UPPSALA
UNIVERSITET

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Pharmacy 243

Pharmacometrics to improve clinical benefit assessment in oncology

EMILIE SCHINDLER

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2018

ISSN 1651-6192
ISBN 978-91-513-0191-4
urn:nbn:se:uu:diva-336420

Dissertation presented at Uppsala University to be publicly examined in B/B42, Biomedicinskt centrum, Husargatan 3, Uppsala, Friday, 16 February 2018 at 09:15 for the degree of Doctor of Philosophy (Faculty of Pharmacy). The examination will be conducted in English. Faculty examiner: Rik de Greef (Certara).

Abstract

Schindler, E. 2018. Pharmacometrics to improve clinical benefit assessment in oncology. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Pharmacy* 243. 73 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-513-0191-4.

The high attrition rate in oncology drug development calls for new approaches that would increase the understanding of drugs' efficacy and safety profiles. This thesis focuses on the development of pharmacometric models to characterize and quantify the relationships between drug exposure, circulating and imaging biomarkers, adverse effects, overall survival (OS), and patient-reported outcomes (PROs).

In axitinib-treated metastatic renal cell carcinoma patients, exposure-driven changes in soluble VEGF receptor 3 were linked to tumor size dynamics, which could in turn predict OS better than biomarker- or hypertension-related predictors. In sunitinib-treated gastro-intestinal stromal tumor (GIST) patients, the tumor metabolic response was sensitive to sunitinib dosing schedule and a substantial inter-lesion variability was quantified. A more pronounced decrease in tumor metabolism for the lesion that best responds to treatment after one week was predictive of longer OS. In imatinib-treated GIST patients, tumor volume better detected size changes of liver metastases and were slightly more predictive of OS than conventional tumor diameters, while tumor density had no predictive value.

A new modeling approach, the minimal continuous-time Markov model (mCTMM), was developed to facilitate the analysis of ordered categorical scores with Markovian features, e.g. fatigue or hand-foot syndrome grades. The mCTMM is applicable when existing approaches are not appropriate (non-uniform assessment intervals) or not easily implemented (variables with large number of categories).

An item response theory pharmacometric framework was established to describe longitudinal item-level data of a PRO questionnaire, the Functional Assessment of Cancer Therapy-Breast (FACT-B). Four correlated latent well-being variables characterized the multi-dimensional nature of FACT-B. When applied to data from breast cancer patients, the progression of physical well-being was typically better in patients treated with ado-trastuzumab emtansine (T-DM1) than with capecitabine-plus-lapatinib-treated patients. No relationship was identified between T-DM1 exposure and any of the latent variables.

In summary, the developed models advance the use of pharmacometrics in assessing the clinical benefit of anti-cancer therapies. They provide a quantitative understanding of the desired and adverse responses to drugs, and their relationships to exposure and long-term clinical outcome. Such frameworks may help to early assess response to therapy and optimize dosing strategies for investigational or existing therapies.

Keywords: nonlinear mixed effect models, NONMEM, pharmacokinetics, pharmacodynamics, VEGF, SLD, targeted therapies, IRT, FDG-PET, SUVmax

Emilie Schindler, Department of Pharmaceutical Biosciences, Box 591, Uppsala University, SE-75124 Uppsala, Sweden.

ISSN 1651-6192
ISBN 978-91-513-0191-4
urn:nbn:se:uu:diva-336420 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-336420)

Don't play what's there, play what's not there.

Miles Davis

To my family

List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I **Schindler E**, Amantea MA, Karlsson MO, Friberg LE. (2017) A pharmacometric framework for axitinib exposure, efficacy, and safety in metastatic renal cell carcinoma patients. *CPT Pharmacometrics Syst Pharmacol.* 6(6):373-382.

II **Schindler E**, Amantea MA, Karlsson MO, Friberg LE. (2016) PK-PD modeling of individual lesion FDG-PET response to predict overall survival in patients with sunitinib-treated gastrointestinal stromal tumor. *CPT Pharmacometrics Syst Pharmacol.* 5(4):173-81.

III **Schindler E**[*], Krishnan SM[*], Mathijssen R, Ruggiero A, Schiavon G, Friberg LE. (2017) Pharmacometric modeling of liver metastases' diameter, volume, and density and their relation to clinical outcome in imatinib-treated patients with gastrointestinal stromal tumors. *CPT Pharmacometrics Syst Pharmacol.* 6(7):449-457. [*] Shared first authorship

IV **Schindler E**, Karlsson MO. (2017) A minimal continuous-time markov pharmacometric model. *AAPS J.*, 19(5):1424–1435.

V **Schindler E**, Friberg LE, Lum BL, Wang B, Quartino A, Li C, Girish S, Jin JY, Karlsson MO. A pharmacometric analysis of patient-reported outcomes in breast cancer patients through item response theory. *In manuscript.*

Additional papers not included in this thesis:

Panoilia E, **Schindler E**, Samantas E, Aravantinos G, Kalofonos HP, Christodoulou C, Patrinos GP, Friberg LE, Sivolapenko G. (2015) A pharmacokinetic binding model for bevacizumab and VEGF165 in colorectal cancer patients. Cancer Chemother Pharmacol., 75(4):791-803.

Bender BC, **Schindler E**, Friberg LE. (2015) Population pharmacokinetic-pharmacodynamic modelling in oncology: a tool for predicting clinical response.Br J Clin Pharmacol., 79(1):56-71.

