult to comprehend but should

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	The difference in %	Median	Interquartile range	
							• still be rationalized to ensure the authors aren't just using a convenience analysis that shows the results they want. • I believe this needs to be included. Those developing the analysis criteria should have enough background in this area to determine the

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	The difference in %	Median	Interquartile range	
							appropriate approaches to use, and would then be able to guide individuals with less experience in population analysis. Validation of PK structural models is critical to the quality of the analysis that is performed. • I believe this is important information, but agree with the comment that it may be above the readers' level of knowledge

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	The difference in %	Median	Interquartile range	
Q27) *Appendix I: Were the baseline characteristics of the included participants provided? All the following variables should be clearly defined for all participants sex, race, age, weight, height, concomitant diseases, co-medication, smoking habits, covariates, the severity of illness, residual, renal function, and hepatic function. Authors should describe if `participants are taking any medications that may interact with the analyzed medication. The Acute Dialysis Quality Initiative (ADQI) minimum reporting criteria by ADQI should be followed in case of including participants on dialysis. a) Operational	Where applicable, were the relevant baseline characteristics of the participants adequately described? Examples: sex, race, age, weight, height, concomitant disease, administered medications, smoking status, pregnancy, severity of illness that may affect pharmacokinetics parameters, renal function, and hepatic		Agree: (12) 70.58% Neither Agree Nor Disagree: (2) 11.76% Disagree: (3) 17.64%		4	1	• I don't feel this appendix is helpful. An appendix on PK methods would be more helpful (eg. bootstra, goodness of fit, etc) • I think Point #5 above is not complete... • I don't feel qualified to comment on #1; #2

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	The difference in %	Median	Interquartile range	
characteristics • Membrane/dialyser/filter and area • A measure of time actually spent on therapy • Delivery device Access and blood flow • Anticoagulation • Replacement fluid composition and administration • Dialysis fluid composition and administration b) Patient characteristics • Measure of time actually spent on therapy • Surgical/trauma/medical/other • Measure of	function. Note: This question will be further clarified by adding Appendix-2 Patient Demographics. Appendix-2 is attached in the invitation email for you to refer to. This Appendix will be included in the final version of the appraisal tool. Summary of comments from round-2: 1- Authors need to report subject demographics as this indicates						• is confusing and needs clarification; #3 and 4 are fine (a bit extensive); #5 is incomplete. • Consider changing to "Child-Pugh" Class What type of globulins would you like provided?

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	The difference in %	Median	Interquartile range	
• severity of illness • Cointerventions • Integrated hemodynamic status and vasopressor treatment • Outcomes In case of including participants suffering from renal failure the following information should be provided: • Cause • Plasma creatinine concentration/creatinine clearance • Plasma electrolytes • Hemoglobin concentration • Plasma protein level • Time and the nature of last dialysis • Existence of clinical edema	that the study is appropriately conducted based on the Inclusion/Exclusion criteria						

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	The difference in %	Median	Interquartile range	
• Existence of peripheral neuropathy In case of including participants suffering from hepatic cirrhosis the following information should be provided: • Cause • Child's Classification (34) • Prothrombin time, platelet count • Albumin and globulin levels In case of including participants suffering from Critically ill patients the following information should be provided: • Clinical description • Apache II score • Plasma creatinine and electrolyte Concentrations • Presence/absence of renal failure							

Question	Modified question	Consensus					Comments
		Percentage of agreement round-1	Percentage of agreement round-2	The difference in %	Median	Interquartile range	
• Presence/absence of liver failure In case of including participants suffering from Thermal injury the following information should be provided: • Regular hematocrit • Preoperative plasma albumin and globulin level In the case of Bioequivalence studies, the following criteria should be fulfilled: • Nonsmoker healthy volunteers (males/females) with a body weight that is ± 20% of the standard and with age between 18 to 55 years old should be enrolled as long as possible.							

*This question is added only to round-2 based on the comments of the panelist

4.2.3.3. Consensus through round-3

In round 3 of the modified Delphi process, 15/25 (60%) experts responded to the survey. Round 3 survey composed of 4 questions: the consent form sent again with 3 potential questions related to the appraisal process of clinical pharmacokinetics studies (Figures 8 and 9). Of 3 potential clinical pharmacokinetics critical appraisal tool items, all 3 items reached ≥80% consensus from experts for inclusion and were retained in Table 21.

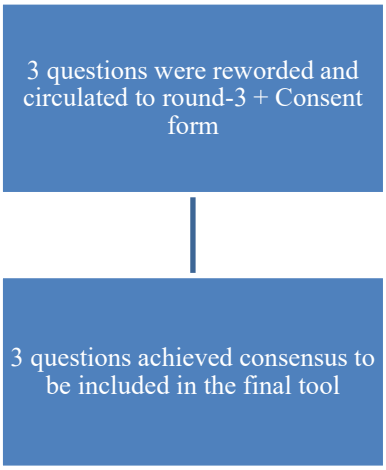

Figure 8. Flowchart of round-3 questions

Table 21: Questions that meet consensus for inclusion through round-3

Question	Consensus					Comments
	Percentage of agreement round-2	Percentage of agreement round-3	Difference in %	Median	Interquartile range	
Q1) Was the rationale behind the purpose of the study reasonable and comprehensible? Summary of comments from round-2: 1- Some readers might not be knowledgeable about the discussed drug, and they need a well-written background and rationale to appraise the quality of the study.	Agree: (8) 42.10% Neither Agree Nor Disagree: (2) 10.52% Disagree: (9) 47.36%	Agree: (13) 86.67% Neither Agree Nor Disagree: (0) 0% Disagree: (9) 13.33%	Agree: 44.57 Neither Agree Nor Disagree: -10.52% Disagree: -34.03%	4	0	
Q2) Was the described population pharmacokinetics approach validation method appropriate for the analysis? 1- Basic internal method 2- Advanced internal method 3- External model evaluation Note: Please refer to Appendix-1 Model Evaluation attached in the invitation email for further clarification. This appendix will be provided in the final version of the appraisal tool. Summary of comments from round-2: 1- Although this information might be beyond the level of reader's knowledge to judge if the used population	Agree: (11) 57.89% Neither Agree Nor Disagree: (5) 26.31% Disagree: (3) 15.78%	Agree: (15) 100% Neither Agree Nor Disagree: (0) 0% Disagree: (0) 0%	Agree: -42.11% Neither Agree Nor Disagree: -26.31% Disagree: -15.78%	4	1	

Question	Consensus					Comments
	Percentage of agreement round-2	Percentage of agreement round-3	Difference in %	Median	Interquartile range	
pharmacokinetics approach and validation are appropriate or not, it is important to be included.						
Q3) Where applicable, were the relevant baseline characteristics of the participants adequately described? Examples: sex, race, age, weight, height, concomitant disease, administered medications, smoking status, pregnancy, severity of illness that may affect pharmacokinetics parameters, renal function, and hepatic function. Note: This question will be further clarified by adding Appendix-2 Patient Demographics. Appendix-2 is attached in the invitation email for you to refer to. This Appendix will be included in the final version of the appraisal tool. Summary of comments from round-2: 1- Authors need to report subject demographics as this indicates that the study is appropriately conducted based on the Inclusion/Exclusion criteria.	Agree: (12) 70.58% Neither Agree Nor Disagree: (2) 11.76% Disagree: (3) 17.64%	Agree: (14) 93.33% Neither Agree Nor Disagree: (0) 0% Disagree: (1) 6.67 %	Agree: 22.75% Neither Agree Nor Disagree: -11.76% Disagree: -10.97%	5	1	

4.2.3.4. Selection of the rating scale

The selection of the final rating scale for the clinical pharmacokinetics critical appraisal tool was done based on the suggestions provided by the panelists. In round 1, we suggested 4 options like Yes, No, I Don't Know; 3-points Likert-scale (Good, Fair, Poor); 4-points Likert-scale (Excellent, Good, Poor, Fair) and 5-Point Likert-scale (Excellent, Above Average, Average, Below Average, Very Poor). A high percentage of panelists chose the use of a 4-point Likert-scale and Yes, No, I Don't Know. Additionally, Panelists suggested the addition of 'Not Applicable' as another option and comment box for the end-users to add their comments. One of the panelists also suggested to relate each question to a certain outcome like trustworthy, reliability and reproducibility of the results, sensitivity and specificity of the assay, whether the population is appropriate for the study, generalizability of the results to the population of interest, etc. Thus, these options were formulated and added to round 2 survey and sent to participants. An almost equal number of participants selected the last two options, Yes, No, I Don't Know and Not Applicable, and to relate each question to a certain outcome (Table 22). Based on discussion with the team members, the last two options should be combined together to put a rating scale for each question and to correlate each question or group of questions to a certain outcome to help end-users to come up with final conclusions. This idea came out after we were done with the modified Delphi process so we could not send out the idea to the participants again to achieve their consensus. Therefore, Yes, No, I Don't Know and Not Applicable was the final used rating scale in the developed clinical pharmacokinetics critical appraisal tool.

Table 22: Selection of the rating scale

Question	Options disclosed to round-1	Number of panelists (%)	Received comments from round-1	Options disclosed to round-2	Number of panelists (%)
Q26) In the final version of this proposed tool, which rating system do you feel would be the best method to help potential users appraise a clinical pharmacokinetics study?	Yes, No, I Don't know	26.32%	Suggestions for better options from the panelists: 1) Yes, partial yes, no	4 Point Likert-scale: Excellent, Good, Fair, Poor, Not applicable and a comment box for notes beside each question.	2 (10.53%)
	3 Point Likert-scale: Good, Fair, Poor	10.53%	2) Likert-scale would be very subjective - there do not recommend it. The information is either there or not.	Yes, No, I Don't Know, Not Applicable and a comment box for notes beside each question.	8 (42.11%)
	4 Point Likert-scale: Excellent, Good, Poor, Fair	31.58%	3) 4-point Likert-scale with additional column for "not applicable" plus include box for notes beside each question	Each question in the appraisal tool would relate to an outcome of whether or not the results are trust worthy, applicable to the population of interest, generalizable, etc. (e.g. reliability, reproducibility, sensitivity, specificity of the assay; whether the population is appropriate for the study, etc.). Upon completion of the appraisal tool, the user can look at the overall outcomes and decide on how to use/apply the study results.	9 (47.37%)
	5 Point Likert-scale: Excellent, Above Average, Average, Below Average, Very Poor	21.05%	4) I prefer AMSTAR rating system: Yes, No, Can't Answer, Not Applicable		
			5) Yes, No, Not described (with a comment box)		

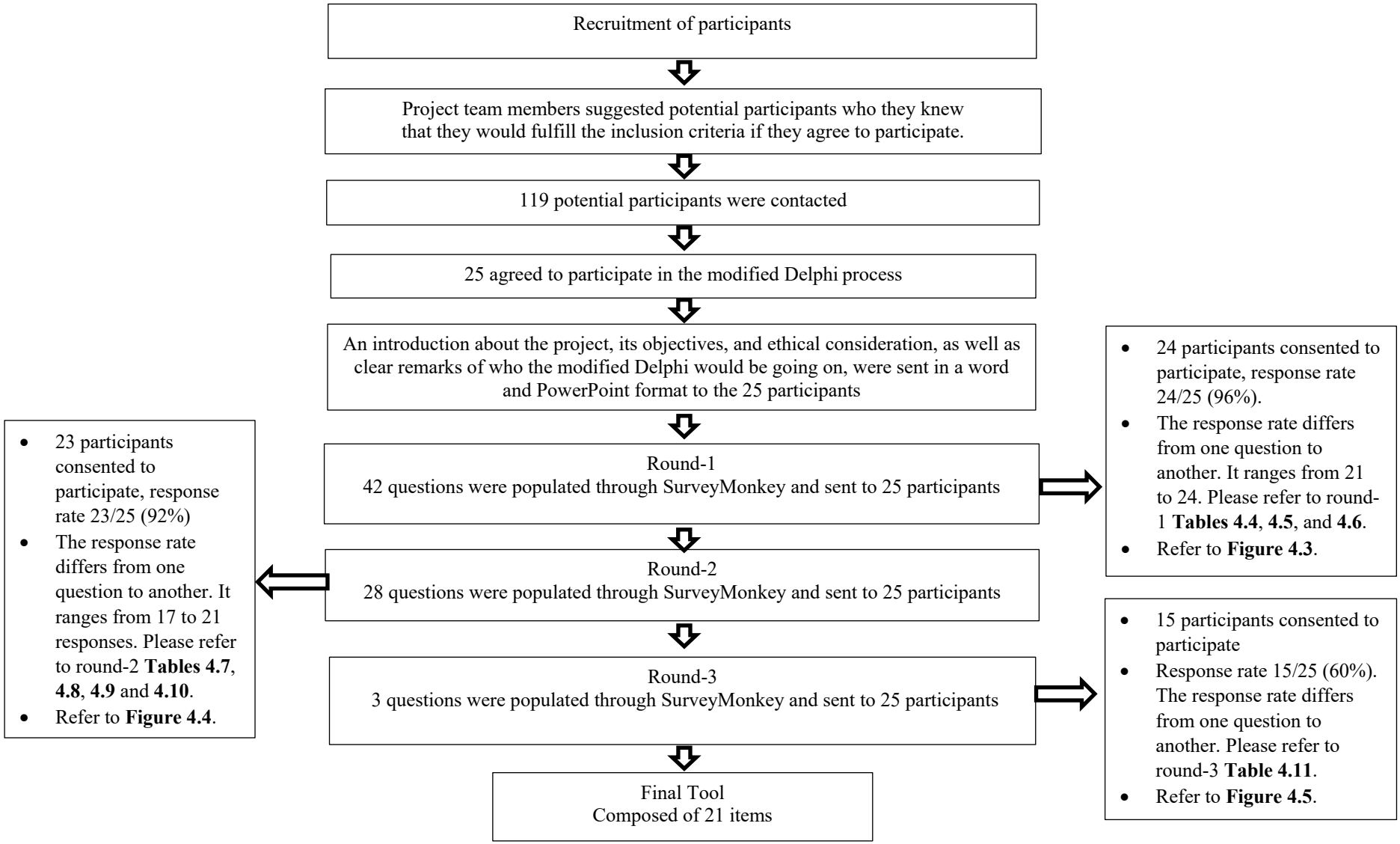

Figure 9. The modified Delphi flow chart of all rounds.

4.2.4. The final format of the clinical pharmacokinetics critical appraisal tool

Appendix G represents the draft of the final tool which is composed of 5 sections (background, design, sampling, applied statistics, and results) with a total of 21 questions. The final rating scale that was used to help the end-users to assess the quality of published clinical pharmacokinetics studies was “Yes, No, I Do Not Know and Not Applicable”. Furthermore, a comment box was added to help end-users to add any other comments.

4.3. Phase III results

In this phase, the validity and reliability of the developed clinical pharmacokinetics critical appraisal tool were assessed.

4.3.1. Results of the face and content validity

A psychometric expert from the College of Pharmacy at Qatar University revised the developed tool. Several modifications were suggested related to the used English and scientific language and formatting of the formulate the questions. Table 23 represents the proposed and applied adjustments. The provided comments were accepted or rejected after discussion with the project's main supervisor. The critical appraisal tool that was sent to the expert was provided in Appendix G. The modified version of the clinical pharmacokinetics critical appraisal tool was presented in Appendix H.

Table 23: Face and content validity results

Questions	Received comments from the psychometric experts	Modifications
Evaluator's name: _____ Name of the evaluated article: _____ Estimated time of evaluating the paper: _____ Score: _____/21_____	If there are "Not Applicable" items during the evaluation, will you account for these by adjusting the denominator? Example, if I marked 6 items as Not Applicable, should the total be out of 15 or still 21?	Evaluator's name: _____ Name of the evaluated article: _____ - Estimated time of evaluating the paper: _____ Score: _____/21_____
Clinical Pharmacokinetics Critical Appraisal Tool (CPKCAP)	I suggest using "Pharmacokinetics"	Clinical Pharmacokinetics Critical Appraisal Tool (CPKCAP)
Appraising Background	-	-
1-Was a clear description of the objectives of the study provided? Authors should provide a clear statement of the objectives of the research to clarify the purpose and the scope of the study.	No suggestions	1. Was a clear description of the objectives of the study provided? • Authors should provide a clear statement of the objectives of the research to clarify the purpose and the scope of the study.

Questions	Received comments from the psychometric experts	Modifications
2- Was a valid and comprehensive rationale provided to support the purpose of the study?	A scientific reviewer may sometimes be able to determine a clear rationale, but not necessarily its validity. So I will suggest changing “valid” to “clear”.	2. Was a clear and comprehensive rationale provided to support the purpose of the study?
Appraising Design	Study Design and Experimental	Appraising Study Design and Experimental Methods
3-Was the chosen study design appropriately selected and justified?	A design may be appropriately selected, but not justified by the author. If a design is well-known to be the best to address a certain research inquiry/question, then I may not have the space to justify it and it may be redundant. So I suggest to keep this as “appropriately selected”. Unless if you mean justified from the perspective of the reviewer (not the author). If the latter is the case, then I will add “from your perspective”	3. Was the chosen study design appropriately selected and justified?