Contents

Abbreviations

AUC	Area under the concentration-time curve
BCS	Breast cancer subscale
b.i.d.	Twice a day
CI	Confidence interval
C_{min}	Minimum concentration
CV	Coefficient of variation
CT	Computed tomography
CTMM	Continuous-time Markov model
dBP	Diastolic blood pressure
DTMM	Discrete-time Markov model
EBE	Empirical Bayes estimate
ECOG	Eastern Cooperative Oncology Group
E_{max}	Maximum effect
FDG	fluorodeoxyglucose
FDG-PET	fluorodeoxyglucose positron emission tomography
FACT-B	Functional assessment of cancer therapy
FOCE(-I)	First order conditional estimation (with interaction)
GIST	Gastro-intestinal stromal tumor
HER2	Human epidermal growth factor receptor 2
HFS	Hand-foot syndrome
HR	Hazard ratio
HRQoL	Health related quality of life
HU	Hounsfield
ICC	Item characteristic curve
IDR	Indirect response model
IIV	Inter-individual variability
ILV	Inter-lesion variability
IPP	Individual pharmacokinetic parameters
IPPSE	Individual pharmacokinetic parameters standard errors
IRT	Item response theory
LABC	Locally advanced breast cancer
mBC	Metastatic breast cancer
mCTMM	Minimal continuous time Markov model
MET	Mean equilibration time
MID3	Model-informed drug discovery and development
mRCC	Metastatic renal cell carcinoma

MRI	Magnetic resonance imaging
MRT	Turnover time
MTD	Maximum transaxial diameter
NCI-CTCAE	National Cancer Institute's Common Terminology Criteria for Adverse Events
NLME	Nonlinear mixed effects
OFV	Objective function value
OS	Overall survival
pcVPC	Prediction-corrected visual predictive check
PD	Pharmacodynamics
PDGFR	Platelet-derived growth factor receptor
PET	Positron emission tomography
PFS	Progression-free survival
PK	Pharmacokinetics
PKPD	Pharmacokinetic-pharmacodynamic
pNET	Pancreatic neuroendocrine tumor
PO	Proportional odds
PPP&D	Population pharmacokinetic parameters and data
PRO	Patient-reported outcome
PsN	Perl-speaks-NONMEM
q3w	Every 3 weeks
q.d.	Once daily
RCC	Renal cell carcinoma
RECIST	Response evaluation criteria in solid tumors
RSE	Relative standard error
RUV	Residual unexplained variability
SCM	Stepwise covariate model building procedure
sKIT	Soluble stem cell factor receptor
SLD	Sum of longest diameters
SUV_{mean}	Mean standardized uptake value
SUV_{max}	Mean standardized uptake value
sVEGFR	Soluble vascular endothelial growth factor receptor
T-DM1	Ado-trastuzumab emtansine
TKI	Tyrosine kinase inhibitor
TTE	Time-to-event
TTP	Time-to-progression
V_{actual}	Actual volume
VEGF	Vascular endothelial growth factor
VEGFR	Vascular endothelial growth factor receptor
$V_{ellipsoid}$	Ellipsoidal volume
VPC	Visual predictive checks

Introduction

Evidence of *clinical benefit,* including improvement in survival and in a patients' quality of life, is essential for regulatory approval. In anticancer drug development, high attrition rate is a major issue: the likelihood of approval by the U.S. Food and Drug Administration is less than 40% for drugs in phase III.[1] Most of the late phase attritions are due to a lack of efficacy, suggesting that dose-response relations are not always well understood from earlier phases and that phase III designs are not optimized.[2]

Targeted therapies interfere with molecular targets that are involved in the growth, progression and spread of cancer. Over the past decade, these therapies have come to the forefront of cancer treatment. Traditional strategies for selecting doses (maximum tolerated dose) and evaluating the efficacy of cytotoxic chemotherapies (e.g. objective tumor response) may not be suitable for targeted agents, stressing the need for new approaches.[3] The value of model-based approaches has been recognized to improve clinical trial design and analysis, and better predict outcomes.[4, 5] Model-informed drug discovery and development (MID3) enables the integration of quantitative information collected across the drug development, including *pharmacokinetic* (PK) and *pharmacodynamic* (PD) data.[6] By increasing the understanding of the exposure-response relationships for the desired and adverse effects, MID3 approaches have the potential to guide rationale selection of dose levels and dosing schedules. MID3 also provides a framework for comparing candidate biomarkers through the quantification of their relation to drug exposure and clinical outcome. These may help to early identify patients that are the most likely to benefit from treatment. Besides their applications in drug development, model-based approaches may also serve as a tool for optimizing treatment in routine clinical practice, e.g. by individualizing doses.[7]

Efficacy endpoints and biomarkers in cancer trials

Definitions

A *clinical endpoint* is defined as a "characteristic or variable that reflects how a patient feels, functions or survives".[8] Clinical endpoints are used as direct measures of clinical benefit and have been the basis for drug approval. A *biomarker* is a "characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to therapeutic intervention". Biomarkers serve for the diagnosis and staging of a disease, for evaluating the disease prognosis, and for predicting and monitoring the response to a treatment. *Surrogate endpoints* are a subset of biomarkers that can predict clinical benefit and substitute for a clinical endpoint. Established surrogate endpoints may be used to support regular approval.

Traditional endpoints

The preferred endpoint for assessing clinical benefit of anti-cancer therapies is *overall survival* (OS), defined as the time from randomization into a trial until death from any cause. It is considered precise and unambiguous.[9] Yet, its determination usually requires a large number of patients and/or long study duration for cancers with good prognosis. In addition, it may be affected by the trial design (e.g. crossover) and subsequent therapies.

A number of surrogate endpoints that inform on tumor response and disease progression in a shorter timeframe compared to OS are commonly used to support decisions and drug approval.[9, 10, 11] The Response Evaluation Criteria in Solid Tumors (RECIST) are the most widely adopted criteria for assessing objective changes in tumor burden.[10] They are based on a categorization of the change in *sum of longest diameters* (*SLD*) of target lesions, assessed via computed tomography (CT) or magnetic resonance imaging (MRI), and on new lesions' appearance. In the latest version of RECIST (v 1.1), up to five target lesions may be defined, with a maximum of two per organ. RECIST defines four response categories: complete response, partial response, stable disease, and progressive disease. These are further used to define *progression-free survival* (*PFS*), a surrogate endpoint corresponding to the time from randomization until objective tumor progression or death (Figure 1).[9, 12]

Although PFS and other RECIST-based endpoints are easy to implement, they all result in a loss of valuable quantitative information due to the categorization of tumor response. Moreover, using SLD implies that information on individual lesion's (lack of) response within an individual is disregarded. Additionally, since these criteria are not based on the full tumor size time-

course, variations of the response over time are partly neglected. Analyzing longitudinal tumor data on a continuous scale using mathematical models has the potential to overcome these limitations.[13]

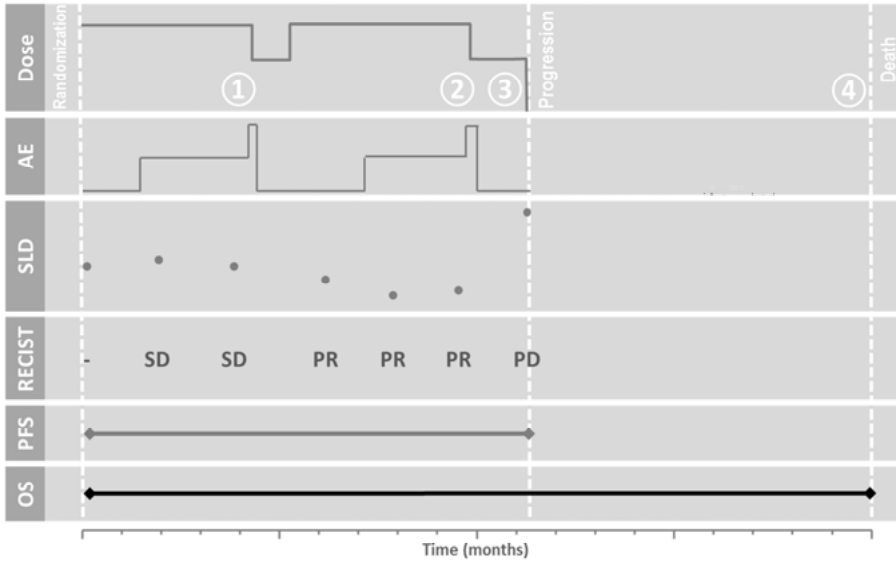

Figure 1. Schematic representation of the data and traditional endpoints collected during a clinical trial for a hypothetical patient who underwent transient dose reductions due to adverse events (AE, ①,②), discontinued therapy due to progressive disease (PD, ③) after first experiencing stable disease (SD) and partial response (PR), and died during follow-up (④).

RECIST criteria were established when anti-cancer treatments mostly consisted of cytotoxic chemotherapies. Growing evidence indicates that the above mentioned endpoints are suboptimal for assessing the efficacy of modern targeted therapies, for which prolonged OS is seen even though tumor shrinkage is rare or delayed.[14]

New biomarkers for targeted therapies

Thanks to the development of molecular biomarkers related to the mechanism of action of targeted therapies and of novel imaging techniques (e.g. functional and volumetric imaging) new efficacy measures are becoming available. These measures may in the future supplement or replace SLD-based endpoints but require stringent validation and their relation to drug exposure and OS must be confirmed.[15, 16] Using markers that enables early assessment of therapeutic response is crucial to distinguish between responders and non-responders, and therefore tailor therapy accordingly to

save costs and prevent patients from being administered with ineffective therapies.

Circulating biomarkers for anti-angiogenic drugs

Angiogenesis is the process whereby new blood vessels form and has an essential role in tumor growth and spread by ensuring adequate supply of oxygen and nutrients and removal of waste products.[17] Antiangiogenics are targeted compounds that keep new blood vessels from forming, thereby causing the tumor to starve, stop growing or shrink.[12] Most of them block the vascular endothelial growth factor (VEGF) pathway, for example through the inhibition of VEGF receptors (VEGFR)-1, -2, -3 which are upregulated in numerous solid tumors.[18] Examples of such agents include the tyrosine kinase inhibitors (TKI) sunitinib, axitinib and pazopanib. Anti-VEGF agents have demonstrated significant improvement in OS in chemotherapy-refractory tumors, such as renal cell carcinoma (RCC) and gastro-intestinal stromal tumors (GIST).[19]

Soluble forms of VEGFR (sVEGFR)-1, -2, -3 result from alternative splicing or shedding of the cell surface receptors and can be detected in plasma.[20] Changes in VEGF or sVEGFR-1, -2, -3 plasma levels during anti-VEGF therapy are thought to reflect systemic effects of the drug, although mechanisms behind the drug-induced change in their concentrations are not fully understood. The stem cell factor receptor (c-KIT) is inhibited by sunitinib and imatinib, and the levels in its soluble form (sKIT) decrease during sunitinib therapy in GIST.[21] Such biomarkers can therefore be important for evaluating the activity of anti-angiogenic agents and for assessing the degree of target inhibition caused by treatment. Moreover, circulating biomarkers are attractive as they are minimally invasive and can be collected easily and repeatedly over the course of a therapy. However, the impact of these changes on tumor response and their predictive ability on long-term clinical outcome remain unclear in most cancers.[14, 20]

Modeling analyses can be used to gain insight on these relationships, e.g. by allowing differences between patients and variables with different dynamics to be described. Hansson et al. developed an integrated modeling framework for imatinib-resistant GIST treated with sunitinib.[22, 23] sVEGFR-3 and sKIT dynamics, together with sunitinib exposure, were predictive of SLD time-course. Moreover, more pronounced decreases in sVEGFR-3 levels and lower baseline SLD were predictive of better OS. In another modeling analysis in hepatocellular carcinoma, sunitinib-induced inhibition of sVEGFR-2 release was linked to tumor growth inhibition and predictive of time-to-progression (TTP). Conversely, in sunitinib-treated metastatic renal cell carcinoma (mRCC) patients, sVEGFR-2 baseline, but not sVEGFR-2 or -3 dynamics, were predictive of PFS, and in metastatic colorectal cancer none of the sVEGFR-related metrics could predict TTP.