Questions	Received comments from the psychometric experts	Modifications
4-Was the dosing (dose, route of administration, dosing interval) of the drug in the study justified for the intended study? Example: Authors should justify the use of single-dose versus steady-state analysis	1- I think it is better to be explicit here by writing it as (i.e. dose, route of administration, and dosing interval). 2- After reading the example, I think the criterion is too stringent. I will expect the investigators to declare the complete dosing approach (i.e. dose, route, interval), but not necessary to justify it all the times. I do not see it as necessary to justify the use. As an example, if I am making comparison between extended-interval dosing	4. Was the dosing (i.e. dose, route of administration, and dosing interval) of the drug in the study justified for the intended study? Examples: • Authors should justify the use of single-dose versus steady-state analysis.

Questions	Received comments from the psychometric experts	Modifications
	vs. multiple conventional dosing of aminoglycosides, do I need to justify? The literature about post-antibiotic effect associated with the extended dosing is well documented in the literature.	
5-Were the endpoints of the study appropriate to answer the objectives of the study?	1- outcome measures 2- I suggest “address the objectives”. Alternatively, you could use “answer the research question” or both (e.g. …. To address the research objectives or answer the research question	5. Were the outcome measures endpoints of the study appropriate to address the objectives of the study?

Questions	Received comments from the psychometric experts	Modifications
6-Were the exclusion criteria of participants included AND appropriate for the intended outcomes of the study? • The exclusion criteria should be relevant to assist with decreasing significant confounders (e.g. co-administration of drugs, organ impairment, and special populations) that may impact outcomes.		6. Were the exclusion criteria of participants included AND appropriate for the intended outcomes of the study? • The exclusion criteria should be relevant to assist with decreasing significant confounders (e.g. co-administration of drugs, organ impairment, and special populations) that may impact outcomes.
7-Where applicable, were the relevant baseline characteristics of the participants adequately described? Examples: sex, race, age, weight, height, concomitant disease, administered medications, smoking status, pregnancy, severity of illness that may affect pharmacokinetics parameters, renal function, and hepatic function. Note: Note: Please refer to Appendix-2 Patient Demographics for further clarification.	1- I suggest paraphrasing to "Were the relevant baseline characteristics of the study participants adequately described as applicable?" 2- Note is repeated twice. Consider deleting one of the two. 3- My concern with this item is that we do not describe the characteristics under the Design, but rather as part of the Results. So I wonder if this	7. Where applicable, were the relevant baseline characteristics of the participants adequately described? Examples: Sex, race, age, weight, height, concomitant disease, administered medications, smoking status, pregnancy, severity of illness that may affect pharmacokinetics parameters, renal function, and hepatic function. • Note: Please refer to Appendix-1 Patient Demographics for further clarification.

Questions	Received comments from the psychometric experts	Modifications
	criterion belongs to this domain.	
8-Were plausible interacting covariates described a priori or in post hoc evaluation? Examples: demographic variables, laboratory values, concomitant medications, and relevant disease states to the drug being studied.	1- The word a priori should be Italicize 2- Same concern as above.	8. Were plausible interacting covariates described *a priori* or in post hoc evaluation? Examples: Demographic variables, laboratory values, concomitant medications, and relevant disease states to the drug being studied.
9-Was the description of the used sample analysis methods or citations of prior validation studies provided in the publication or affiliated appendix? Example: Chromatography type. Detection type. Assay characteristics: mobile phase composition, gradient and flow rate, chromatographic column (packing material, dimensions). Analytical runtime. Operating temperature and detection parameters. Validation method: specificity, recovery, linearity and sensitivity, the stability of the assay and its reproducibility.	1- To avoid confusion with methodological or statistical sampling analysis, I will use the term "biological sample analytical methods". I think you get my message and can paraphrase as necessary. 2- Are these combined? Please check formatting carefully.	9. Was the description of the used biological sample analytical methods sample analysis methods or citations of prior validation studies provided in the publication or affiliated appendix? Examples: • Chromatography type. • Detection type. • Assay characteristics: mobile phase composition, gradient and flow rate, chromatographic column (packing material, dimensions). • Analytical runtime. • Operating temperature. • Detection parameters. • Validation method: specificity, recovery, linearity and sensitivity, the stability of the assay and its reproducibility.

Questions	Received comments from the psychometric experts	Modifications
	3- What is unclear to me is that, how much info is needed for me to evaluate this as Yes or No? In other words, what is considered "sufficient" or "insufficient" description?	
Appraising Sampling	This term needs to be further clarified. It may be interpreted as "sampling of subjects" or "analytical sampling – e.g. serum or blood samples".	
10- Was the method of data sampling appropriate for the study? Examples: first vs. second order absorption, and lag time. Evaluating for nonlinearity requires multiple dose levels and a complete profile is recommended. Researchers obtain these data from previously conducted studies with completed concentration-time profile (e.g. phase I studies). The method of data sampling should reference previously validated quantitative bioanalytical methods and if those are	This is really unclear to me, although I have some good understanding of clinical pharmacokinetics and TDM. Again, do you mean "sampling of subjects" or "sampling of analytics"? I do not feel the example was very helpful and would advise you to reflect on either clarifying/refining the	10. Was the method of data sampling of analytics appropriate for the study? Examples: • First vs. second order absorption, and lag time. • Evaluating for nonlinearity requires multiple dose levels and a complete profile is recommended. • Researchers obtain these data from previously conducted studies with completed concentration-time profile (e.g. phase I studies).

Questions	Received comments from the psychometric experts	Modifications
not available then the full description or defense of data sampling should be included.	question or the examples.	The method of data sampling should reference previously validated quantitative bioanalytical methods and if those are not available then the full description or defense of data sampling should be included.
11- Was a clear description of the sampling site and the sampling interval (the exact times at which samples are obtained) provided and justified? Example: Sampling site should be consistent for all subjects in the study. Arterial sampling is preferable during frequent sampling schedule. Arterial sampling is more representative of the delivered concentration to the effect site in the case of peripheral elimination. Arterial sampling is preferable when administering a drug that has a short duration of action or fast onset of action. Sampling interval should not exceed the expected half-life of the studied exponential phase (fast distribution, slow distribution and elimination).	1- To me, these are two independent parameters. The question is double-barreled as the authors may describe one but not the other. I suggest split this into two criteria. 2- May be nice bulleting will make your tool look more tidy and more appealing to users? Apply throughout the tool please.	11. Was a clear description of the sampling site provided and justified? Examples: • Sampling site should be consistent for all subjects in the study. • Arterial sampling is preferable during frequent sampling schedule. • Arterial sampling is more representative of the delivered concentration to the effect site in the case of peripheral elimination. • Arterial sampling is preferable when administering a drug that has a short duration of action or fast onset of action

Questions	Received comments from the psychometric experts	Modifications
12- Was the number of half-lives elapsed within the sampling period appropriate for the analyzed drug?	This is unclear to me? Are we referring to “reaching steady state (i.e. 3 – 5 half lives)? If this this the case, the criterion needs to be made clearer. Example: Was sampling initiated after the study drug reached steady state (i.e. 3 – 5 half lives)? I am not trying to undermine the question or mislead you, but just to say	12. Was the number of half-lives elapsed within the sampling period appropriate for the analyzed drug? Examples: Sampling interval should not exceed the expected half-life of the studied exponential phase (fast distribution, slow distribution and elimination).

Questions	Received comments from the psychometric experts	Modifications
	that you should try to make it clearer, if possible.	
13- Were sample storage conditions appropriate and described in a manner that could be accurately replicated? Examples: sample storage temperature, use and description of anticoagulants, stabilizers, centrifugation etc.	What about if they are appropriate, but not described in details or vice versa?	13. Were sample storage conditions appropriate and described in a manner that could be accurately replicated? Examples: Sample storage, temperature, use and description of anticoagulants, stabilizers, centrifugation etc.
14- If applicable, was there a clear description of the pharmacokinetics model, its development, validation and justification for use? It is recommended to provide the following details about the selected modeling process: Description of studies from which dataset was driven Model structure Validated software for the pharmacokinetics analysis Criteria for accepting valid model's parameters Fitting procedure defined prior to the initiation of the analysis. A reasonable assumption based on which the scheme for	1- I believe there are some punctuation problems here. Please check places where "commas" are needed.	14. If applicable, was there a clear description of the pharmacokinetics model, its development, validation and justification for use? It is recommended to provide the following details about the selected modeling process: • Description of studies from which dataset was driven • Model structure • Validated software for the pharmacokinetics analysis • Criteria for accepting valid model's parameters

Questions	Received comments from the psychometric experts	Modifications
weighting is considered to be appropriate and the transformation of data [e.g. logarithmic transformation to achieve the homoscedastic (constant) variance requirements] should be provided.		• Fitting procedure defined prior to the initiation of the analysis. • A reasonable assumption based on which the scheme for weighting is considered to be appropriate and the transformation of data [e.g. logarithmic transformation to achieve the homoscedastic (constant) variance requirements] should be provided.
14- If applicable, was there a clear description of the pharmacokinetics model, its development, validation and justification for use? It is recommended to provide the following details about the selected modeling process: Description of studies from which dataset was driven Model structure Validated software for the pharmacokinetics analysis Criteria for accepting valid model's parameters Fitting procedure defined prior to the initiation of the analysis. A reasonable assumption based on which the scheme for weighting is considered to be appropriate and the	2- I believe there are some punctuation problems here. Please check places where "commas" are needed.	14. If applicable, was there a clear description of the pharmacokinetics model, its development, validation and justification for use? It is recommended to provide the following details about the selected modeling process: • Description of studies from which dataset was driven • Model structure • Validated software for the pharmacokinetics analysis • Criteria for accepting valid model's parameters • Fitting procedure defined prior to the initiation of the analysis.

Questions	Received comments from the psychometric experts	Modifications
transformation of data [e.g. logarithmic transformation to achieve the homoscedastic (constant) variance requirements] should be provided.		• A reasonable assumption based on which the scheme for weighting is considered to be appropriate and the transformation of data [e.g. logarithmic transformation to achieve the homoscedastic (constant) variance requirements] should be provided.
15- Was the described population pharmacokinetics approach validation method appropriate for the analysis? 1- Basic internal method 2- Advanced internal method 3- External model evaluation Note: Please refer to Appendix-1 Model Evaluation for further clarification.		15. Was the described population pharmacokinetics approach validation method appropriate for the analysis? 1- Basic internal method 2- Advanced internal method 3- External model evaluation Note: Please refer to Appendix-2 Model Evaluation for further clarification.
16- Were the essential pharmacokinetics parameters required to make the results applicable in clinical settings addressed?	1- or included?	16. Were the essential pharmacokinetics parameters required to make the results applicable in clinical settings included?

Questions	Received comments from the psychometric experts	Modifications
Consider: Total clearance (CL), Fraction of dose excreted unchanged in urine (fe), Volume of distribution at steady state (Vss), Volume of distribution during the terminal phase (VZ), Blood/plasma concentration ratio, Terminal half-life ($t_{1/2}$ Z), Fraction of the unbound drug in plasma (fu), Bioavailable fraction of dose (F), Absorption rate constant (Ka),C_{min}, C_{max}, t_{max}, EC50, Ke0, Hill coefficient, or gamma, AUC and bioavailability for the two drug formulation	2- If examples are too many, they tend to distract the user. Unless if you think, you feel it is necessary to provide an exhaustive list.	Examples: • Total clearance (CL), Volume of distribution at steady state (Vss), Blood/plasma concentration ratio, Terminal half-life (t1/2 Z), Fraction of the unbound drug in plasma (fu), Absorption rate constant (Ka),Cmin, Cmax, tmax, , AUC, etc.
17- Were the pharmacokinetics equations used to calculate the patient's pharmacokinetics parameters disclosed or cited within the article? Example: creatinine clearance, body weight calculations, Michaelis Menten, volume of distribution, patient weight: total body weight vs. ideal body weight.	1- Presented 2- But these are not equations by themselves? So if I provide Sawchuk-Zaske equation or at least mention it, but did not mention the method of calculating Clcr, what should be the evaluation (Yes or No)? 3- Same as above. Do I have to declare equations for all common parameters	17. Were the pharmacokinetics equations used to calculate the patient's pharmacokinetics parameters presented or cited within the article? Examples: • Equations used to calculate the following pharmacokinetics parameters: creatinine clearance, body weight calculations, Michaelis Menten, volume of distribution

Questions	Received comments from the psychometric experts	Modifications
Appraising Applied Statistics 18- Were the chosen statistical tests and software to perform the statistical analysis appropriate to achieve the study objectives? 19- Were all patients enrolled in the study accounted for? Example: Description of patient screening, enrollment, run-in or wash out phases, study period and follow-up periods are adequately described. Any loss to follow-up or withdrawals are described.	including Clcr and $t_{1/2}$?	18. Were the chosen statistical tests and software to perform the statistical analysis appropriate to achieve the study objectives? 19. Were all patients enrolled in the study accounted for? Examples: • Description of patient screening, enrollment, run-in or wash out phases, study period and follow-up periods are adequately described.

Questions	Received comments from the psychometric experts	Modifications
		Any loss to follow-up or withdrawals are described.
20- In the event of missing data or outliers, was the process for analysis justified and appropriate?		20. In the event of missing data or outliers, was the process for analysis justified and appropriate?
21- Were appropriate summary statistics to describe centrality and variance used to document the pharmacokinetics results? Example: Descriptive statistics such as confidence interval, standard deviation, mean, median, range, interquartile range, standard error and trimmed rang	present?	21. Were appropriate summary statistics to describe centrality and variance used to present the pharmacokinetics results? Examples: • Descriptive statistics such as confidence interval, standard deviation, mean, median, range, interquartile range, standard error and trimmed range

4.3.2. Reliability testing

4.3.2.1. Assessment of the interrater agreement for clinical pharmacokinetics critical appraisal tool

4.3.2.1.1. Assessment of the interrater agreement per question for clinical pharmacokinetics critical appraisal tool

Tables 27 and 28 represent the Kappa values and percentage of agreement on each item included in the final clinical pharmacokinetics critical appraisal tool. Each of the following items 13, 16, 18, and 21 scored a Kappa of an almost substantial agreement to perfect agreement at 0.815, 0.783, 0.651, and 1, respectively, and a percentage of agreement of more than 90% percentage of agreement. Furthermore, items 1, 4, 6, 7, 8, 9, 11,12, 14, 17, 19 , and 20 scored a Kappa of fair agreement to moderate agreement at 0.474, 0.38, 0.605, 0.225, 0.360, 0.211, 0.609, 0.255, 0.351, 0.552 0.340, and 0.348, respectively and percentage of agreement between 70% to 93%, *p*-value ≤ 0.05 except question 7, 9, and 12 their *p*-value were not significant. Items 1,4, and 20 scored paradox values of Kappa 0.474, 0.38, and 0.348, respectively despite having a high percentage of agreement at 93.3%, 83.3%, and 83.3%, respectively. Paradox values of Kappa occur due the effect of prevalence due to the skewness of the answers toward either Yes or No and the unbalance totals of the marginals. Therefore, the prevalence index and the bias index were calculated through using *Equations 4 and 5*, respectively, to determine their effects on questions 1, 4, and 20, presented in Tables 24, 25, and 26. The prevalence index and the bias index for question 1 were 0.86, and 0.0667, respectively. The prevalence index and the bias index for question 4 were 0.7, and 0.16, respectively. The prevalence index and the bias index for question 20 were 0.7, and 0.033, respectively. The values of Kappa for the three questions were corrected through using *Equation 6*.

Table 24: Question 1 in the clinical pharmacokinetics critical appraisal tool

	Observer K		
Observer O	Yes	No	Total
Yes	1	2	3
No	0	27	27
Total	1	29	30

Table 25: Question 4 in the clinical pharmacokinetics critical appraisal tool

	Observer K		
Observer O	Yes	No	Total
Yes	2	5	7
No	0	23	23
Total	2	28	30

Table 26: Question 20 in the clinical pharmacokinetics critical appraisal tool

	Observer K		
Observer O	Yes	No	Total
Yes	23	2	25
No	3	2	5
Total	26	4	30

The corrected values of Kappa for questions 1, 4, and 20 are 0.86, 0.667, and 0.667, respectively. Items 3 and 10 scored a Kappa of less than the chance of agreement at -0.098 and -0.017, and percentage of an agreement at 80% and 60%, respectively. Item15 scored a Kappa of less than a chance of agreement to slight agreement at 0.195, and percentage of an agreement at 53.33%. Item 3 scored relatively low Kappa; however, the percentage of agreement on this item was 80%. Items 2 and 5 have no Kappa value despite having a percentage of agreement of 100 % and 96.65%, respectively. This was due to the skewness of the distribution of the reviews, that is, higher number of raters selected the score 'yes' (item 2), 'yes' (item 5), respectively. Item 3 had a negative value of Kappa which shows strong disagreement between raters despite that the percentage of agreement was 80%. This occurs because high number of raters selected 'yes' (item 3). On the other hand, item 10 had negative values of Kappa and consistently the percentage of agreement on this item was 60% therefore both Kappa and percentage of agreement show that there was disagreement between raters on question 10. There is no enough information about negative values of Kappa available in the literature to understand the cause behind its occurrences.