Imaging biomarkers

A key question considered by the RECIST Working Group concerns the use of alternative imaging biomarkers such as tumor metabolic activity, tumor volume and tumor density.[10] They are particularly appealing as early indicators of treatment effect for targeted drugs.

Tumor metabolic activity

Tissue metabolic activity can be assessed by [^{18}F]-fluorodeoxyglucose (FDG) positron emission tomography (PET). FDG is a radio-labeled glucose analogue taken up by cells and accumulating in cells with a high glycolytic activity, including some cancer cells such as in GIST. FDG-PET has the advantage over CT to differentiate between viable tumor tissue and scar or fibrotic tissue.[24] The degree of FDG accumulation in a tissue can be quantified by the *standardized uptake value (SUV)*, calculated as shown in Figure 2.[25] Various FDG-PET metrics have been described, including SUV_{max} and SUV_{mean} (Figure 2). Owing to difficulties in defining the boundaries of the region of interest (ROI), SUV_{max} is often considered as more reliable and accurate than SUV_{mean}. Due to tumor heterogeneity SUV_{max} does not always correlate with SUV_{mean}.

Clinical research has demonstrated that metabolic changes during therapy are correlated with patient outcome in various solid tumors.[26] Response assessment by FDG PET/CT has been formalized by the European Organization for Research and Treatment of Cancer (EORTC) PET Study Group[27], and more recently by the PET Response Criteria in Solid Tumors (PERCIST), which provides a detailed framework for lesion selection, region of interest definition, and response classification[24]. In GIST patients, SUV decrease has been reported as early as 24 h after the start of imatinib treatment[28] and in another study at the first measurement, seven days after the start of sunitinib[29]. These early changes in tumor metabolic activity often appeared weeks or months earlier than changes in tumor size. However, to date the published results largely diverge concerning the predictive value of the FDG-PET response with respect to PFS or OS in GIST. Moreover, no consensus exists about the time intervals at which the response evaluation of targeted therapies by FDG-PET should be assessed.[30]

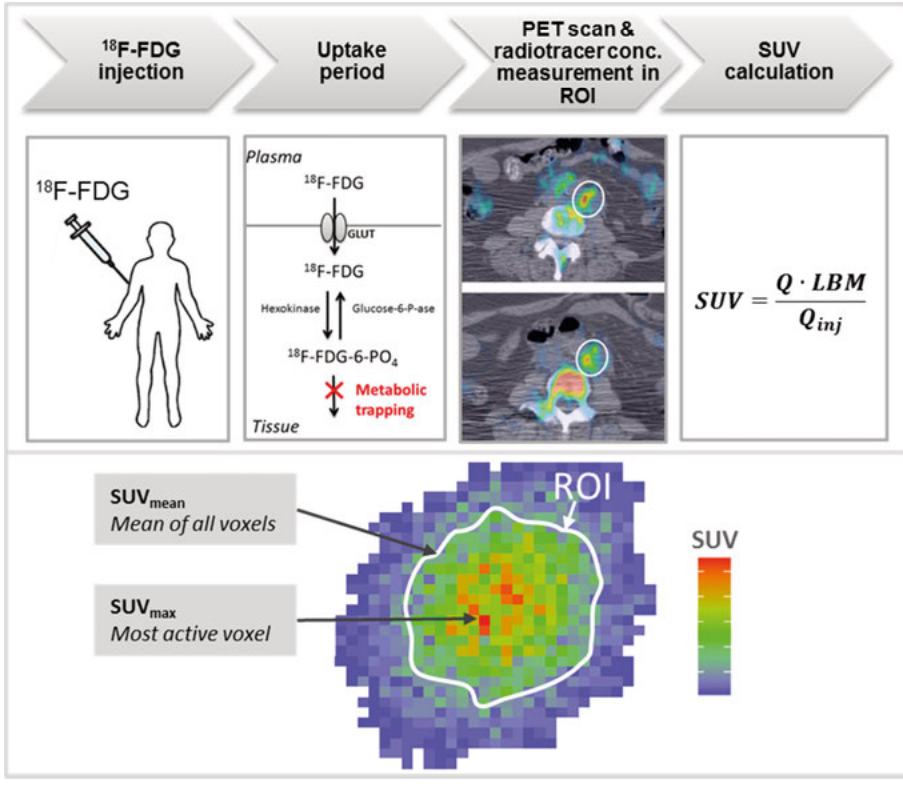

Figure 2. Schematic representation of FDG-PET assessment (top), and SUV metrics (bottom). GLUT: glucose transporter. ROI : region of interest; Q: radiotracer activity in the ROI; LBM: lean body mass; Q_{inj}: injected activity. PET scan image was originally published in JNM by Vanderhoek et al. by the Society of Nuclear Medicine and Molecular Imaging, Inc.[25]

Tumor volume

Tumor volume response may be more informative than unidimensional measurements when assessing irregularly shaped tumors with non-uniform size changes, such as head and neck cancers[31], pancreatic ductal adenocarcinomas[32], lung cancer[33] and GIST[34, 35]. In fact, RECIST may not be sensitive at detecting patients with partial response in non-spherical tumors that shrink along their short axis. The tumor volume can be measured with semi-automated procedures, or approximated by geometric formulas (ellipsoidal volume) based on unidimensional measures along three axes (Figure 3).

In the analysis by Schiavon et al. in imatinib-treated GIST patients, ellipsoidal three-dimensional measurements discriminated size changes of liver metastases better than unidimensional measurements.[25, 26] Moreover, three-dimensional measurements tended to be better associated with OS than unidimensional measurements when using traditional statistical analysis.

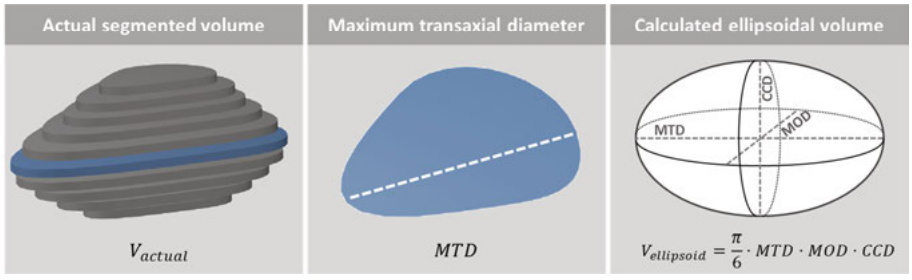

Figure 3. Uni-dimensional and three-dimensional tumor measurements. MTD: maximum transaxial diameter MOD: maximum orthogonal diameter; CCD: craniocaudal diameter

Tumor density

Following treatment with antiangiogenic agents, some tumors may initially increase in size while demonstrating a decrease in density and metabolic activity. This may be explained by drug-induced tumor necrosis, central cavitation and hemorrhage that do not reflect true tumor progression. Biological changes in tumors are not accounted for by RECIST, leading to an underestimation of tumor response, especially at early stages. Choi et al. proposed a criterion based on a categorization of changes in tumor size and in tumor density. *Tumor density* (in Hounsfield units, HU) can be readily measured on CT scans.[36] Density-based criteria have been associated with long-term benefit in other clinical settings, e.g. advanced hepatocellular carcinoma treated with sorafenib[37], mRCC patients treated with sunitinib[38], and metastatic breast cancer treated with bevacizumab[39]. Contrasting results have been reported regarding the predictive value of Choi criteria for OS in GIST.

Adverse events

Adverse effects related to cancer therapy are typically graded by the National Cancer Institute's Common Terminology Criteria for Adverse Events (NCI-CTCAE) according to their intensity, as exemplified in Table 1.[40]

For inhibitors of the VEGF pathway, an association between adverse effects related to their mechanism of action and clinical outcome has been shown. The most common adverse effect include hypertension, proteinuria [41] and hand-foot syndrome (HFS)[42]. In particular, treatment-related increase in blood pressure may indicate a successful inhibition of angiogenesis and has been associated with improved PFS and OS in mRCC[43] and GIST patients[44] receiving sunitinib. In sunitinib-treated GIST, the results from a pharmacometric analysis showed that neutrophil count and diastolic blood pressure (dBP) time-courses could, in combination, predict OS.

Table 1. *Examples of adverse event grading according to NCI-CTCAE version 3.0.*[a]

Adverse event	Grade 0 (No event)	Grade 1 (Mild)	Grade 2 (Moderate)	Grade 3 (Severe)	Grade 4 (Life-threatening/ disabling)
Fatigue	No event	Mild fatigue over baseline	Moderate or causing difficulty performing some ADL[b]	Severe fatigue interfering with ADL[b]	Disabling
Hand-foot syndrome	No event	Minimal skin changes or dermatitis (e.g., erythema) without pain	Skin changes (e.g., peeling, blisters, bleeding, edema) or pain, not interfering with function	Ulcerative dermatitis or skin changes with pain interfering with function	-

[a] Used in Paper III; most recent version is 4.0. [b] Activities of daily life.

Patient-reported outcomes

The introduction of targeted therapies has led to a dramatic increase in clinical response rates and OS duration. Patients who tolerate and respond well to a targeted drug may stay on therapy for years, turning cancer into a chronic illness. Interest has increased in incorporating patient-reported outcome (PRO) data in clinical trials and routine clinical practice.[45, 46] PRO measures are reports about how a patient feels and functions in relation to a disease and its treatment; PROs come directly from the patient without interpretation of the response by a clinician or any third party.[47] PROs are often collected as questionnaires consisting of several items (questions), which inform on single or multiple aspects of health related quality of life (HRQoL). These include disease symptoms, drug-related toxicities, physical functioning, and social and psychological well-being.

The Functional Assessment of Cancer Therapy-Breast (FACT-B) is a multi-dimensional PRO questionnaire consisting of 36 items grouped into five subscales: the physical, social/family, emotional and functional well-being subscales, which form the FACT-General scale, and an additional breast cancer-specific subscale (BCS)[48]. For each item, patients can respond "Not at all", "A little bit", "Somewhat", "Quite a bit" and "Very much". These answers are converted into scores ranging from 0 to 4, where higher scores reflect better well-being.

Scoring of multi-item PRO such as FACT-B traditionally relies on sum of the item scores, from which criteria based on expert opinion, such as time to symptom worsening, can be derived. While being quick to compute and easy to interpret, these composite scores result in a loss of information at both longitudinal and item levels. Moreover, technical challenges related to missing data often arise.

Pharmacometrics

Pharmacometrics is defined as "the science of developing and applying mathematical and statistical methods to characterize, understand and predict drug's PK, PD and biomarker-outcomes behavior".[49] In oncology, PD encompasses a wide variety of observation types, including plasma biomarkers, tumor size, adverse effects and time-to-event (TTE) outcomes (e.g. OS). Changes in these variables can occur on widely different time-scales. Central to pharmacometrics, population pharmacokinetic-pharmacodynamic (PKPD) models provide a quantitative description of the dose-concentration-effects relationships and integrate the multi-level drug action in a unified framework, often involving a set of algebraic or differential equations.[2, 50, 51, 52] By including prior knowledge on the physiology, disease, pharmacology and drug's mechanism of action, mechanistically and physiologically relevant models can be developed. Such models may be used for clinical trials simulations to predict clinical outcome in untested scenarios such as untested doses or new dosing schedules. Among their many applications throughout the lifecycle of drug development, pharmacometric models can be used to predict clinical outcome from preclinical and early clinical data, inform clinical trial design (sample size, optimal time points for data collection), guide dose and schedule selection, identify sources of variability and suggest strategies for dose adjustment in patient subgroups.[6] Additionally, even though its use remains scarce in oncology routine clinical practice, pharmacometrics has potential applications as a tool toward optimizing the benefit–risk profile through dose adaptation strategies for individualized dosing.[2, 7]

Pharmacometric analyses mostly rely on *nonlinear mixed effects* (NLME) models, also known as population models, to describe PKPD relationships.