Table 27: Assessment of the interrater agreement per question for clinical pharmacokinetics critical appraisal tool

Type of measurement	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11
Kappa Value	K 0.474	-	K -0.098	K 0.38	-	K 0.605	K 0.255	K 0.360	K 0.211	K -0.017	K 0.609
% of agreement	28/30*100 = 93.3%	30/30*100 = 100%	24/30*100 = 80%	25/30*100 = 83.33%	29/30*100 = 96.67%	24/30 * 100 = 80%	22/30*100=73.33%	21/30*100= 70%	21/30*100 = 70%	18/30*100= 60%	27/30*100= 90%
p-value	0.002*	-	0.568	0.008*	1	0.001*	0.900	0.035*	0.232	0.900	0.001*

*significant *p-value* ≤ 0.05

Table 27: continued

Type of measurement	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	Q21
Kappa Value	K 0.255	K 0.815	K 0.351	K 0.195	K 0.783	K 0.552	K 0.651	K 0.340	K 0.348	K 1
% of agreement	22/30*100 = 73.33%	28/30*100 = 93.33%	22/30*100= 73.33%	16/30*100= 53.33%	29/30*100= 96.67%	24/30*100 = 80%	29/30*100= 96.67%	23/30*100= 76.67%	25/30*100= 83.33%	30/30*100= 100%
p-value	0.163	0.000*	0.028*	0.134	0.000*	0.000*	0.000*	0.035*	0.055*	0.000*

*significant *p-value* ≤ 0.05

Table 28: Level of agreement per question for clinical pharmacokinetics critical appraisal tool

Value of Kappa	Level of agreement	Questions
Less than 0	Less than a chance of agreement	Q3, Q10
0.01 - 0.20	Slight agreement	Q15
0.21 - 0.40	Fair agreement	Q4, Q7, Q8,Q9,Q12, Q14, Q19, Q20
0.41 - 0.60	Moderate agreement	Q1,Q6, Q11, Q17
0.61 - 0.80	Substantial agreement	Q16, Q18
0.81 – 0.99	Almost perfect agreement	Q13, Q21

4.3.2.1.2. Assessment of the interrater agreement per paper for clinical pharmacokinetics critical appraisal tool

Tables 29 and 30 represent the Kappa values and percentage of agreement of 4 raters who appraise 30 selected papers using the developed clinical pharmacokinetics critical appraisal tool. Each of the following papers P3, P4, P6, P9, P12, P15, P17, P18, P19, P21, P22, P25, and P27 scored a Kappa of almost substantial agreement to perfect agreement at 0.859, 0.897, 0.667, 0.829, 0.632, 0.859, 0.699, 0.64, 0.696, 0.615, 0.667, 0.774 and 0.632, respectively, more than 85% percentage of agreement and p-value < 0.005. Papers P1, P5, P7 , P10, P11,P13, P20, P23, P24, P26, P28 and P29 scored a Kappa of fair agreement to moderate agreement at 0.538, 0.417, 0.481, 0.588, 0.475, 0.323, 0.314, 0.357, 0.444, 0.222 and 0.533, respectively and percentage of agreement > 70%. Papers P2, P8, P14 and P16 scored a Kappa of slightly agreement at 0.176, 0.152, 0.152, and 0.173, and percentage of agreement at 61.90%, 61.90%, 61.90%, and 76.19%, respectively. It has been noticed that raters were able to differentiate between articles of high quality and articles of low quality as raters gave scores ranges from 2 to 20 out of 21. This proves the validity and reliability of the developed tool. Additionally, raters concluded that this tool could not be used to critically appraise reviews; however, it could be used to appraise prospective, retrospective clinical pharmacokinetics studies, population pharmacokinetics studies, bioequivalence and drug interaction clinical pharmacokinetics studies.

Table 29: Level of agreement per paper for clinical pharmacokinetics critical appraisal tool

Value of Kappa	Level of agreement	Papers
Less than 0	Less than chance of agreement	
0.01 - 0.20	Slight agreement	P2, P8, P14, P16
0.21 - 0.40	Fair agreement	P7, P20, P23, P24, P28
0.41 - 0.60	Moderate agreement	P1, P5, P10, P11 P13, P26, P29
0.61 - 0.80	Substantial agreement	P6, P12, P17, P18, P19, P21, P22, P25, P27, P30
0.81 - 0.99	Almost perfect agreement	P3, P4, P9, P15

Table 30: Assessment of the interrater agreement per paper for clinical pharmacokinetics critical appraisal tool

Paper number	Name of the article	Rater-1	Rater - 2	Time by rater-1	Time by rater-2	Kappa value	*p*-value	Percentage of agreement
1	A randomized, placebo-controlled, single ascending-dose study to assess the safety, tolerability, pharmacokinetics, and immunogenicity of subcutaneous tralokinumab in Japanese healthy volunteers	SP	KJ	16 min	15 min	K 0.538	0.011*	17/21*100 = 80.95%
2	Infliximab Pharmacokinetics are Influenced by Intravenous Immunoglobulin Administration in Patients with Kawasaki Disease	OR	KJ	20 min	15 min	K 0.176	0.375	13/21*100 = 61.90%
3	An Open-Label Crossover Study of the Pharmacokinetics of the 60-mg Edoxaban Tablet Crushed and Administered Either by a Nasogastric Tube or in Apple Puree in Healthy Adults	OR	KJ	45 min	10 min	K 0.859	0.000*	20/21*100 = 95.238%
4	Identification of Cytochrome P450-Mediated Drug–Drug Interactions at Risk in Cases of Gene Polymorphisms by Using a Quantitative Prediction Model	SP	KJ	23 min	10 min	K 0.897	0.000*	20/21* 100 = 95.238%

*significant *p*-value ≤ 0.05

Paper number	Name of the article	Rater-1	Rater - 2	Time by rater-1	Time by rater-2	Kappa value	*p*-value	Percentage of agreement
5	Pharmacokinetics, Safety and Tolerability of Oral Semaglutide in Subjects with Renal Impairment	OR	KJ	30 min	15 min	K 0.417	0.019*	17/21*100 = 80.95%
6	Phase I Clinical Study of ZYAN1, A Novel Prolyl-Hydroxylase (PHD) Inhibitor to Evaluate the Safety, Tolerability, and Pharmacokinetics Following Oral Administration in Healthy Volunteers	OR	KJ	90 min	10 min	K 0.667	0.002*	18/21*100= 85.71%
7	Population pharmacokinetics and exposure–response modeling and simulation for evolocumab in healthy volunteers and patients with hypercholesterolemia	OR	KJ	50 min	15 min	K 0.386	0.075	16/21*100 = 76.19%
8	Population Pharmacokinetics Analysis of Alirocumab in Healthy Volunteers or Hypercholesterolemic Subjects Using a Michaelis–Menten Approximation of a Target-Mediated Drug Disposition Model—Support for a Biologics License Application Submission: Part I	SP	KJ	21 min	15 min	K 0.152	0.368	13/21*100 = 61.904%

*significant *p*-value ≤ 0.05

Paper number	Name of the article	Rater-1	Rater - 2	Time by rater-1	Time by rater-2	Kappa value	*p*-value	Percentage of agreement
9	Pharmacokinetics of the B-Cell Lymphoma 2 (Bcl-2) Inhibitor Venetoclax in Female Subjects with Systemic Lupus Erythematosus	OR	KJ	20 min	15 min	K 0.829	0.000*	20/21*100 = 95.2380%
10	Population Pharmacokinetics of Volasertib Administered in Patients with Acute Myeloid Leukaemia as a Single Agent or in Combination with Cytarabine	OR	KJ	45 min	15 min	K 0.481	0.022*	17/21*100 = 80.95%
11	A Population Pharmacokinetics and Pharmacodynamic Analysis of Abemaciclib in a Phase I Clinical Trial in Cancer Patients	OR	KJ	60 min	15 min	K 0.588	0.003*	18/21*100 = 85.7142%
12	Higher Midazolam Clearance in Obese Adolescents Compared with Morbidly Obese Adults	OR	KJ	30 min	15min	K 0.632	0.004*	18/21*100 = 85.7142%
13	Pharmacokinetics Optimization of Everolimus Dosing in Oncology: A Randomized Crossover Trial	SP	KJ	24 min	15 min	K 0.475	0.030*	17/21*100=80.95%
14	Population pharmacokinetics–pharmacodynamics of oral everolimus in patients with seizures associated with tuberous sclerosis complex	SP	KJ	16 min	15 min	K 0.152	0.475	13/21 *100 = 61.90%

*significant *p*-value ≤ 0.05

Paper number	Name of the article	Rater-1	Rater - 2	Time by rater-1	Time by rater-2	Kappa value	*p*-value	Percentage of agreement
15	Population Pharmacokinetics and Optimal Sampling Strategy for Model-Based Precision Dosing of Melphalan in Patients Undergoing Hematopoietic Stem Cell Transplantation	OR	KJ	75 min	15 min	K 0.859	0.000*	20/21*100 = 95.23%
16	Effects of Mild to Severe Hepatic Impairment on the Pharmacokinetics of Sonidegib: A Multicenter, Open-Label, Parallel-Group Study	SP	AS		55 min	K 0.173	0.361	16/21*100 =76.1904%
17	Pharmacokinetics of MHAA4549A, an Anti-Influenza A Monoclonal Antibody, in Healthy Subjects Challenged with Influenza A Virus in a Phase IIa Randomized Trial	SP	AS	18 min	50 min	K 0.690	0.001*	18/21*100 = 85.7142%
18	Clinical Pharmacokinetics and Dose Recommendations for Posaconazole in Infants and Children	SP	AS	17 min	40 min	K 0.64	0002*	18/21*100 = 85.7142%
19	Population Pharmacokinetics Modeling of Olaratumab, an AntiPDGFRa Human Monoclonal Antibody, in Patients with Advanced and/or Metastatic Cancer	OR	AS	45 min	35 min	K 0.696	0.001*	19/21*100 = 90.47%

*significant *p*-value ≤ 0.05

Paper number	Name of the article	Rater-1	Rater - 2	Time by rater-1	Time by rater-2	Kappa value	*p*-value	Percentage of agreement
20	Pharmacokinetics of dexmedetomidine during analgosedation in ICU patients	SP	AS	21 min	30 min	K 0.323	0.129	18/21*100 = 85.71%
21	Pharmacokinetics of ADS-5102 (Amantadine) Extended Release Capsules Administered Once Daily at Bedtime for the Treatment of Dyskinesia	OR	AS	30 min	50 min	K 0.615	0.004*	17/21*100 = 80.95%
22	Effect of Semaglutide on the Pharmacokinetics of Metformin, Warfarin, Atorvastatin and Digoxin in Healthy Subjects	SP	AS	18 min	30 min	K 0.667	0.002*	18/21*100 = 85.7142%
23	Population Pharmacokinetics Modeling of JNJ-53718678, a Novel Fusion Inhibitor for the Treatment of Respiratory Syncytial Virus: Results from a Phase I, Double-Blind, Randomized, Placebo-Controlled First-in-Human Study in Healthy Adult Subjects	SP	AS	23 min	30 min	K 0.314	0.115	16/21*100 = 76.1904%
24	Characterization of the Pharmacokinetics of Vilaprisan: Bioavailability, Excretion, Biotransformation, and Drug–Drug Interaction Potential	SP	AS	16 min	25 min	K 0.357	0.102	15/21*100= 71.4285%

*significant *p*-value ≤ 0.05

Paper number	Name of the article	Rater-1	Rater - 2	Time by rater-1	Time by rater-2	Kappa value	*p*-value	Percentage of agreement
25	Piperacillin Population Pharmacokinetics and Dosing Regimen Optimization in Critically Ill Children with Normal and Augmented Renal Clearance	OR	AS	30 min	40 min	K 0.774	0.000*	20/21*100=95.2380%
26	Safety and Tolerability of Intravenous Valproic Acid in Healthy Subjects: A Phase I Dose-Escalation Trial	OR	AS	60 min	32 min	K 0.444	0.040*	16/21*100 = 76.1904%
27	Population Pharmacokinetics of GemtuzumabOzogamicin in Pediatric Patients with Relapsed or Refractory Acute Myeloid Leukemia	SP	AS	27 min	25 min	K 0.632	0.0004*	18/21*100 = 85.7142%
28	Clinical Pharmacokinetics and Mass Balance of Veliparib in Combination with Temozolomide in Subjects with Nonhematologic Malignancies	SP	AS	22 min	25 min	K 0.222	0.292	15/21*100 = 71.4285%
29	Model-Based Therapeutic Drug Monitoring of Infliximab Using a Single Serum Trough Concentration	SP	AS	26 min	23 min	K 0.533	0.015	17/21*100 = 80.9523%
30	The Ontogeny of UDP-glucuronosyltransferase Enzymes, Recommendations for Future Profiling Studies and Application Through Physiologically Based Pharmacokinetics Modelling	OR	AS	30 min	17 min	K 0.774	0.000*	20/21*100= 95.2380%

4.3.2.2. Assessment of the intra-rater agreement for clinical pharmacokinetics critical appraisal tool

The four raters had a Kappa value of the substantial agreement to almost perfect agreement after critically appraising 5 papers two times with a gap of 14 days between the two times as presented in Table 31. The overall intra-rater agreement was substantial which shows that the developed tool is stable over time.

Table 31: Assessment of the intra-rater agreement for clinical pharmacokinetics critical appraisal tool

Raters	Value of Kappa	Level of agreement
Rater-1	K 0.631	Substantial agreement
Rater-2	K 0.924	Almost perfect agreement
Rater-3	K 0.745	Substantial agreement
Rater-4	K 0.905	Almost perfect agreement
Overall agreement	K 0.802	Substantial agreement

4.3.3. Assessment of the feasibility of using clinical pharmacokinetics critical appraisal tool

The use of clinical pharmacokinetics critical appraisal tool proved easy to apply, as it took an average of 28.5 minutes. **Table 32** represents the average time required by each rater to appraise the 30 articles using a clinical pharmacokinetics critical appraisal tool.

Table 32: Assessment of the feasibility of using clinical pharmacokinetics critical appraisal tool

Raters	Average time
Rater-1	34 minutes
Rater-2	21 minutes
Rater-3	14 minutes
Rater-4	45 minutes
Mean	28.5 minutes

CHAPTER 5: DISCUSSION AND CONCLUSION

5.1. Introduction

One of the main goals of Qatar's national health vision by 2022 is to practice evidence-based medicine (EbM). Clinicians practice EbM through tailoring medical decisions for each patient by integrating their clinical skills and expertise with the best available up-to-date research (2). The critical appraisal process is central to the practice of EbM. Critical appraisal is used to thoroughly evaluate published articles. Different types of appraisal tools were developed, like design-specific and generic critical appraisal tools (9). However, the published appraisal tools are not highly specific to determine the methodological quality and validity of clinical pharmacokinetics studies. The validity of the study varies based on the methodological criteria that usually differ based on the study design. Nevertheless, most of the available critical appraisal tools lack the presence of items that help in analyzing the published articles in depth (63).

The application of pharmacokinetics in the clinical setting is considered to be an integral part of providing pharmaceutical care services that are delivered by the pharmacist (65). Patient outcomes are improved by the appropriate application of the clinical pharmacokinetics principles from the published evidence-based information to decrease the following events: mortality, morbidity, length of treatment and hospital stay (LOS), adverse effects of medications, and economic burden (66-70). Despite the importance of applying pharmacokinetics principles in a clinical setting, researchers found, through conducting studies to assess the quality of reporting clinical pharmacokinetics studies, that the reporting quality of clinical pharmacokinetics was low. The first study was a systematic review which evaluated the quality of reporting of pharmacokinetics studies of antibiotics in patients with

sepsis receiving continuous renal replacement therapy. In this systematic review, it was found that all of the identified articles did not report the required information that was essential for end-users to interpret the reported results. Furthermore, it was noticed that 20% of the published pharmacokinetics trials did not contain the fundamental pharmacokinetics parameters (71). Consequently, reporting guidelines for clinical pharmacokinetics studies (The ClinPK Statement) were issued to assess and guide researchers while reporting their clinical pharmacokinetics studies. A Yes/No checklist composed of 24 items was developed to guide researchers while writing their research and ensuring the reporting of the minimum required information in the published clinical pharmacokinetics studies (72). While offering a valuable guideline for reporting findings of clinical pharmacokinetics studies, the ClinPK statement guidelines did not cover all the dimensions of quality of trials including design, conduct, analysis, clinical relevance, quality of reporting, and results validity. Therefore, there is a gap in the knowledge, as there is no available critical appraisal tool that aids clinicians in appraising and determining the quality of the published clinical pharmacokinetics studies.