Models for continuous variables

For continuous data, such as drug concentration or tumor size, a NLME model describes the observation y_{ij} (the dependent variable) for individual i at observation time t_{ij} (the independent variable) as:

$$y_{ij} = f\left(t_{ij}, \theta, X_i, \eta_i\right) + \varepsilon_{ij} \quad (1)$$

The term *mixed effects* refers to the combination of *fixed effects*, represented by a vector of parameters θ characterizing the typical individual, and *random effects* describing the variability in the population.[53] The latter can be partitioned into the inter-individual variability (IIV) and the residual unexplained variability (RUV). The IIV is represented by a vector of individual-specific random effect parameters η_i. The RUV, represented by the random variable ε_{ij}, is the deviation between the observations and model predictions. η_i and ε_{ij} are generally assumed to be normally distributed with mean zero and respective variance-covariance matrix Ω and Σ. Additional levels of

variability may be considered based on the data. X_i is a matrix of patient-specific covariates (e.g. dose, genotype), that may or may not vary over time and explain part of the observed variability. Importantly, by simultaneously fitting data from multiple subjects, NLME models allow analyzing sparse data (e.g. few samples per subject) and unbalanced data (e.g. different sampling times).

For example, models describing tumor size time-courses often utilize ordinary differential equations.[52] Claret et al. proposed a model that incorporates a first-order tumor growth rate (denoting exponential growth), a drug effect driven by drug exposure inducing tumor size reduction, and an exponential decrease in drug effect over time denoting resistance appearance or tumor regrowth.[54]

Time-to-event models

Time to clinical outcome data (e.g. OS) can be described by parametric TTE models, in which the hazard function characterizes the instantaneous rate at which an event occurs (*h(t)*, Equation 2), and the survival function indicates the probability of surviving beyond time *t* (*S(t)*, Equation 3).

$$h(t) = h_0(t) \cdot e^{f(\boldsymbol{\theta}, \boldsymbol{X_i})} \quad (2)$$

$$S(t) = e^{-\int_0^t h(t)dt} \quad (3)$$

$h_0(t)$ represents the baseline hazard, which may vary over time. $\boldsymbol{\theta}$ is a vector of estimated coefficients that inform on the size of the impact of a vector of patient-specific explanatory variables $\boldsymbol{X_i}$ on the hazard. Investigated predictors may include baseline patient characteristics (e.g. baseline tumor size) or time-varying variables (e.g. biomarker or tumor response time-courses).

Models for ordered categorical data

NCI-CTCAE grading of adverse effects gives rise to ordered categorical data, which have mostly been modeled using proportional odds (PO) models. When adverse effects are collected frequently, e.g. daily, the grades for two consecutive assessments may be inter-correlated beyond what is expected after accounting for standard predictors (time, dose, etc.). This property, referred to as Markov property, is handled by Markov models, including discrete-time Markov models (DTMM) and continuous-time Markov models (CTMM). Advantages and limitations of these models are summarized in Table 2 and their schematic representations are provided in Figure 4.

Table 2. *Advantages and limitations of existing models for ordered categorical data*

Model	Advantages	Limitations
PO	Easy to implement Short run times Easy introduction of predictors	Does not account for Markov elements
DTMM	Transitions can occur between any two states Accounts for Markov elements Applicable to data with uniform observation intervals No assumption about intermediate states	Number of estimated parameters increases dramatically with number of states Not optimal when non-uniform time intervals Complex introduction of predictors
CTMM	Accounts for Markov elements Applicable to data with uniform and non-uniform observation intervals Fewer estimated parameters compared to DTMM	Transitions only occur between neighboring states Long runtimes (differential equations), not practically applicable when large number of states Complex introduction of predictors

In a PO model, the probability that an observation Y_{ij} is greater than a score k ($P(Y_{ij} \geq k)$) is described as a function $f(\boldsymbol{\theta}, \boldsymbol{X_i})$ of patient-specific explanatory variables (vector $\boldsymbol{X_i}$), and their associated coefficients (vector $\boldsymbol{\theta}$), and of an intercept α_k. The probability of scoring exactly k ($P(Y_{ij}=k)$) can then be derived as in Table 3 (exemplified for a 4-state variable, such as fatigue or HFS).

Table 3. *Parameter and equations for a PO model describing a 4-state variable.*

Score k	Intercept α_k	$P(Y_{ij} \geq k)$	$P(Y_{ij} = k)$
0	-	1	$1 - P(Y_{ij} \geq 1)$
1	α_1	$\frac{1}{1 + e^{\alpha_1 + f(\boldsymbol{\theta}, \boldsymbol{X_i})}}$	$P(Y_{ij} \geq 1) - P(Y_{ij} \geq 2)$
2	$\alpha_2 = \alpha_1 + b_2$	$\frac{1}{1 + e^{\alpha_2 + f(\boldsymbol{\theta}, \boldsymbol{X_i})}}$	$P(Y_{ij} \geq 2) - P(Y_{ij} \geq 3)$
3	$\alpha_3 = \alpha_1 + b_2 + b_3$	$\frac{1}{1 + e^{\alpha_3 + f(\boldsymbol{\theta}, \boldsymbol{X_i})}}$	$P(Y_{ij} \geq 3)$

Estimated parameters include α_1 (unbound), b_2 and b_3 (positive), and $\boldsymbol{\theta}$.

DTMM describe the probability $P(Y_{ij}=k_j | Y_{i(j-1)}=k_{j-1})$ of transiting from a score k_{j-1} at the previous observation ($Y_{i(j-1)}$) to a score k_j at the current observation (Y_{ij}), in a time-independent manner. For a 4-state variable, the DTMM can be obtained by extending the PO model equations in Table 3, with α_1, b_2, b_3, and the effect of predictors $f(\boldsymbol{\theta}, \boldsymbol{X_i})$ being conditional on the previous score.