This project was, therefore, conducted to gain a better understanding of quality markers of clinical pharmacokinetics studies and to develop a specific critical appraisal tool for clinical pharmacokinetics studies. To date, this is the first study conducted to identify clinical pharmacokinetics studies quality markers and to develop specific critical appraisal tool for these types of research. The project was accordingly designed to address the following primary question: What are the quality markers of clinical pharmacokinetics studies needed to develop clinical pharmacokinetics critical appraisal tool to help end-users to appraise the quality of the published articles in this field?

5.2. Explanation and exploration

The developed clinical pharmacokinetics critical appraisal tool is the first of its kind to assess the quality of published clinical pharmacokinetics studies via consensus by key pharmacokinetics stakeholders. This tool guides researchers through answering 21 questions to determine the quality of different types of published clinical pharmacokinetics studies like retrospective and prospective clinical pharmacokinetics studies, population pharmacokinetics, bioequivalence, and drug interaction studies. This tool was developed in format similar to other critical appraisal tools like a measurement tool to assess systematic reviews (AMSTAR) to facilitate the appraising process. Clinical pharmacokinetics critical appraisal tool is composed of four sections: appraising background, appraising study design and experimental methods, appraising applied statistics, and appraising results. There is an importance for each of the included items in the checklist.

Item 1: Was a clear description of the objectives of the study provided?

The main goal behind including this item was to ensure the provision of a clear statement about the research objectives and to clarify the purpose and the scope of the conducted study (72).

Item 2: Was a clear and comprehensive rationale provided to support the purpose of the study?

As in the ClinPK statement, this item was added because the rational is one of the leading quality markers that help in assessing the quality of the published study by evaluating the value of the new generated information to the existing knowledge in the filed (72).

Item 3: Was the chosen study design appropriately selected and justified?

One of the leading quality cornerstones is the selection of the study design to help in achieving the study objectives. There are several determinants based on which researchers have to select suitable study designs like drug administration route, mode, and schedule method, site, and timing of obtaining biological sampling (85). For example, crossover design is preferable when researchers are planning to compare new drugs with gold standard ones. This type of study design is preferred in case of studying the effect of short-lived and reversible medications on treating symptomatic chronic diseases. On the other hand, this study design is not recommended to study unstable conditions (82).

Item 4: Was the dosing (i.e., dose, route of administration, and dosing interval) of the drug in the study justified for the intended study?

Describing the dose of the investigated medication, dosing interval and route of administration is highly essential as this will affect the pharmacokinetics profile of the studied medication and the generalizability of the results (72).

Item 5: Were the outcome measures endpoints of the study appropriate to address the objectives of the study?

Item 6: Were the exclusion criteria of participants included AND appropriate for the intended outcomes of the study?

Item 7: Where applicable, were the relevant baseline characteristics of the participants adequately described?

The answers to item 5, 6 and 7 will affect the external validity and the generalizability of the final results on the population who have the same characteristics of the included participants in the study. Item 6 was added because the exclusion criteria is an essential element through which readers could determine the quality of a published article. The reported exclusion criteria should be relevant to assist with decreasing

significant confounders (e.g. co-administration of drugs, organ impairment, and special populations) that may impact outcomes; otherwise, bias might be introduced. This item was not included in the ClinPK statement. Item 7 was added because the following participant's characteristics: sex, race, age, weight, height, concomitant disease, administered medications, smoking status, pregnancy, the severity of illness, renal function, and hepatic function have high impact on the pharmacokinetics parameters. Therefore, including them in an article is highly relevant to assess the quality of the article. These items were not mentioned in the ClinPK statement reporting guidelines, and instead the researchers are asked if the "Eligibility criteria of study participants are described" which is not enough to assess the quality of an article (72).

Item 8: Were plausible interacting covariates described a priori or in post hoc evaluation?

Covariates like demographic variables, laboratory values, concomitant medications, and relevant disease states to the drug being studied affect the studied pharmacokinetics parameters. End-users should evaluate if the mentioned covariates in the study were clinically relevant and statistically significant (89) .

Item 9: Was the description of the used biological sample analytical methods sample analysis methods or citations of prior validation studies provided in the publication or affiliated appendix?

We elaborated on this item by providing an example of a bioanalytical method and how it should be described; for example, chromatography and the detection type. Additionally, we described the assay characteristics: mobile phase composition, gradient and flow rate, chromatographic column (packing material, dimensions), analytical runtime, operating temperature, and detection parameters. We

recommended specific elements that the end-users should look for to assess the validation process that was conducted by the researchers like specificity, recovery, linearity and sensitivity, stability of the assay, and its reproducibility. This question was mentioned from the reporting point of view in the ClinPK statement, but the previously mentioned elaboration and examples of what the end-users should look for to assess the quality of what is exactly reported were missing in the ClinPK statement (72).

Item10: Was the method of data sampling of analytics appropriate for the study?

There are three types of sampling schedule:

- Sparse sampling: this method is commonly used in population pharmacokinetics studies. Three to five samples are collected from each patient.
- Traditional sampling: this method is commonly used in pharmacokinetics studies. After one to two minutes from injecting the medications, researchers start to collect 10 to 20 blood samples.
- Early intensive sampling: this method is characterized by collecting many numbers of samples for the first minutes then the samples should be collected in intervals similar to the traditional sampling. The midpoint of timing since the collection of the blood samples started should be used to perform pharmacokinetics calculations. This method leads to an accurate estimation of peak concentration (86).

Therefore sampling methods should be aligned with the study design as well as the study goals. This is one of the questions that differentiate the critical appraisal tool from the ClinPK reporting checklist.

Item 11: Was a clear description of the sampling site provided and justified?

End-users should consider the following points while evaluating the selected sampling site:

- The sampling site should be consistent for all subjects in the study.
- Arterial sampling is preferable during frequent sampling schedule.
- Arterial sampling is more representative of the delivered concentration to the effect site in the case of peripheral elimination.
- Arterial sampling is preferable when administering a drug that has a short duration of action or a fast onset of action.
- Samples should be collected from the opposite arm if the medication is administered intravenously to avoid any contamination.
- Researchers should mention any case of deviation occurred while collecting the samples (85).

Considering all these points, the answer to this question will guide the end-user to decide on the validity, reliability, and generalizability of the results on a similar population.

Item 12: Was the number of half-lives elapsed within the sampling period appropriate for the analyzed drug?

This item is unique and differentiates this critical appraisal tool from the ClinPK statement reporting guidelines (72). Researchers should collect samples within a time that covers more than 3 half-lives within the terminal phase of disposition. The sampling interval should not exceed the expected half-life of the studied exponential phase (fast distribution, slow distribution, and elimination) (84). From the recorded time points, end-users will be able to determine if the results generated from extensive

sampling are available. Additionally, these time points will help in understanding and determining pharmacokinetics parameters like clearance (84).

Item 13: Were sample storage conditions appropriate and described in a manner that could be accurately replicated?

Describing the storage conditions of the samples is highly critical because this will affect the analysis of the pharmacokinetics parameters. End-users should evaluate the selected anticoagulant and determine if it did not interfere with the analyzed medication or its metabolites. Furthermore, end-users should assess whether researchers added stabilizers to the blood collecting tubes prior to collecting samples and if the cooled the collected sample before adding the stabilizer (86).

Item 14: If applicable, was there a clear description of the pharmacokinetics model, its development, validation, and justification for use?

Item 15: Was the described population pharmacokinetics approach validation method appropriate for the analysis?

These two questions were specifically related to population pharmacokinetics since it was known that there was no acceptable modeling method. Thus, in the examples below, it was suggested what components of the description should be provided by the researchers. Additionally, different validation methods were explained and provided in the attached appendix so users who do not have enough knowledge to appraise this part can understand after reading the provided appendix. These two questions differentiate the clinical pharmacokinetics critical appraisal tool from the ClinPK statement reporting guidelines (72).

Item 16: Were the essential pharmacokinetics parameters required to make the results applicable in clinical settings included?

It is highly essential to report the fundamental pharmacokinetics parameters like total clearance (CL), volume of distribution at steady state (Vss), blood/plasma concentration ratio, terminal half-life ($t_{1/2}$ Z), fraction of the unbound drug in plasma (fu), absorption rate constant (Ka), C_{min}, C_{max}, t_{max}, AUC, etc. Reporting the fundamental parameters will help end-users to interpret the reported results (71).

Item 17: Were the pharmacokinetics equations used to calculate the patient's pharmacokinetics parameters presented or cited within the article?

Several equations could be used to calculate each of the previously listed variables including the Cockroft–Gault equation or the Modification of Diet in Renal Disease (MDRD) formula to calculate creatinine clearance or glomerular filtration rate. There are many variations of the previously mentioned equations. Thus the out product of these formulas might be different; therefore, researchers should disclose the formula that they used to grantee the external validity and generalizability of the study results (72).

Item 18: Were the chosen statistical tests and software to perform the statistical analysis appropriate to achieve the study objectives?

The statistical tests should be selected based on the "Principle of Statistical Analysis of Clinical Test (1998)." Researchers should mention the method chosen and disclose the details in the appendix. Researchers should clearly state the used software to analyze the data and to handle the outliers (54). Additionally, the selection of specific software depends mainly on the study problem as well as the experience of the researchers (86).

Item 19: Were all patients enrolled in the study accounted for?

Item 20: In the event of missing data or outliers, was the process for analysis justified and appropriate?

Researchers should precisely describe patient screening, enrollment, run-in or wash out phases, study period, any loss to follow-up or withdrawals. Additionally, researchers should determine a priori how they are going to manage missing data since it represents a source of bias. This is because deleting participants with missing data leads to the imprecise estimation of the pharmacokinetics parameters. Therefore, the best option, in this case, is to impute the missing data by calculating the mean, median or mode of uncorrelated data. Additionally, the maximum likelihood procedures like deriving regression models could be used to determine each predictor (56).

The definition of the outlier is arbitrary. Therefore, researchers should clearly define an outlier in their protocol from a statistical point of view. If researchers did not specify the exact method of dealing with the outliers due to the exploratory nature of the population pharmacokinetics studies, they should exclude the outliers from the data set before building the model and restrict the conclusion to the defined population after removing the outlier (56).

Item 21: Were appropriate summary statistics to describe centrality and variance used to present the pharmacokinetics results?

Descriptive statistics such as confidence interval, standard deviation, mean, median, range, interquartile range, standard error, and trimmed range of drug concentrations and pharmacokinetics parameters should be calculated using a suitable statistical method to the selected study design (54).

5.3. Validity and reliability testing

This clinical pharmacokinetics critical appraisal tool was developed to evaluate the

validity and reliability of published clinical pharmacokinetics studies. Therefore, the developed clinical pharmacokinetics critical appraisal tool should pass through the same validation process to ensure that it is effective, valid and reliable in distinguishing between poor and good quality published clinical pharmacokinetics studies.

5.3.1. Validity testing

The content and face validity were evaluated, and modifications were applied based on the comments received from an expert opinion like what was done in the development of the tool to assess the cognitive skills of evidence-based practice in student health professionals. The convergent validity of this tool was not tested due to the absence of a gold standard tool to compare with the newly developed critical appraisal tool.

5.3.2. Reliability testing

After testing the reliability of the developed clinical pharmacokinetics critical appraisal tool. Eight questions (Q1, Q6, Q11, Q13, Q17, Q16, Q18, and Q21) achieved a Kappa value of moderate to almost perfect agreement and *p-value* ≤ 0.005. However, 9 questions (Q4, Q7, Q8, Q9, Q12, Q14, Q15, Q19, and Q20) achieved a range of Kappa value within less than a chance of agreement to a fair agreement. Kappa values for Q2 and Q5 were not detectable. Additionally, we got negative values of Kappa for Q3 and Q10. Thus, the percentage of agreement was calculated for all of the questions to compare to the Kappa values. This is because it was noticed that Kappa values were un-representatively low due to the skewedness of the ratings toward one of the following categories ‘Yes’ or ‘No’ only. This effect was known as a prevalence or striking paradox and it occurred in this study due to the nature of the used coding system (140). It has been reported in one of the studies that the highest

value of Kappa was inspected when Pe has the smallest value with fixed values of Po. The paradox is detectable with large values of Pe as the large values of Pe would be converted into small values of K in the correction process (141). For example, in this study in question 4, when two observers, K and O, selected a binary rating scale Yes / No, the results appear in 2x2 table (Table 33) as following:

Table 33: Question 4 in the clinical pharmacokinetics critical appraisal tool

	Observer K		
Observer O	Yes	No	Total
Yes	2	5	7
No	0	23	23
Total	2	28	30

The percentage of observed agreement in this case:

Chance-corrected observed agreement (Po) calculated by using Equation (2):

Po = (2+23)/30= 0.833

Chance-corrected perfect agreement (Pe) calculated by using Equation (3):

Pe = $(2*7 + 28*23)/30^2$=0.731

Thus, Kappa calculated through using Equation (1):

K= (0.833 - 0.731)/(1 – 0.733) = 0.382

Therefore, Kappa values could be 2 folds higher from one situation to another for different values of Pe and fixed values of Po. Thus, Kappa values for questions 1, 4,

and 20 were corrected by using the prevalence-adjusted-bias-adjusted Kappa to 0.86, 0.667, and 0.667, respectively. The percentage of agreement of Q2 and Q5 was more than 80%, but their Kappa values were undetectable. Thus, the five (Q1, Q2, Q4, Q5, Q20) questions proved to be reliable and were included in the final tool without modifications. Compared with the other instruments, like AMSTAR, this striking paradox occurred in question 4 and 7 (127). On the other hand, the percentage of agreement of the other seven questions Q7, Q8, Q9, Q12, Q14, Q15, and Q19 that achieved a range of Kappa values within less than a chance of agreement to a fair agreement, was below 80%. Landis and Koch agreed that Kappa values less than 0 were unaccepted (135). All of the 7 questions had Kappa values more than zero so the percentage of agreement will be ignored in this case as a measurement of reliability. This phenomenon happened because the included raters have different backgrounds with 3 faculty members and 1 Masters' student. The mathematical formulation of Kappa considers the correction of chance agreement. Therefore, Kappa value can detect the error and the true score in every observable measurement, which is not the case by using the percentage of agreement. The quality of an instrument improved through increasing its reliability and this happened through reducing the measurement error, thus Kappa values were used instead of the percentage of agreement (136). Additionally, the comments that were provided by the four reviewers were revised to determine if any of them provided any comment that would help in clarifying these questions, rewording them or removing them from the final draft. There were no comments provided by the 4 reviewers that would help in modifying these questions. Thus, the questions will be included in the final draft without modifications, this is similar to what researchers had done while testing the reliability of AMSTAR in question 7 (127). It is recommended to assess the reliability of these questions in

future studies using more than 30 studies. On the other hand, questions 3 and 10 may require further revision and modifications in future studies because they have negative values of Kappa which show that there was disagreement between raters and their percentages of agreement were not acceptable.

There was a moderate to almost perfect inter-rater agreement on 20 papers out of 30 critically appraised papers. However, there was a slight to a fair agreement on other 10 papers (P2, P7, P8, P14, P16, P20, P23, P24, P26, and P28). It has been noticed that raters were able to differentiate between articles of high quality and articles of low quality as raters gave scores ranging from 2 to 20 out of 21. This proves the validity and reliability of the developed tool. Additionally, raters concluded that this tool could not be used to critically appraise review articles; however, it could be used to appraise prospective, retrospective clinical pharmacokinetics studies, population pharmacokinetics studies, bioequivalence and drug interaction clinical pharmacokinetics studies.

Questions were written to help authors to critically appraise the quality of the published clinical pharmacokinetics articles. Below each question, there was a description to help end-users to appraise the article and to give them a hint about what they have to look for to answer the question. However, end-users were given the opportunity to appraise the article based on their knowledge and to determine if all the required information was provided by the author to generalize the results in their setting from their point of view. For instance, in question 16: Were the essential pharmacokinetics parameters required to make the results applicable in clinical settings included? The raters who we selected represent the end-users of the tool and they have different experiences in different settings, which may be why each rater appraised some articles differently.

5.3.3. Intra-rater agreement

All of the 4 raters had Kappa values between substantial agreement to almost perfect agreement which proved the stability of the measure throughout time. We recommended to test the intra-rater reliability on a larger sample size in future studies.

5.4. Feasibility testing

The feasibility was assessed by measuring the average time needed by all of the four raters to complete the clinical pharmacokinetics critical appraisal tool. End-users would need 28.5 minutes to complete this tool which is more than the required time to complete AMSTAR but less than the time needed to complete Sacks' instrument 34.4 minutes tools to assess the quality of systematic reviews (127).

5.5. Comparison between clinical pharmacokinetics critical appraisal tool and reporting checklist

Firstly, the aim of developing the clinical pharmacokinetics critical appraisal tool was to assess the quality of studies, which has several dimensions including study design, conduct, analysis, clinical relevance, results validity and quality of reporting. The ClinPK statement reporting guidelines focused only on assessing the quality of reporting without determining if the used methods were of high quality or not. Assessing the quality of reporting was common between the two tools; therefore, there was a cross-matching between both in some questions. Although there are cross-matching questions between the two tools, the way in which the questions were formulated was different. For example, question 18 in our critical appraisal tool: Were the chosen statistical tests and software to perform the statistical analysis appropriate to achieve the study objectives? Versus item 15 in the ClinPK checklist: "statistical methods including software used as described"(72). In our tool, we are looking for the appropriateness of the used statistical test not only the description because sometimes

the description is provided but the wrong test is selected.

Furthermore, in our critical appraisal tool, we guided end-users by giving them examples about points that they should consider when evaluating whether the methods used by researchers was of high quality. For instance, question 15: Was the described population pharmacokinetics approach validation method appropriate for the analysis? We put bullet points about different approaches of validation, and we provided an appendix to describe the use of these approaches. Therefore, this will ensure that end-users will have the same level of understanding of the question regardless of their level of knowledge about clinical pharmacokinetics. Thus, we expect that end-users will be able to justify their final decision about the quality of the published articles in clinical pharmacokinetics. Additionally, end-users will not leave a question due to lack of knowledge or misunderstanding which might happen while using the reporting checklist due to lack examples compared to those provided in our critical appraisal tool.

5.6. Study strengths and limitations

To our knowledge, this is the first study to identify quality markers of clinical pharmacokinetics studies to develop specific critical appraisal tool for these studies. The findings of this study highlight the quality markers of clinical pharmacokinetics based on pharmacokinetics stakeholders' consensus and based on these quality markers; a critical appraisal tool was developed. One of the main strengths of this study was the robustness of the methodology used to develop clinical pharmacokinetics critical appraisal tool, which comprehensively helped in addressing the study objectives. Firstly, a systematic review was conducted through searching different electronic databases like Embase and PubMed to identify quality markers related to clinical pharmacokinetics studies. Secondly, a modified Delphi was used to

achieve expert consensus to develop a final draft of the clinical pharmacokinetics critical appraisal tool. In the modified Delphi process, we were keen to recruit participants who represent all clinical pharmacokinetics stakeholders (clinicians, researchers, people who work in the academic and industrial sector and policymakers). The inclusion of different stakeholders allowed us to enrich our tool with different perspectives from different end-users. Although the modified Delphi process was anonymous for the research team, the provided comments and suggestions from panelists could hint to their background. For instance, the inclusion of clinicians and policymakers helped us to consider most of the questions and pharmacokinetics parameters that can help evaluate the quality of published clinical pharmacokinetic studies and apply in their practice setting. Question 17: Were the pharmacokinetic equations used to calculate patient pharmacokinetic parameters disclosed or cited within the article? was added based on the panelists' comments. Additionally, panelists who are working in the industrial sector and academia sector helped in improving questions that assess the methodological quality. For example, question 15: Was the described population pharmacokinetics approach validation method appropriate for the analysis? Panelists were divided around including this question in the final tool. Clinicians disagreed on including this question because they mentioned that this is beyond the end-users' level of knowledge. On the other hand, panelists with industrial and academic backgrounds support the inclusion of this question as it assesses the methodological quality study and they suggested the addition of Appendix F as an elaboration for other users.

Furthermore, the consensus criteria that we used were very strong as it depended on three elements: percentage of agreement, median, and the interquartile range. Keeping the participants' identity anonymous from each other gave the opportunity for each of

them to express their opinion regarding each question and prevent the dominance effect from taking place during the process. Finally, assessing the validity and reliability of the developed clinical pharmacokinetics critical appraisal tool ensured that it was robust.

The development of a clinical pharmacokinetics critical appraisal tool adds much value to the ClinPK statement reporting guideline as it complements the checklist by assessing the other dimensions of quality of clinical pharmacokinetics study. This useful tool can be used in the future by clinicians and policymakers to evaluate the quality of published articles to take clinical decisions and develop policies by applying EbM. Furthermore, stakeholders who work in the academic sector can use it to teach students how to appraise this type of study.

Despite these strengths, we need to acknowledge the limitations of this study. First, the participants' identification was anonymous from us as researchers, which prevented us from excluding participants who agreed with every single question without reading it. However, this did not affect the cutoff point of the consensus since we removed 4 participants from the total of people who selected agree. We then calculated the percentage of agreement of every question included in the final tool after round-1 and each met the consensus cut-off point. Furthermore, we received different response rates for every round of the survey. The response rate declined in round-3 despite sending reminders through email to encourage and remind participants about the importance of their participation to us and to fulfill the gap in knowledge. Additionally, in phase III, raters were asked to re-evaluate 5 articles only due to time constraints. Although we got a high value of Kappa that proves that the developed tool was stable over time, this needs to be re-evaluated in future studies. As based on a review of sample size requirements for the design of the reliability study,

the same number of articles used to evaluate the inter-rater reliability should be re-evaluated by the raters to calculate the intra-rater reliability (143). Additionally, the minimum required sample size is 30 comparisons to avoid having confidence interval resulting in no agreement (132),

5.7. Future work and recommendations

There are implications from research, clinical, policy, and education perspectives for this study. Our findings exposed issues related to assessing the quality of clinical pharmacokinetics studies through developing critical appraisal tool for these types of studies. This critical appraisal tool maybe used by clinicians in all clinical settings to appraise clinical pharmacokinetics studies to make sure that they apply EbM of high quality. Secondly, researchers may choose to use this tool as a guide in addition to ClinPK statement reporting guidelines to assess the quality of their written manuscripts and attached supplements before submitting to the targeted journal. Additionally, this tool maybe used to teach students in academic sectors the aspects that they have to look for to assess the quality of clinical pharmacokinetics studies. This tool maybe used by journal editors to assess the quality of the articles to be published. Healthcare policymakers may use this tool to evaluate the quality of studies based on which they are going to make their decision of adding certain medication to the drug formulary. Future studies need to be continuously conducted to modify and improve the currently developed tool so it will evolve with changes in the literature to maintain the practical benefits, since the pharmacokinetics field is dynamic and in continuous advancement.

5.8. Conclusions

This study aimed to determine quality markers for the appraisal of clinical pharmacokinetics studies based on the available literature. We then sought to achieve

expert consensus regarding the identified quality markers of clinical pharmacokinetics studies and to assess the psychometric properties of the developed tool. Through a systematic review of the literature, quality markers were identified within subcategories relating to aspects of high quality in a clinical pharmacokinetics study. These quality markers were used to formulate quality-related questions aimed at appraising clinical pharmacokinetics studies. Through expert consensus over multiple rounds of the modified Delphi process, a final list of questions was developed that allowed for appraisal of quality of a clinical pharmacokinetics study. The psychometrics of the final critical appraisal pharmacokinetics tool were assessed to ensure that the results generated from the tool were valid and reliable. Work presented in this thesis provides the first critical appraisal tool for clinical pharmacokinetics studies.

REFERENCES

1. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. Bmj. 1996;312(7023):71-2.

2. Egger M. Evidence-Based Medicine. In: Wright JD, editor. International Encyclopedia of the Social & Behavioral Sciences (Second Edition). Oxford: Elsevier; 2015. p. 327-31.

3. Burls A. What Is Critical Appraisal? Hayward Medical Communications. Oxford.[Google Scholar]; 2009.

4. Last JM, Harris SS, Thuriaux MC, Spasoff RA. A dictionary of epidemiology: International Epidemiological Association, Inc.; 2001.

5. Abdullah M, Firmansyah MA. Critical appraisal on journal of clinical trials. Acta Med Indones. 2012;44(4):337-43.

6. Young JM, Solomon MJ. How to critically appraise an article. Nature Clinical Practice Gastroenterology & Hepatology. 2009;6(2):82-91.

7. Glasziou PP, Irwig L, Bain CJ, Colditz GA. How to use the evidence: assessment and application of scientific evidence. 2000.

8. Crombie IK, Harvey BJ. The pocket guide to critical appraisal: a handbook for health care professionals. Canadian Medical Association Journal. 1997;157(4):448.

9. Katrak P, Bialocerkowski AE, Massy-Westropp N, Kumar VSS, Grimmer KA. A systematic review of the content of critical appraisal tools. BMC medical research methodology. 2004;4(1):22.

10. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol. 2007;7:10.

11. Public Health Resource Unit (2008) C*ritical Appraisal Skills Programme (CASP)* [Available from: http://www.phru.nhs.uk/Pages/PHD/CASP.htm.

12. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. Jama. 2000;283(15):2008-12.

13. O'Brien P, Silagy C, McCallum J, O'Connell D, Glasziou P, Hill A. How to use the evidence: Assessment and application of scientific evidence. National Health and Medical Research Council, Canberra, Commonwealth of Australia. 2000:1-84.

14. Joanna Briggs I. Protocols & work in progress. 2008 Available: http://www. joannabriggs. edu. au/pubs/systematic_reviews_prot. php. Accessed; 2009.

15. Greenhalgh T. How to read a paper: Papers that summarise other papers (systematic reviews and meta-analyses). Bmj. 1997;315(7109):672-5.

16. Auperin A, Pignon JP, Poynard T. critical review of meta-analyses of randomized clinical trials in hepatogastroenterology. Alimentary pharmacology & therapeutics. 1997;11(2):215-25.

17. Barnes DE, Bero LA. Why review articles on the health effects of passive smoking reach different conclusions. Jama. 1998;279(19):1566-70.

18. Beck CT. Use of meta-analysis as a teaching strategy in nursing research courses. Journal of Nursing Education. 1997;36(2):87-90.

19. Carruthers SG, Larochelle P, Haynes RB, Petrasovits A, Schiffrin EL. Report of the Canadian Hypertension Society Consensus Conference: 1. Introduction. CMAJ: Canadian Medical Association Journal. 1993;149(3):289.

20. Oxman AD, Guyatt GH, Singer J, Goldsmith CH, Hutchison BG, Milner RA, et al. Agreement among reviewers of review articles. Journal of clinical epidemiology. 1991;44(1):91-8.

21. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Int J Surg. 2010;8(5):336-41.

22. Smith AF. An analysis of review articles published in four anaesthesia journals. Canadian journal of anaesthesia. 1997;44(4):405.

23. L'AbbÉ KA, Detsky AS, O'Rourke K. Meta-analysis in clinical research. Annals of internal medicine. 1987;107(2):224-33.

24. Mulrow CD. The medical review article: state of the science. Annals of internal medicine. 1987;106(3):485-8.

25. Cook DJ, Sackett DL, Spitzer WO. Methodologic guidelines for systematic reviews of randomized control trials in health care from the Potsdam Consultation on Meta-Analysis. Journal of clinical epidemiology. 1995;48(1):167-71.

26. Cranney A, Tugwell P, Shea B, Wells G. Implications of OMERACT outcomes in arthritis and osteoporosis for cochrane metaanalysis. The Journal of rheumatology. 1997;24(6):1206-7.

27. Guyatt GH, Sackett DL, Sinclair JC, Hayward R, Cook DJ, Cook RJ, et al. Users' guides to the medical literature: IX. A method for grading health care recommendations. Jama. 1995;274(22):1800-4.

28. Gyorkos TW, Tannenbaum TN, Abrahamowicz M, Oxman AD, Scott EAF, Millson ME, et al. An approach to the development of practice guidelines for community health interventions. Canadian Journal of Public Health. 1994;85(1):S8.

29. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Oncology Research and Treatment. 2000;23(6):597-602.

30. Oxman AD, Cook DJ, Guyatt GH, Bass E, Brill-Edwards P, Browman G, et al. Users' guides to the medical literature: VI. How to use an overview. Jama. 1994;272(17):1367-71.

31. Pogue J, Yusuf S. Overcoming the limitations of current meta-analysis of randomised controlled trials. The Lancet. 1998;351(9095):47-52.

32. Irwig L, Tosteson ANA, Gatsonis C, Lau J, Colditz G, Chalmers TC, et al. Guidelines for meta-analyses evaluating diagnostic tests. Annals of internal medicine. 1994;120(8):667-76.

33. FOCUS Critical Appraisal Tool: A Manual for SIGN Guideline Developers [Available from: http://www.focus project.org.uk.

34. Continuing Professional Development: A Manual for SIGN Guideline Developers [Available from: http://www.sign.ac.uk.

35. Learning and Development Public Health Resources Unit [Available from: http://www.phru.nhs.uk/.

36. Verhagen AP, de Vet HCW, de Bie RA, Kessels AGH, Boers M, Bouter LM, et al. The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. Journal of clinical epidemiology. 1998;51(12):1235-41.

37. DerSimonian R, Charette LJ, McPeek B, Mosteller F. Reporting on methods in clinical trials. New England Journal of Medicine. 1982;306(22):1332-7.

38. Clark O, Castro Filho AA, Jv DB. Interrater agreement of Jadad's scale. Cochrane. 2001;1:op031.

39. Oremus M, Wolfson C, Perrault A, Demers L, Momoli F, Moride Y. Interrater reliability of the modified Jadad quality scale for systematic reviews of Alzheimer's disease drug trials. Dementia and geriatric cognitive disorders. 2001;12(3):232-6.

40. Sindhu F, Carpenter L, Seers K. Dovelopment of a tool to rate the quality assessment of randomized controlled trials using a Delphi technique. Journal of advanced nursing. 1997;25(6):1262-8.

41. Moher D, Schulz KF, Altman DG, Consort G. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. Elsevier; 2001.

42. Carey TS, Boden SD. A critical guide to case series reports. LWW; 2003.

43. Greenhalgh T, Taylor R. How to read a paper: papers that go beyond numbers (qualitative research). BMj. 1997;315(7110):740-3.

44. Mays N, Pope C. Assessing quality in qualitative research. Bmj. 2000;320(7226):50-2.

45. Evans M, Pollock AV. A score system for evaluating random control clinical trials of prophylaxis of abdominal surgical wound infection. British journal of surgery. 1985;72(4):256-60.

46. Charnock D. The DISCERN handbook. Quality criteria for consumer health information on treatment choices Radcliffe: University of Oxford and The British Library. 1998.

47. Wagner JG. History of pharmacokinetics. Pharmacol Ther. 1981;12(3):537-62.

48. Bauer LA. Clinical Pharmacokinetic and Pharmacodynamic Concepts. Applied Clinical Pharmacokinetics, 3e. New York, NY: McGraw-Hill Medical; 2015.

49. Rowland M, Tozer TN. Clinical pharmacokinetics: concepts and applications: Lippincott Williams & Wilkins; 1980.

50. Ducharme MP. Drug Elimination, Clearance, and Renal Clearance. In: Shargel L, Yu ABC, editors. Applied Biopharmaceutics & Pharmacokinetics, 7e. New York, NY: McGraw-Hill Education; 2016.

51. Tam VH. Application of Pharmacokinetics to Clinical Situations. In: Shargel L, Yu ABC, editors. Applied Biopharmaceutics & Pharmacokinetics, 7e. New York, NY: McGraw-Hill Education; 2016.

52. Tozer TN, Rowland M. Introduction to pharmacokinetics and pharmacodynamics: the quantitative basis of drug therapy: Lippincott Williams & Wilkins; 2006.

53. Yue CS, Ducharme MP. Empirical Models, Mechanistic Models, Statistical Moments, and Noncompartmental Analysis. In: Shargel L, Yu ABC, editors. Applied Biopharmaceutics & Pharmacokinetics, 7e. New York, NY: McGraw-Hill Education; 2016.

54. Clinical pharmacokinetic studies of pharmaceuticals. 2001 [Available from: http://www.nihs.go.jp/phar/pdf/ClPkEng011122.pdf.

55. Aarons L. Population pharmacokinetics: theory and practice. British journal of clinical pharmacology. 1991;32(6):669-70.

56. Sun H, Fadiran EO, Jones CD, Lesko L, Huang SM, Higgins K, et al. Population pharmacokinetics. A regulatory perspective. Clinical Pharmacokinetics. 1999;37(1):41-58.

57. Ette EI, Williams PJ. Population pharmacokinetics I: background, concepts, and models. Ann Pharmacother. 2004;38(10):1702-6.

58. Committee ICHS. Studies in Support of Special Populations: Geriatrics. ICH harmonized tripartite guideline. 1993.

59. Akondi RB, Chodavarapu RK, Devarakonda KR. Population Pharmacokinetics of Amikacin in Indian Pediatric Patients. Clinical Research and Regulatory Affairs. 2008;25(3):173-82.

60. Karlsson MO, Sheiner LB. The importance of modeling interoccasion variability in population pharmacokinetic analyses. Journal of pharmacokinetics and biopharmaceutics. 1993;21(6):735-50.

61. Sheiner LB, Beal SL. Evaluation of methods for estimating population pharmacokinetic parameters. III. Monoexponential model: routine clinical pharmacokinetic data. Journal of pharmacokinetics and biopharmaceutics. 1983;11(3):303-19.

62. Chow S-C. Bioavailability and Bioequivalence in Drug Development. Wiley interdisciplinary reviews Computational statistics. 2014;6(4):304-12.

63. Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, et al. Evaluating non-randomised intervention studies. Health technology assessment (Winchester, England). 2003;7(27):iii-x.

64. Crowe M, Sheppard L. A review of critical appraisal tools show they lack rigor: Alternative tool structure is proposed. Journal of Clinical Epidemiology. 2011;64(1):79-89.