In the CTMM, $P(Y_{ij}=k)$ are defined by ordinary differential equations (e.g. Equations 4-7 for a 4-state variable). CTMM typically assume that transitions (observed or not observed) only occur between neighboring states. The transfer rate constants λ govern the rate at which the probability amounts

distribute between states, between two observation times. At the time of observation, the probability of the actual observed score is set to 1, while the other probabilities are set to 0.

$$\frac{dP(Y=0)}{dt} = \lambda_{10} \cdot P(Y=1) - \lambda_{01} \cdot P(Y=0) \quad (4)$$

$$\frac{dP(Y=1)}{dt} = \lambda_{01} \cdot P(Y=0) + \lambda_{21} \cdot P(Y=2) - (\lambda_{10} + \lambda_{12}) \cdot P(Y=1) \quad (5)$$

$$\frac{dP(Y=2)}{dt} = \lambda_{12} \cdot P(Y=1) + \lambda_{32} \cdot P(Y=3) - (\lambda_{21} + \lambda_{23}) \cdot P(Y=2) \quad (6)$$

$$\frac{dP(Y=3)}{dt} = \lambda_{23} \cdot P(Y=2) - \lambda_{32} \cdot P(Y=3) \quad (7)$$

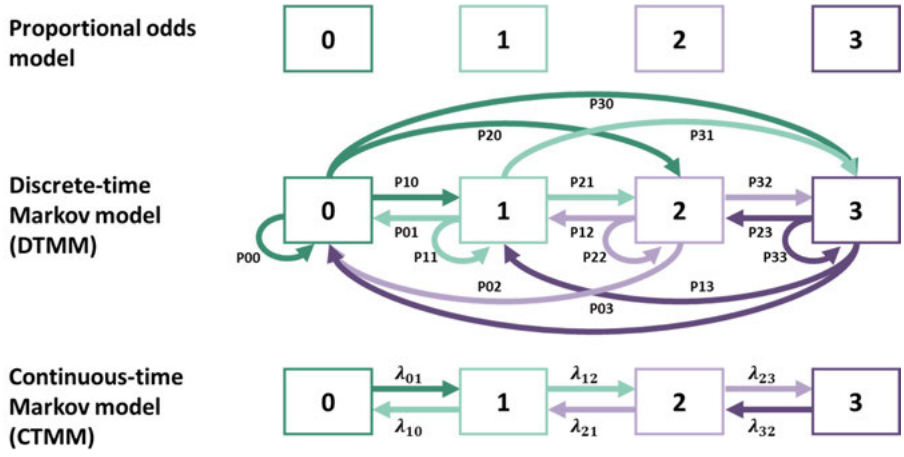

Figure 4. Schematic representation of existing models for ordered categorical data. P_{yx}: transition probability from state x to state y. λ_{xy}: transfer rate constant from state x to state y. Adapted from Schindler and Karlsson (2017).[55]

Item response theory models

Item response theory (IRT) is a modern test theory, originally developed in psychometrics and education to measure psychological traits or proficiency. In recent years, IRT has gained in popularity and acceptance in PRO research.[56, 57] IRT models describe the probability of a patient's response to each item of a questionnaire as a function of a latent (unobserved) variable, which represents the construct that the questionnaire aims to measure. The relationship between the outcome of a question and the latent variable can be visualized with item characteristic curves (ICCs, Figure 5).

In a NLME context, parameters describing the shape of the ICCs can be modelled as fixed effects and the latent variable can be described as an individual-specific random effect. Effects of covariates and drug exposure can be investigated on the latent variable using longitudinal questionnaire data.

IRT implemented in a NLME framework has been applied in Alzheimer's disease[58], multiple sclerosis[59], schizophrenia[60], and Parkinson's disease[61, 62] and has demonstrated more precise and effective analysis of composite score data with increased power to detect drug effect compared to total score-based methods.

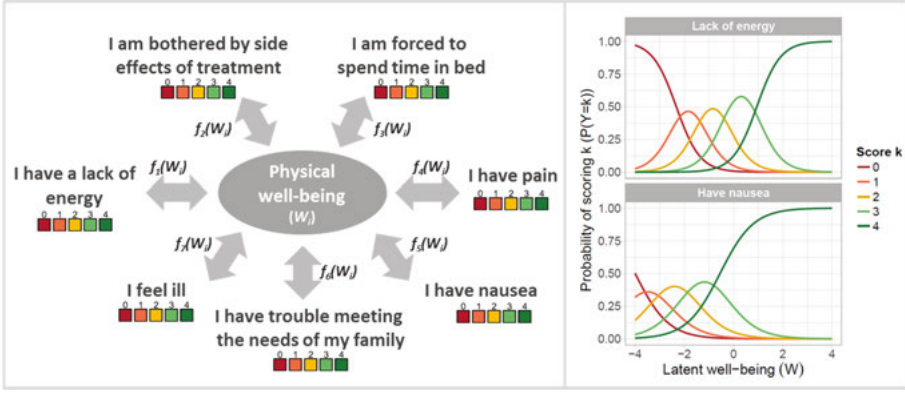

Figure 5. Schematic representation of an IRT model exemplified for the physical well-being subscale of FACT-B questionnaire (left), and example of item characteristic curves for two items (right). Higher scores denote better well-being.

Modeling of multiple variables

Simultaneous analysis of several variables (e.g. PK and PD, or biomarker and TTE) is generally considered as the reference method as it allows for a precise estimation of all model parameters[63]. However, it may not always be possible due to long run times and model instability. Alternative sequential PK-PD approaches include the individual PK parameters (IPP) method, where the empirical Bayes estimates (EBEs) of the individual PK parameters are used to predict PK, assuming that they are estimated without error. In the population PK parameters and data (PPP&D) approach, populations PK parameters are fixed to their final estimates and individual PK parameters are predicted simultaneously with PD parameters, using both PK and PD data.[64] Finally, the IPP standard error (IPPSE) method is an extension of the IPP method, accounting for the uncertainty in individual PK parameters.[65] When using a sequential modeling approach, the use of PPP&D and IPPSE that minimize the bias in parameter uncertainty should be preferred over IPP approach. The latter may underestimate the effect of a biomarker on the TTE outcome.