65 Cies JJ, Varlotta L. Clinical pharmacist impact on care, length of stay, and cost in pediatric cystic fibrosis (CF) patients. Pediatric pulmonology. 2013;48(12):1190-4.

66. Ensom MHH, Davis GA, Cropp CD, Ensom RJ. Clinical pharmacokinetics in the 21st century. Clinical pharmacokinetics. 1998;34(4):265-79.

67. American Society of Health-System P. American Society of Health-System Pharmacists Statement on the Pharmacist's Role in Substance Abuse Prevention,

Education, and Assisstance. Washington, DC: American Society of Health-System Pharmacists. 2015.

68. Murphy R, Chionglo M, Dupuis LL. Impact of a pharmacist-initiated therapeutic drug monitoring consult service for children treated with gentamicin. The Canadian Journal of Hospital Pharmacy. 2007;60(3).

69. Shenfield GM. Therapeutic drug monitoring beyond 2000. British journal of clinical pharmacology. 1998;46(2):93-4.

70. Touw DJ, Neef C, Thomson AH, Vinks AA. Cost-effectiveness of therapeutic drug monitoring: a systematic review. Therapeutic drug monitoring. 2005;27(1):10-7.

71. Li A, Gomersall CD, Choi G, Tian Q, Joynt GM, Lipman J. A systematic review of antibiotic dosing regimens for septic patients receiving continuous renal replacement therapy: do current studies supply sufficient data? Journal of antimicrobial chemotherapy. 2009;64(5):929-37.

72. Kanji S, Hayes M, Ling A, Shamseer L, Chant C, Edwards DJ, et al. Reporting Guidelines for Clinical Pharmacokinetic Studies: The ClinPK Statement. Clinical Pharmacokinetics. 2015;54(7):783-95.

73. The Qatar National

Cancer Strategy [Available from: http://www.nhsq.info/app/media/880.

74. Tan T, Al-Khalaqi A, Al-Khulaifi N. Qatar national vision 2030. Sustainable Development: An Appraisal from the Gulf Region. 2014;19:65.

75. Jüni P, Altman DG, Egger M. Systematic reviews in health care: assessing the quality of controlled clinical trials. BMJ [Internet]. 2001 [cited 2011 Jan 21]; 323 (7303): 42-6.

76. Elwood M. Critical appraisal of epidemiological studies and clinical trials: Oxford University Press; 2017.

77. Haynes RB, Sackett DL, Richardson WS, Rosenberg W, Langley GR. Evidence-based medicine: How to practice & teach EBM. Canadian Medical Association Journal. 1997;157(6):788.

78. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS medicine. 2009;6(7):e1000100.

79. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC medical research methodology. 2007;7(1):10.

80. Higgins JPT, Green S. Cochrane handbook for systematic reviews of interventions: John Wiley & Sons; 2011.

81. Soliman A, Pawluk S, Wilby K. Identification of quality markers for clinical pharmacokinetic studies: a systematic review. . PROSPERO 2018

82. Mills EJ, Chan A-W, Wu P, Vail A, Guyatt GH, Altman DG. Design, analysis, and presentation of crossover trials. Trials. 2009;10:27.

83. Sager JE, Yu J, Ragueneau-Majlessi I, Isoherranen N. Physiologically based pharmacokinetic (PBPK) modeling and simulation approaches: A systematic review of published models, applications, and model verification. Drug Metabolism and Disposition. 2015;43(11):1823-37.

84. SOP 14: Population pharmacokinetic analysis. Onkologie. 2003;26 Suppl 6:60-6.

85. SOP 13: Pharmacokinetic data analysis. Onkologie. 2003;26 Suppl 6:56-9.

86. Viby-Mogensen J, Ostergaard D, Donati F, Fisher D, Hunter J, Kampmann JP, et al. Pharmacokinetic studies of neuromuscular blocking agents: Good clinical research practice (GCRP). Acta Anaesthesiologica Scandinavica. 2000;44(10):1169-90.

87. Vaara S, Pettila V, Kaukonen KM. Quality of pharmacokinetic studies in critically ill patients receiving continuous renal replacement therapy. Acta anaesthesiologica Scandinavica. 2012;56(2):147-57.

88. Sherwin CMT, Kiang TKL, Spigarelli MG, Ensom MHH. Fundamentals of population pharmacokinetic modelling: Validation methods. Clinical Pharmacokinetics. 2012;51(9):573-90.

89. Vozeh S, Steimer JL, Rowland M, Morselli P, Mentré F, Balant LP, et al. The use of population pharmacokinetics in drug development. Clinical Pharmacokinetics. 1996;30(2):81-93.

90. Gotta V, Dao K, Rodieux F, Buclin T, Livio F, Pfister M. Guidance to develop individual dose recommendations for patients on chronic hemodialysis. Expert review of clinical pharmacology. 2017;10(7):737-52.

91. Sun HE, Fadiran EO, Jones CD, Lesko L, Huang S-M, Higgins K, et al. Population pharmacokinetics. Clinical pharmacokinetics. 1999;37(1):41-58.

92. Nakai K, Fujita M, Ogata H. International harmonization of bioequivalence studies and issues shared in common. Yakugaku zasshi : Journal of the Pharmaceutical Society of Japan. 2000;120(11):1193-200.

93. Bonate PL, Ahamadi M, Budha N, de la Pena A, Earp JC, Hong Y, et al. Methods and strategies for assessing uncontrolled drug-drug interactions in population pharmacokinetic analyses: results from the International Society of Pharmacometrics (ISOP) Working Group. J Pharmacokinet Pharmacodyn. 2016;43(2):123-35.

94. Dykstra K, Mehrotra N, Tornøe CW, Kastrissios H, Patel B, Al-Huniti N, et al. Reporting guidelines for population pharmacokinetic analyses. Journal of Clinical Pharmacology. 2015;55(8):875-87.

95. Wade JR, Edholm M, Salmonson T. A guide for reporting the results of population pharmacokinetic analyses: a Swedish perspective. The AAPS journal. 2005;7(2):45.

96. Schoonenboom J, Johnson RB. How to Construct a Mixed Methods Research Design. Kolner Z Soz Sozpsychol. 2017;69(Suppl 2):107-31.

97. Creswell J. Research design: Qualitative, quantitative, and mixed methods approaches. . Thousand Oaks, CA: Sage Publications; 2014. 127 p.

98. Powell C. The Delphi technique: myths and realities. J Adv Nurs. 2003;41(4):376-82.

99. Delbecq AL, Van de Ven AH. A group process model for problem identification and program planning. The Journal of Applied Behavioral Science. 1971;7(4):466-92.

100. Van de Ven AH, Delbecq AL. The nominal group as a research instrument for exploratory health studies. American journal of public health. 1972;62(3):337-42.

101. Waggoner J, Carline JD, Durning SJ. Is there a consensus on consensus methodology? Descriptions and recommendations for future consensus research. Academic Medicine. 2016;91(5):663-8.

102. Lopez-Olivo MA, Suarez-Almazor ME. Developing guidelines in musculoskeletal disorders. Clinical and experimental rheumatology. 2007;25(6):S28.

103. Harrison LL. Health Care Research by Degrees. Journal of Advanced Nursing. 1993;18(11):1849-.

104. Rowe G, Wright G. The Delphi technique as a forecasting tool: issues and analysis. International journal of forecasting. 1999 Oct 1;15(4):3 75.

105. Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CF, Askham J, et al. Consensus development methods, and their use in clinical guideline development. Health Technol Assess. 1998;2(3):i-iv, 1-88.

106. Hasson F, Keeney S. Enhancing rigour in the Delphi technique research. . Technological Forecasting and Social Change. 2011 Nov 1;7 8(9):695.

107. Avella RJ. Delphi panels: Research design, procedures, advantages, and challenges. I nternational Journal of Doctoral Studies. 2016 11(1):321

108. Nair R, Aggarwal R, Khanna D. Methods of formal consensus in classification/diagnostic criteria and guideline development. Semin Arthritis Rheum. 2011;41(2):95-105.

109. Barrios M, Villarroya A, Borrego A, Olle C. Response rates and data quality in web and mail surveys administered to PhD holders. Social Science Computer Review. 2011 May;29(2):220.

110. SurveyMonkey LLC. SurveyMonkey®. Palo Alto (CA): SurveyMonkey, LLC. 2012.

111. Allen PJ, Roberts LD. The ethics of outsourcing online survey research. International Journal of Technoethics 2010 Jul 1;1(3):48.

112. Culley JM. Use of a computer-mediated Delphi process to validate a mass casualty conceptual model. Comput Inform Nurs. 2011;29(5):272-9.

113. Adler M, Ziglio E. Gazing into the oracle: The Delphi method and its application to social policy and public health. : Jessica Kingsley Publishers; 1996. 252 p.

114. Cresswell JW, Plano Clark VL. Designing and conducting mixed method research. 2nd Sage. Thousand Oaks, CA. 2011;201.

115. Delbecq AL, Van de Ven AH, Gustafson DH. Group techniques for program planning: A guide to nominal group and Delphi processes: Scott Foresman; 1975.

116. Hsu C-C, Sandford BA. The Delphi technique: making sense of consensus. Practical assessment, research & evaluation 2007 Aug;12 (10):8.

117. Skulmoski GJ, Hartman FT, Krahn J. The Delphi method for graduate research. Journal of Information Technology Education: Research 2007 Jan 1;6:21.

118. Lissitz RW, Green SB. Effect of the number of scale points on reliability: A Monte Carlo approach. Journal of Applied Psychology. 1975;60(1):10.

119. Preston CC, Colman AM. Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. Acta Psychol (Amst). 2000;104(1):1-15.

120. Holey EA, Feeley JL, Dixon J, Whittaker VJ. An exploration of the use of simple statistics to measure consensus and stability in Delphi studies. BMC medical research methodology. 2007;7(1):52.

121. Becker GE, Roberts T. Do we agree? Using a Delphi technique to develop consensus on skills of hand expression. J Hum Lact. 2009;25(2):220-5.

122. Schneider P, Evaniew N, Rendon JS, McKay P, Randall RL, Turcotte R. Moving forward through consensus: protocol for a modified Delphi approach to determine the top research priorities in the field of orthopaedic oncology. BMJ open. 2016 May 1 6(5).

123. Garland R. The mid-point on a rating scale: Is it desirable. Marketing bulletin. 1991 May;2(1):70.

124. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. Journal of clinical epidemiology. 2007;60(1):34-42.

125. Souza ACd, Alexandre NMC, Guirardello EdB. Psychometric properties in instruments evaluation of reliability and validity. Epidemiologia e Serviços de Saúde. 2017;26(3):649-59.

126. Roach KE. Measurement of health outcomes: reliability, validity and responsiveness. JPO: Journal of Prosthetics and Orthotics. 2006;18(6):P8-P12.

127. Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. Journal of clinical epidemiology. 2009;62(10):1013-20.

128. Osborne JW, Costello AB. Sample size and subject to item ratio in principal components analysis. Practical assessment, research & evaluation. 2004;9(11):8.

129. Tsang S, Royse CF, Terkawi AS. Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine. Saudi journal of anaesthesia. 2017;11(Suppl 1):S80.

130. Portney LG, Watkins MP. Foundations of clinical research: applications to practice: Pearson/Prentice Hall Upper Saddle River, NJ; 2009.

131. Nunnally JC, Bernstein IH. The theory of measurement error. Psychometric theory. 1994;3:209-47.

132. McHugh ML. Interrater reliability: the kappa statistic. Biochemia medica. 2012;22(3):276-82.

133. Cohen J. A coefficient of agreement for nominal scales. Educational and psychological measurement. 1960;20(1):37-46.

134. Banerjee M, Capozzoli M, McSweeney L, Sinha D. Beyond kappa: A review of interrater agreement measures. Canadian journal of statistics. 1999;27(1):3-23.

135. Landis JR, Koch GG. The measurement of observer agreement for categorical data. biometrics. 1977:159-74.

136. Nurjannah I, Siwi SM. Guidelines for analysis on measuring interrater reliability of nursing outcome classification. International Journal of Research in Medical Sciences. 2017;5(4):1169-75.

137. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. Journal of chiropractic medicine. 2016;15(2):155-63.

138. Bland JM. Measuring health and disease: Cohen's Kappa. University of York Department of Health Sciences. 2008.

139. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Physical therapy. 2005;85(3):257-68.

140. Hallgren KA. Computing inter-rater reliability for observational data: an overview and tutorial. Tutorials in quantitative methods for psychology. 2012;8(1):23.

141. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. Journal of clinical epidemiology. 1990;43(6):543-9.

142. Polit DF, Beck CT. Fundamentos de pesquisa em enfermagem: avaliação de evidências para a prática da enfermagem: Artmed Editora; 2011.

143. Shoukri MM, Asyali MH, Donner A. Sample size requirements for the design of reliability study: review and new results. Statistical Methods in Medical Research. 2004;13(4):251-71.

APPENDIX A: Cover letter

Background:

Critical appraisal is an important process to determine and to assess the transparency, the scientific value, and the relevance of published articles. Critical appraisal tools are usually used to guide the critical appraisal process. Several techniques and formats are used to develop critical appraisal tools like question and answer format, checklist format, numeric scales and summary scores [1].

Scientists classified critical appraisal tools broadly into two categories: generic and research design-specific. Tools which are classified as design-specific consists of themes which are related to unique methodological issues for the research design [2, 3]. Moreover, researchers developed generic critical appraisal tools aiming to improve the research consumer's ability to appraise quantitative and qualitative studies to be able to come up with reliable evidence from them [4]. Some studies were conducted to modify different kind of tools like AMSTAR study that was developed to assess the quality of the systematic review methods based on the previously developed tools, other empirical evidence and expert consensus [5]. In addition to the development of other tools like PRISMA for systematic reviews and meta-analysis that was developed after noticing that the poor reporting of the systematic reviews key information [6].

Despite the acknowledgment of the significant value of critical appraisal tools [5], there is no available 'gold standard' critical appraisal tool specifically for clinical pharmacokinetic (PK) studies. Researchers noticed that 20% of the published trials did not specify basic pharmacokinetic parameters such as Volume of distribution and total clearance which are fundamental requirements for all drug dosing [7]. Consequently, pharmacokinetic reporting guidelines were published to guide researchers to conduct pharmacokinetic studies and to ensure the reporting of the minimum required basic

information. They developed “Yes/No” checklist that was composed of 24 items which were relevant to the majority of clinical pharmacokinetic studies [8]. However, this checklist did not allow readers to assess the methodological quality of the reported items since there were no numeric scales that provide a profile of strengths and weaknesses of each item of interest. Furthermore, this checklist did not assess the other dimensions of quality which were related to the study design, conduct, and analysis, clinical relevance, and result validity [9].

In the first Phase:

We did a systematic review in the first phase of this master project to identify the quality markers which were related to clinical pharmacokinetic studies to fulfill this gap in knowledge.

In the second phase:

Research question:

What are the important quality markers to be considered while developing clinical PK critical appraisal tool?

Goals:

To develop an inventory of items to be included in critical appraisal tool that assesses the quality of clinical pharmacokinetic studies based on experts’ consensus.

Specific Objectives:

To achieve expert consensus regarding the identified quality markers for appraising clinical pharmacokinetic studies through using modified Delphi method.

Ethical consideration:

The study protocol and other relevant documents will be submitted to Qatar University-Institutional Review Boards (QU-IRB) to obtain an ethical approval. All participants will be asked to sign a consent before the initiation of the study to show their agreement

in participating in the development of consensus. Moreover, quasi-anonymity and confidentiality of panelists' answers will be granted. All the data that will be collected through conducting Delphi method will be entered and stored in a laptop that is secured with a password under the custody of the study investigators (MSc student, MSc supervisor and Co-supervisor).

References:

1. Rosella L, Bowman C, Pach B, Morgan S, Fitzpatrick T, Goel V. The development and validation of a meta-tool for quality appraisal of public health evidence: Meta Quality Appraisal Tool (MetaQAT). Public Health. 2016 Jul 1;136:57-65.
2. Crombie I. The Pocket Guide to Critical Appraisal: A Handbook for Health Care Professionals London: BMJ Publishing Group; 1996.
3. Elwood J. Critical Appraisal of Epidemiological Studies and Clinical Trials. 2nd ed. Oxford: Oxford University Press; 1998.
4. Sackett D, Richardson W, Rosenberg W, Haynes R. Evidence Based Medicine. How to Practice and Teach EBM. London: Churchill Livingstone; 2000.
5. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol. 2007;7:10.
6. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epidemiol. 2009;62(10):e1-34.

7. Touw DJ, Neef C, Thomson AH, Vinks AA. Cost-effectiveness of therapeutic drug monitoring: a systematic review. Ther Drug Monit. 2005;27(1):10–7.
8. Kanji S, Hayes M, Ling A, Shamseer L, Chant C, Edwards DJ,et al. Reporting guidelines for clinical pharmacokinetic studies: the ClinPK statement. Clinical pharmacokinetics. 2015 Jul 1;54(7):783-95.
9. Juni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. Bmj. 2001;323(7303):42-6.

APPENDIX B: Rules of modified Delphi

Problem statement:

There is a gap in the knowledge, as there is no available critical appraisal tool that aids clinicians in appraising and determining the quality of the published clinical pharmacokinetic studies. Since pharmacokinetic principles are applied in clinical practice settings to help in reducing mortality, length of treatment, length of hospital stays (LOS), morbidity, adverse effects of drug therapy, and cost-savings, it is important to develop such a tool to enhance the guidance of evidence-based medicine application in practice.

Research question:

What aspects of quality can be used to appraise a clinical pharmacokinetic trial from the experts' point of view?

Goals:

To develop an inventory of items to be included in a critical appraisal tool that assesses the quality of clinical pharmacokinetic studies based on experts' consensus

Specific Objectives:

To achieve expert consensus regarding the quality markers identified from the systematic review that was conducted in phase I of this project through using modified Delphi method.

Ethical consideration:

The study protocol and other relevant documents was submitted to Qatar University-Institutional Review Boards (QU-IRB) to obtain an ethical approval. All participants will be asked to sign a consent before the initiation of the study to show their agreement in participating in all the rounds of the modified Delphi process. Moreover, quasi-anonymity and confidentiality of panelists' answers will be granted. All the data

that will be collected through conducting Delphi method will be entered and stored in a laptop that is secured with a password under the custody of the study investigators (MSc student, MSc supervisor and Co-supervisor).

Criteria for selecting participants:

Inclusion criteria of the expert panelists:

- Participants who have an academic position that reflects their direct involvement in the research field of clinical pharmacokinetics
- Clinicians who have experience in application of clinical pharmacokinetic principles in their clinical practice. These individuals should have experience in interpreting the findings of clinical pharmacokinetic studies and applying these to their patients
- Pharmaceutical industry researchers with experience in clinical pharmacokinetics
- Individuals in health regulation who assess clinical pharmacokinetic studies when making decisions for their respective health authorities
- Participant should be willing to participate in all rounds

This diversity will help in ensuring that the selected panelists are going to be representative of most of the stakeholders and will reduce the incidence of selection bias.

Sampling method:

- Purposeful sampling by using different strategies like criterion-i convenience strategy, and heterogeneous sampling [3].

Number of participants: 25 participants will be included in the study.

Number of rounds and timing:

There will be three – four rounds. Participants will be given 2 weeks to answer each

questionnaire with two subsequent reminders to be sent via email [4]. The updated questionnaire will be sent to the panelists after one week, during which investigators are going to modify and to remove items from the questionnaire based on the panelists' comments before the second round. The same will be repeated for the third round.

Questionnaire design and administration:

SurveyMonkey will be used to develop a questionnaire for each round in the Delphi process and to collect responses. The questionnaire will be divided into 7 main sections (title/abstract, introduction, methods, results, discussion, conclusion, others). Several items will be included under each section. Participants will be asked to rate their agreement on including each item in a critical appraisal tool for clinical pharmacokinetics on 5 Likert-scale ranging from (5) strongly agree, (4) agree, (3) neither agree or disagree, (2) disagree, (1) strongly disagree. Participants will be given the opportunity to add any modification to the listed items or to suggest the addition of a new item [5].

Criteria for consensus:

In Delphi methodology, there was no agreement between scientists on what constitutes consensus [3]. In healthcare, the most commonly used consensus measurements were level of agreement, median scores, and interquartile ranges [3,6].

Table 1: Consensus Thresholds

Inclusion	• More than 75% of participants provide positive result on 5-Point Likert-scale (through selecting 4 or 5 on 5 points Likert-scale). • A median score of more than or equal to 4 on 5-Point Likert-scale. • Interquartile-range of less than or equal to 1.
Exclusion	• Less than 75% of participants provide negative result on 5-Point Likert-scale (through selecting 1 or 2 on 5 points Likert-scale). • A median score of less than or equal 2 on 5-Point Likert scaleLikert-scale. • Interquartile-range of more than 1.
Non-Consensus	• If the item does not meet any of the inclusion or exclusion consensus threshold, it will be modified based on the panelists' comments and circulated for the second round. Then, the change in the distribution of responses of each item will be measured between the rounds. If the change in the distribution of responses is more than 15% between rounds, the item will be modified again based on the panelists' comments and recirculated to the next round. On the other hand, if the change in the distribution of responses is 15% or less between rounds, the item will be excluded because this shows that there is stability [7,8,9].

References:

1. Jorm AF. Using the Delphi expert consensus method in mental health research. Australian & New Zealand Journal of Psychiatry. 2015 Oct;49(10):887-97.
2. Rowland M, Tozer T. Clinical pharmacokinetics and pharmacodynamics: concepts and applications. 4th ed. USA: Lippincott Williams and Wilkins; 2010
3. Keeney, Sinead, Hasson, Felicity, & McKenna, Hugh (2006). Consulting the oracle: Ten lessons from using the Delphi technique in nursing research. Journal ofAdvanced Nursing, 53(2), 205–212.
4. Suri H. Purposeful sampling in qualitative research synthesis. Qualitative Research Journal. 2011 Aug 3;11(2):63-75.

5. Garland R. The mid-point on a rating scale: Is it desirable. Marketing bulletin. 1991 May;2(1):66-70.
6. von der Gracht, Heiko A. (2012). Consensus measurement in Delphi studies: Review and implications for future quality assurance. Technological Forecasting and Social Change, 79, 1525–1536.
7. Scheibe M, Skutsch M, Schofer J. IV. C. Experiments in Delphi methodology. In: Linstone H, Turoff M (eds) The Delphi method: techniques and applications. 2002:257-81.
8. Chneider P, Evaniew N, Rendon JS, McKay P, Randall RL, Turcotte R. Moving forward through consensus: protocol for a modified Delphi approach to determine the top research priorities in the field of orthopaedic oncology. BMJ open. 2016 May;6(5).
9. Garland R. The mid-point on a rating scale: Is it desirable. Marketing bulletin. 1991 May;2(1):70.

APPENDIX C: Ethical approval for Phase II

جامعة قطر
QATAR UNIVERSITY

Qatar University Institutional Review Board
QU-IRB

APP-31/05/2018/QU

October 3, 2018

Dr. Shane Pawluk
College of Pharmacy
Qatar University
Tel.: 4403-5619
Email: shane.pawluk@qu.edu.qa

Dear Dr. Shane Pawluk,

Sub.: Research Ethics Review Exemption
Ref.: Project titled, "Strengthening the quality of reporting of clinical pharmacokinetic studies: development and validation of a quality reporting tool for clinical pharmacokinetic research"

We would like to inform you that your application along with the supporting documents provided for the above proposal, is reviewed and having met all the requirements, has been exempted from the full ethics review.

Please note that any changes/modification or additions to the original submitted protocol should be reported to the committee to seek approval prior to continuation.

Your Research Ethics Approval No. is: **QU-IRB 970-E/18**
Kindly refer to both your application tracking number and the IRB Approval number in all your future correspondence pertaining to this project.

Best wishes,

Dr. Mashael Al-Shafai
Chairperson, QU-IRB

Institutional Review Board
(IRB)
Office Of Academic Research

Qatar University-Institutional Review Board (QU-IRB), P.O. Box 2713 Doha, Qatar
Tel +974 4403-5307 (GMT +3hrs) email: QU-IRB@qu.edu.qa

APPENDIX D: 64 formulated questions out of the systematic review

Title					
Was the title concisely reflected as the discussed topic in the paper? The title should reflect the name of the analyzed medication, and comparator (if applicable), the targeted patient population, and the study design	1	2	3	4	5
Abstract					
Was a summary of the article provided to the readers within the abstract? A brief description of the knowledge gap, the objectives, summary of the methods, the results of the primary objectives and the main conclusion should be provided.	1	2	3	4	5
Background					
Was a comprehensive introduction provided about the analyzed drug that showed the rationale behind the conduction of that study? Authors may provide information related to: • Stages of the analyzed drug development. • Known aspects of the drug's absorption, distribution, metabolism and elimination. • Previous studies. • What will be added to the existing body of knowledge with their proposed study.	1	2	3	4	5
Was a clear description of the objectives of the study clearly provided? Authors should provide a clear statement of the objectives of the research to clarify the purpose and, the scope of the study, so readers will know if the study matches their interest or not	1	2	3	4	5
Methods					
Study Design					
Was the chosen study design appropriately selected and justified? Example: • Immediate release formulation, single dose study design is recommended. • Sustained release formulation and medication with a long half-life or high intra-patient variability, parallel study design or steady-state design are recommended. • Bioequivalence studies, crossover design is recommended with a washout period between the administered interventions.	1	2	3	4	5
Was the method used for drug dosing appropriate and/or referenced?	1	2	3	4	5
Were the endpoints of the study clearly stated? The endpoints should be directly related to the objectives.	1	2	3	4	5
Was the eligibility criteria of participant inclusion provided? The inclusion criteria should allow authors to choose representative participants appropriately who are representative of the targeted population to answer the main study question	1	2	3	4	5
Were the exclusion criteria of participants provided?	1	2	3	4	5

Exclusion criteria should not be restricted to an extent that interferes with the generalizability of the study results					
Were the study setting/location clearly described? Authors should reflect on the setting and/or location, as this will help practitioners utilize the results of the research.	1	2	3	4	5
During the last week, how many days did you forget to take your pills?	1	2	3	4	5
If applicable, was the used method to generate the random allocation sequence described? Example: Simple randomization Block randomization Stratified randomization Unequal randomization	1	2	3	4	5
If applicable, was the used method of allocation concealment described?	1	2	3	4	5
Were any of the participants, the investigators or the individuals who analyzed the data blinded while the study was being conducted?	1	2	3	4	5
Was the method of data sampling provided? Example: • Extensive data sampling is essential to select the most appropriate structural model number of compartments, first Vs second order absorption, and lag time. • To determine the linearity of pharmacokinetics sparse data sampling is recommended. Researchers obtain these data from previously conducted studies with completed concentration-time profile e.g. phase I studies.	1	2	3	4	5
Was a clear description of both the intervention and if applicable the comparator provided? Medication name, dose, dose units, schedule or frequency, route of administration, starting and stopping date of administering the medication, the reason for use	1	2	3	4	5
Was a clear description of the *sampling site* and the *sampling interval* (the exact times at which samples are obtained) provided and justified? Example: • Arterial sampling is preferable during frequent sampling schedule. • Arterial sampling is more representative of the delivered concentration to the effect site in case of peripheral elimination. • Arterial sampling is preferable when administering a drug that has a short duration of action or fast onset of action. • Sampling interval should not exceed the expected half-life of the studied exponential phase (fast distribution, slow distribution and elimination). 3-4 samples are the minimum number of samples to be collected. • Bioequivalence studies: researchers should continue to collect samples until 80% of the AUC is covered.	1	2	3	4	5
Was a description of participant follow-up clearly described? Example: Monitoring parameters (e.g. signs and symptoms of disease or side effects of the given medication, lab data, etc.) to be collected in each period should documented.	1	2	3	4	5
Was the storage condition of samples clearly described? Example:					

• Anticoagulant • Stabilizers • Centrifugation • Temperature • Labeling	1	2	3	4	5
Was there a clear description of the pharmacokinetic model, its development, and justification for use? It is recommended to provide the following details about the selected modeling process: • Description of studies from which dataset was driven • Model structure Validated software for the PK analysis • Criteria for accepting valid model's parameters • Fitting procedure defined prior to the initiation of the analysis. A reasonable assumption based on which the scheme for weighting is considered to be appropriate and the transformation of data (e.g. logarithmic transformation to achieve the homoscedastic (constant) variance requirements) should be provided.	1	2	3	4	5
Was a list of interacting covariates (demographic variables, laboratory values, co-medication, environmental factors and disease states) provided a priori? • The covariates should be predetermined based on biological plausibility.	1	2	3	4	5
Was a detailed description of the used apparatus provided? Example: • Chromatography type • Detection type • Assay characteristics: mobile phase composition, gradient and flow rate, chromatographic column (packing material, dimensions), • Analytical runtime • Operating temperature and detection parameters. • Validation method: specificity, recovery, linearity and sensitivity, the stability of the assay and its reproducibility.	1	2	3	4	5
Were the used pharmacokinetic equations to calculate different parameters (e.g. creatinine clearance) disclosed within the article?	1	2	3	4	5
Was the method used to estimate the area under the curve (AUC) and the area under the first moment curve (AUMC) provided? Example: • AUC and AUCM can be estimated by using the linear trapezoidal rule in case of increasing or equal concentrations. • AUC and AUCM can be estimated by using the log-linear trapezoidal rule for decreasing concentrations.	1	2	3	4	5
Were the used weight metrics in the pharmacokinetic calculation and drug dosing provided?	1	2	3	4	5
Was the used population pharmacokinetic approach and validation method described? Example: *Population PK approach* • Standard two-stage • Naive pooling of data • Mixed-effects modeling *Basic internal methods* • Goodness-of-fit plots/diagnostic plots. • Uncertainty in parameter estimates	1	2	3	4	5

• Model sensitivity to outliers *Advanced internal methods* • Data splitting • Bootstrap • Cross validation • Simulations such as visual or posterior predictive checks (PPCs) • *External model evaluation* (validation dataset observations compared with model predictions).					
Did the authors justify the selection of the key models at different stages of the development process? Justification of key models' selection at different stages of the development process through using Goodness-of-fit (GOF) plots: **C)** Predicted data versus observed data (PRED versus DV; a line of identity and a trendline should be included) **D)** PRED versus weighted residuals (WRES; zero line and a trend line should be included), **E)** Time versus WRES (a zero line and a trend line should be included). Time can be both time after dose and continuous time (time in the study).	1	2	3	4	5
Was the study approved by a regional Research Ethics board?	1	2	3	4	5
Applied statistics					
Was a detailed description or reference of the specific level of statistical significance and the sample size calculations provided before the initiation of the study to ensure adequate power for detecting differences of interest?	1	2	3	4	5
Were the chosen statistical tests and software to perform the statistical analysis appropriate to achieve the study objectives?	1	2	3	4	5
Results					
Was a patient flow diagram provided? Example: • Number of patients who enrolled in each arm of the trial • Description of withdrawals	1	2	3	4	5
Were the baseline characteristics of the included participants provided? All the following variables should be clearly defined for all participants': sex, race, age, weight, height, concomitant diseases, co-medication, smoking habits, covariates, the severity of illness, residual, renal function, and hepatic function.	1	2	3	4	5
How did you rate your performance during last year in your work?	1	2	3	4	5
Was the method used to handle missing data during the results analysis described and justified?	1	2	3	4	5
Was the method used to handle outliers during analysis provided?					

• Authors should explain the physiological/study events which result in excluding the data from the analysis. • Outlying data should be included in the final population PK model and their effect on different PK parameters should be documented.	1	2	3	4	5
Was an appropriate measure of precision (e.g. descriptive statistics confidence interval, standard deviation, mean, median, range, interquartile range, and trimmed range) used to document the pharmacokinetic results?	1	2	3	4	5
Were the essential pharmacokinetic parameters required to perform dose calculations in practice setting documented? Consider: Total clearance (CL), Fraction of dose excreted unchanged in urine (fe), Volume of distribution at steady state (Vss), Volume of distribution during the terminal phase (VZ), Blood/plasma concentration ratio, Terminal half-life (t1/2 Z), Fraction of unbound drug in plasma (fu), Bioavailable fraction of dose (F), Absorption rate constant (ka), Cmin, Cmax, tmax, EC50, ke0, Hill coefficient, or gamma	1	2	3	4	5
Discussion					
Were the study limitations described by the authors consistent with those identified within the study?	1	2	3	4	5
Was the provided interpretation consistent with the displayed results?	1	2	3	4	5
Did the authors compare their observed results with the results of other relevant studies and if they could be generalized to the targeted population?	1	2	3	4	5
Were recommendations of future studies provided?	1	2	3	4	5
Conclusion					
Was the provided conclusion supported by the observed results? Authors should provide a summary of the observed results. Authors should not provide any new information in conclusion.	1	2	3	4	5
Others					
	1	2	3	4	5
Did the authors disclose any funding resources?	1	2	3	4	5
Did the authors disclose any conflict of interest?	1	2	3	4	5

APPENDIX E: Appendix-1 in the final tool

This appendix has been added based on comments. It is not intended as a question to be answered by users of this tool, but more as a reference should one be required. Please rate your agreement with having this appendix available to users of this tool.

All the following variables should be clearly defined for all participants: sex, race, age, weight, height, concomitant diseases, co-medication, smoking habits, covariates, the severity of illness, residual, renal function, and hepatic function. Authors should describe if participants are taking any medications that may interact with the analyzed medication.

The Acute Dialysis Quality Initiative (ADQI) minimum reporting criteria by ADQI should be followed in case of including participants on dialysis.

a) Operational characteristics

- Membrane/ dialyzer/filter and area
- A measure of time spent on therapy
- Delivery device
- Access and blood flow
- Anticoagulation
- Replacement fluid composition and administration
- Dialysis fluid composition and administration

b) Patient characteristics

- A measure of time spent on therapy

- Surgical/trauma/medical/other
- A measure of the severity of illness
- Cointerventions
- Integrated hemodynamic status and vasopressor treatment
- Outcomes

In the case of including participants suffering from renal failure the following information should be provided:

- Cause
- Plasma creatinine concentration/creatinine clearance
- Plasma electrolytes
- Hemoglobin concentration
- Plasma protein level
- Time and the nature of last dialysis
- Existence of clinical edema
- Existence of peripheral neuropathy

In case of including participants who have hepatic cirrhosis the following information should be provided:

- Cause
- Child's Pugh Score
- Prothrombin time, platelet count
- Albumin and globulin levels

In case of including critically ill participants following information should be provided:

- Clinical description
- Apache II score
- Plasma creatinine and electrolyte concentrations
- Presence/absence of renal failure
- Presence/absence of liver failure

In case of including participants suffering from thermal injury the following information should be provided:

- Regular hematocrit
- Preoperative plasma albumin and globulin level

In the case of Bioequivalence studies, the following criteria should be fulfilled:

- Nonsmoker healthy volunteers (males/females) with a body weight that is ± 20% of the standard and with age between 18 to 55 years old should be enrolled.

APPENDIX F: Appendix-2 in the final tool

Model evaluation / validation and diagnostics	Methods	Description of use
Basic internal methods		
Goodness-of-fit plots/diagnostic: assessments of the goodness of fit statistics/ plots	The following predicted data should be plotted versus the observed data: PRED (population prediction) vs. DV (dependent variable), PRED vs. WRES (weighted residuals) or CWRES (conditional weighted residuals), and time vs. WRES or CWRES	Different models were selected based on the observed goodness-of-fit in the diagnostic plots.
Internal validation	Data splitting: randomly the data are separated into an index population and a test population	Experts use data splitting to determine if the used model was robust.
Model reliability: to assess parameters uncertainty and its random effect; the plausibility of parameter estimates and their precision Model stability: to identify the extent to which the model is resistant to change	Log-likelihood profiling: mapping the objective function; considered as another method for determining parameter CI	Use the likelihood profiling to enhance the model fit through models evaluation through considering the change in both the objective function and determination of the empirical 95% CI.
Resampling techniques	Bootstrapping technique is used to estimate 95% CIs and standard error of the estimate. Bootstrap generates other plausible data and assesses model structure.	Used for the validation of the final model.
	Jack-Knife techniques: estimates the standard error of estimates.	Cross-validation is done through using Jack-Knife techniques to assess the accuracy and the validity of the pharmacokinetic final model.
	Cross-validation	Cross-validation is used to determine the robustness and ability of the final model to predict data. Randomly, a full dataset was divided
Case-deletion diagnostics	Determine and assess the important outliers' effect	Case-deletion diagnostics are used to evaluate the final model by detecting influential individual and assessing its robustness.
Advanced internal method		
Simulation-based diagnostics	VPC (visual predictive check): this is a plot used to compare the 95% prediction interval to the observed data	VPC: used to determine the accuracy and the performance of the used model for data description.

	NPC (Numerical predictive check): this is used to assess the appropriateness of the model	NPC: used to assess the final model simulation properties.
	NPDE (Normalized prediction distribution error)	NPDE: used to produce 1000 model-prediction concentration for each observation available in the external dataset. After that, the observed concentration compared to the 1000 predicted concentrations.
	PPC (posterior predictive check): this is used to assess and determine the predictive model performance	PPC: used to determine if the used model helps in extensively describing both the covariate disposition and the pharmacokinetic parameters.
External model validation		
External validation	A validation dataset from another study used to test the developed model	The external validation process assesses the developed final model.

APPENDIX G: The clinical pharmacokinetics critical appraisal tool before the content and face validity

Evaluator's name: _____

Name of the evaluated article:
_____

Estimated time of evaluating the paper: _____
Score: _____/21_____

Clinical Pharmacokinetic Critical Appraisal Tool (CPKCAP)	
Appraising Background	
1-Was a clear description of the objectives of the study provided? Authors should provide a clear statement of the objectives of the research to clarify the purpose and the scope of the study.	□ Yes □ No □ I Do Not Know □ Not Applicable Comments: _____
2- Was a valid and comprehensive rationale provided to support the purpose of the study?	□ Yes □ No □ I Do Not Know □ Not Applicable Comments: _____
Appraising Design	
3-Was the chosen study design appropriately selected and justified?	□ Yes □ No □ I Do Not Know □ Not Applicable Comments: _____

4-Was the dosing (dose, route of administration, dosing interval) of the drug in the study justified for the intended study? Example: Authors should justify the use of single-dose versus steady-state analysis	□ Yes □ No □ I Do Not Know □ Not Applicable Comments: _____
5-Were the endpoints of the study appropriate to answer the objectives of the study?	□ Yes □ No □ I Do Not Know □ Not Applicable Comments: _____
6-Were the endpoints of the study appropriate to answer the objectives of the study?	
7-Where applicable, were the relevant baseline characteristics of the participants adequately described? Examples: sex, race, age, weight, height, concomitant disease, administered medications, smoking status, pregnancy, severity of illness that may affect pharmacokinetic parameters, renal function, and hepatic function. Note: Note: Please refer to Appendix-2 Patient Demographics for further clarification.	□ Yes □ No □ I Do Not Know □ Not Applicable Comments: _____

8-Were plausible interacting covariates described a priori or in post hoc evaluation? Examples: demographic variables, laboratory values, concomitant medications, and relevant disease states to the drug being studied.	□ Yes □ NO □ I Do Not Know □ Not Applicable Comments: ————————————
9-Was the description of the used sample analysis methods or citations of prior validation studies provided in the publication or affiliated appendix? Example: Chromatography type. Detection type. Assay characteristics: mobile phase composition, gradient and flow rate, chromatographic column (packing material, dimensions). Analytical runtime. Operating temperature and detection parameters. Validation method: specificity, recovery, linearity and sensitivity, the stability of the assay and its reproducibility.	□ Yes □ No □ I Do Not Know □ Not Applicable Comments: ————————————
Appraising Sampling	
10- Was the method of data sampling appropriate for the study? Examples: first vs second order absorption, and lag time. Evaluating for nonlinearity requires multiple dose levels and a complete profile is recommended. Researchers obtain these data from previously conducted studies with completed concentration-time profile e.g. phase I studies. The method of data sampling should reference previously	□ Yes □ No □ I Do Not Know □ Not Applicable Comments: ————————————

validated quantitative bioanalytical methods and if those are not available then the full description or defense of data sampling should be included.	
11- Was a clear description of the sampling site and the sampling interval (the exact times at which samples are obtained) provided and justified? Example: Sampling site should be consistent for all subjects in the study. Arterial sampling is preferable during frequent sampling schedule. Arterial sampling is more representative of the delivered concentration to the effect site in the case of peripheral elimination. Arterial sampling is preferable when administering a drug that has a short duration of action or fast onset of action. Sampling interval should not exceed the expected half-life of the studied exponential phase (fast distribution, slow distribution and elimination).	□ Yes □ No □ I Do Not Know □ Not Applicable Comments: ————————————
12- Was the number of half-lives elapsed within the sampling period appropriate for the analyzed drug?	□ Yes □ No □ I Do Not Know □ Not Applicable Comments: ————————————
13- Were sample storage conditions appropriate and described in a manner that could be accurately replicated? Examples: sample storage temperature, use and description	□ Yes □ No □ I Do Not Know □ Not Applicable

of anticoagulants, stabilizers, centrifugation	Comments: ————————————
14- If applicable, was there a clear description of the pharmacokinetic model, its development, validation and justification for use? It is recommended to provide the following details about the selected modeling process: Description of studies from which dataset was driven Model structure Validated software for the pharmacokinetic analysis Criteria for accepting valid model's parameters Fitting procedure defined prior to the initiation of the analysis. A reasonable assumption based on which the scheme for weighting is considered to be appropriate and the transformation of data (e.g. logarithmic transformation to achieve the homoscedastic (constant) variance requirements) should be provided.	□ Yes □ No □ I Do Not Know □ Not Applicable Comments: ————————————
15- Was the described population pharmacokinetic approach validation method appropriate for the analysis? 1- Basic internal method 2- Advanced internal method 3- External model evaluation Note: Please refer to Appendix-1 Model Evaluation for further clarification.	□ Yes □ NO □ I Do Not Know □ Not Applicable Comments: ————————————
16- Were the essential pharmacokinetic parameters required to make the results applicable in clinical settings addressed? Consider: Total clearance (CL), Fraction of dose excreted unchanged in urine (fe), Volume	□ Yes □ NO □ I Do Not Know □ Not Applicable Comments: ————————————

of distribution at steady state (Vss), Volume of distribution during the terminal phase (VZ), Blood/plasma concentration ratio, Terminal half-life (t1/2 Z), Fraction of the unbound drug in plasma (fu), Bioavailable fraction of dose (F), Absorption rate constant (Ka),Cmin, Cmax, tmax, EC50, Ke0, Hill coefficient, or gamma, AUC and bioavailability for the two drug formulations.	
17- Were the pharmacokinetic equations used to calculate patient pharmacokinetic parameters disclosed or cited within the article? Example: creatinine clearance, body weight calculations, Michaelis Menten, volume of distribution, patient weight: total body weight vs. ideal body weight.	□ Yes □ No □ I Do Not Know □ Not Applicable Comments: _____
Appraising Applied statistics	
18- Were the chosen statistical tests and software to perform the statistical analysis appropriate to achieve the study objectives?	□ Yes □ No □ I Do Not Know □ Not Applicable Comments: _____
Appraising Results	
19- Were all patients enrolled in the study accounted for? Example: Description of patient screening, enrollment, run-in or wash out phases, study period and follow-up periods are adequately described. Any loss to follow-up or withdrawals are described.	□ Yes □ No □ I Do Not Know □ Not Applicable Comments: _____

20- In the event of missing data or outliers, was the process for analysis justified and appropriate?	□ Yes □ No □ I Do Not Know □ Not Applicable Comments: _____
21- Were appropriate summary statistics to describe centrality and variance used to document the pharmacokinetic results? Example: Descriptive statistics confidence interval, standard deviation, mean, median, range, interquartile range, standard error and trimmed rang	□ Yes □ No □ I Do Not Know □ Not Applicable Comments: _____

APPENDIX H: The modified clinical pharmacokinetics critical appraisal tool after face and content validity

Evaluator's name: _____

Name of the evaluated article: _____

Time to complete evaluation of article: _____

Number of YES responses: _____

Clinical Pharmacokinetics Critical Appraisal Tool (CPKCAP)	
Appraising Background	
1. Was a clear description of the objectives of the study provided? • Authors should provide a clear statement of the objectives of the research to clarify the purpose and the scope of the study.	□ Yes □ No □ I Do Not Know □ Not Applicable Comments: _____
2. Was a clear and comprehensive rationale provided to support the purpose of the study?	□ Yes □ No □ I Do Not Know □ Not Applicable Comments: _____
Appraising Study Design and Experimental Methods	
3. Was the chosen study design appropriately selected and justified?	□ Yes □ No □ I Do Not Know □ Not Applicable Comments: _____
4. Was the dosing (i.e. dose, route of administration, and dosing interval) of the drug in the study justified for the intended study? **Examples:** • Authors should justify the use of single-dose versus steady-state analysis.	□ Yes □ No □ I Do Not Know □ Not Applicable Comments: _____

5. Were the outcome measures endpoints of the study appropriate to address the objectives of the study?	□ Yes □ No □ I Do Not Know □ Not Applicable Comments: _____
6. Were the exclusion criteria of participants included AND appropriate for the intended outcomes of the study? • The exclusion criteria should be relevant to assist with decreasing significant confounders (e.g. co-administration of drugs, organ impairment, and special populations) that may impact outcomes.	□ Yes □ No □ I Do Not Know □ Not Applicable Comments: _____
7. Where applicable, were the relevant baseline characteristics of the participants adequately described? **Examples:** • Sex, race, age, weight, height, concomitant disease, administered medications, smoking status, pregnancy, severity of illness that may affect pharmacokinetic parameters, renal function, and hepatic function. **Note:** Please refer to Appendix-1 Patient Demographics for further clarification.	□ Yes □ No □ I Do Not Know □ Not Applicable Comments: _____
8. Were plausible interacting covariates described *a priori* or in post hoc evaluation? **Examples:** • Demographic variables, laboratory values, concomitant medications, and relevant disease states to the drug being studied.	□ Yes □ No □ I Do Not Know □ Not Applicable Comments: _____
9. Was the description of the used biological sample analytical methods sample analysis methods or citations of prior validation studies provided in the publication or affiliated appendix? **Examples:** • Chromatography type. • Detection type. • Assay characteristics: mobile phase composition, gradient and flow rate, chromatographic column (packing material, dimensions).	□ Yes □ No □ I Do Not Know □ Not Applicable Comments: _____

• Analytical runtime. • Operating temperature. • Detection parameters. • Validation method: specificity, recovery, linearity and sensitivity, the stability of the assay and its reproducibility.	
10. Was the method of data sampling of analytics appropriate for the study? **Examples:** • First vs. second order absorption, and lag time. • Evaluating for nonlinearity requires multiple dose levels and a complete profile is recommended. • Researchers obtain these data from previously conducted studies with completed concentration-time profile (e.g. phase I studies). • The method of data sampling should reference previously validated quantitative bioanalytical methods and if those are not available then the full description or defense of data sampling should be included.	□ Yes □ No □ I Do Not Know □ Not Applicable Comments: _____
11. Was a clear description of the sampling site provided and justified? **Examples:** • Sampling site should be consistent for all subjects in the study. • Arterial sampling is preferable during frequent sampling schedule. • Arterial sampling is more representative of the delivered concentration to the effect site in the case of peripheral elimination. • Arterial sampling is preferable when administering a drug that has a short duration of action or fast onset of action.	□ Yes □ No □ I Do Not Know □ Not Applicable Comments: _____
12. Was the number of half-lives elapsed within the sampling period appropriate for the analyzed drug? **Examples:** • Sampling interval should not exceed the expected half-life of the studied exponential phase (fast distribution, slow distribution and elimination).	□ Yes □ No □ I Do Not Know □ Not Applicable Comments: _____
13. Were sample storage conditions appropriate and described in a manner that could be accurately replicated? **Examples:** • Sample storage, temperature, use and	□ Yes □ No □ I Do Not Know □ Not Applicable Comments:

description of anticoagulants, stabilizers, centrifugation etc.	_____
14. If applicable, was there a clear description of the pharmacokinetic model, its development, validation and justification for use? It is recommended to provide the following details about the selected modeling process: • Description of studies from which dataset was driven • Model structure • Validated software for the pharmacokinetic analysis • Criteria for accepting valid model's parameters • Fitting procedure defined prior to the initiation of the analysis. • A reasonable assumption based on which the scheme for weighting is considered to be appropriate and the transformation of data [e.g. logarithmic transformation to achieve the homoscedastic (constant) variance requirements] should be provided.	□ Yes □ No □ I Do Not Know □ Not Applicable Comments: _____
15. Was the described population pharmacokinetic approach validation method appropriate for the analysis? 4- Basic internal method 5- Advanced internal method 6- External model evaluation **Note**: Please refer to Appendix-2 Model Evaluation for further clarification.	□ Yes □ No □ I Do Not Know □ Not Applicable Comments: _____
16. Were the essential pharmacokinetic parameters required to make the results applicable in clinical settings included? **Examples:** • Total clearance (CL), Volume of distribution at steady state (Vss), Blood/plasma concentration ratio, Terminal half-life ($t_{1/2}$ Z), Fraction of the unbound drug in plasma (fu), Absorption rate constant (Ka),C_{min}, C_{max}, t_{max}, , AUC, etc.	□ Yes □ No □ I Do Not Know □ Not Applicable Comments: _____
17. Were the pharmacokinetic equations used to calculate the patient's pharmacokinetic parameters presented or cited within the article? **Examples:**	□ Yes □ No □ I Do Not Know □ Not Applicable Comments:

• Equations used to calculate the following pharmacokinetic parameters: creatinine clearance, body weight calculations, Michaelis Menten, volume of distribution	_____
Appraising Applied Statistics	
18. Were the chosen statistical tests and software to perform the statistical analysis appropriate to achieve the study objectives?	□ Yes □ No □ I Do Not Know □ Not Applicable Comments: _____
Appraising Results	
19. Were all patients enrolled in the study accounted for? **Examples:** • Description of patient screening, enrollment, run-in or wash out phases, study period and follow-up periods are adequately described. Any loss to follow-up or withdrawals are described.	□ Yes □ No □ I Do Not Know □ Not Applicable Comments: _____
20. In the event of missing data or outliers, was the process for analysis justified and appropriate?	□ Yes □ No □ I Do Not Know □ Not Applicable Comments: _____
21. Were appropriate summary statistics to describe centrality and variance used to present the pharmacokinetic results? **Examples:** • Descriptive statistics such as confidence interval, standard deviation, mean, median, range, interquartile range, standard error and trimmed range	□ Yes □ No □ I Do Not Know □ Not Applicable Comments: _____