Maximum likelihood estimation

Maximum likelihood estimation is a method used to estimate model parameters by finding the parameter values that maximize the likelihood of observing the data given the parameters. For the individual I, the likelihood L_i is derived as a function of a vector of ω (**Ω**) and a vector of σ (**Σ**)[66]:

$$\mathrm{L_i}(\theta, \Omega, \Sigma | y_i) = \mathrm{p}(y_i | \theta, \Omega, \Sigma) = \int \mathrm{p}(y_i | \eta_\mathrm{i}, \theta, \Sigma) \cdot \mathrm{p}(y_i | \eta_\mathrm{i}, \theta, \Omega)\, d\eta \quad (8)$$

Numerical approximations are generally needed for the calculation of individual likelihood. Approximation used in this thesis include the first order conditional estimation method (with interaction), or FOCE(-I), and the second-order Laplacian method. Both are gradient-based methods based on the derivative of the approximation of the log-likelihood. Minimizing the log-likelihood is often easier than maximizing the likelihood. Minus two times the log-likelihood provides the objective function value (OFV). Individual OFVs can be summed to obtain the total OFV.

Studied drugs

Tyrosine kinase inhibitors

Tyrosine kinases are enzymes involved in specific cell signaling pathways necessary for tumor growth and progression, and represent appealing targets for anti-cancer drugs. Numerous TKIs have been introduced in the past decade.

Imatinib mesylate (Gleevec®) inhibits the Breakpoint cluster region–Abelson proto-oncogene kinase (Bcr-Abl), the platelet-derived growth factor receptors (PDGFR-α and PDGFR-β), and KIT.[67] Its indications include the treatment of KIT-positive GIST that cannot be surgically removed and/or have spread to other parts of the body.

Sunitinib malate (Sutent®) exerts both antiangiogenic and antitumor effects through its activity against VEGFR-1, -2 and -3, PDGFR-α and -β, KIT and other tyrosine kinases.[68] Sunitinib is approved multi-nationally (U.S., Europe, etc.) for the treatment of imatinib-resistant or –intolerant GIST, advanced RCC, and highly differentiated unresectable pancreas neuroendocrine tumors (pNET). In GIST and RCC, sunitinib is typically administered an intermittent dosing schedule (50 mg once daily on 4 weeks on treatment followed by 2 weeks off) whereas in pNET it is administered on a continuous dosing schedule (37.5 mg once daily).

Axitinib (Inlyta®) is an oral selective inhibitor of VEGFR-1, -2 and -3.[69] It is approved for the treatment of RCC after failure of one prior systemic therapy at a starting dose of 5 mg twice daily with a possibility of dose escalation based on individual safety and tolerability.

Lapatinib (Tykerb®) is a dual TKI directed against the human epidermal growth factor receptor 2 (HER2) and the epidermal growth factor receptor.[70] Lapatinib is indicated in combination with capecitabine (Xeloda®) for the treatment of advanced or metastatic breast cancer (mBC) patients whose tumors overexpress HER2 and who have received prior therapy including an anthracycline, a taxane, and trastuzumab (Herceptin®).

Ado-trastuzumab emtansine

Ado-trastuzumab emtansine (Kadcyla®), or T-DM1, is an antibody-drug conjugate composed of the cytotoxic agent emtansine or DM1 (microtubule inhibitor) conjugated to trastuzumab, a humanized monoclonal antibody directed against HER2.[71] T-DM1 is indicated as a single agent for the treatment HER2-positive mBC patients who previously received trastuzumab and a taxane, separately or in combination.

Aims

The overall aim of this thesis was to develop pharmacometric approaches that could support the assessment of clinical benefit of anticancer therapies. The developed models aimed to characterize and quantify the relationships between drug exposure, efficacy biomarkers, adverse effects, survival and patient-reported outcomes.

The specific aims are:

- To investigate the relationships between drug exposure, plasma biomarker levels, tumor size and hypertension, and compare their predictive ability for overall survival in advanced renal cell carcinoma patients treated with axitinib.

- To characterize longitudinal lesion-level response assessed by functional (FDG-PET) and morphological (density, three-dimensional) imaging techniques in gastrointestinal stromal tumors (GIST) patients treated with tyrosine kinase inhibitors, and to compare their predictive ability of clinical outcome to uni-dimensional tumor measurements.

- To develop a methodology to facilitate the analysis of ordered categorical adverse effect data with Markovian features.

- To establish an item response theory (IRT) pharmacometric framework that can describe longitudinal item-level patient-reported outcome (PRO) data and be applied to evaluate potential exposure-response relationships, as well as utilize the approach to compare PRO responses of ado-trastuzumab emtansine (T-DM1) and lapatinib-plus-capecitabine treatments for locally advanced/metastatic breast cancer.

Methods

Patients and data

Clinical trial data from several phase I-III studies involving metastatic or advanced cancer patients treated with targeted therapies were used for the development of pharmacometric models. The study designs and available data are outlined in Table 4.

All interventional clinical studies (Papers I, II, IV, V) used in this thesis were conducted in accordance with the ethical principles originating in or derived from the Declaration of Helsinki, and with Good Clinical Practice guidelines. The study protocols were approved by the relevant institutional review board and/or independent ethics committee at each study site. Written informed consent was obtained from all patients.

The non-interventional retrospective study in imatinib-treated GIST patients (Paper III) was approved by the Institutional Review Board at Leuven University Hospitals. The submission to the Ethic Committee at Erasmus University Medical Center in Rotterdam, The Netherlands was not mandatory, since scans performed at the time of the diagnosis and during treatment were retrieved retrospectively, made anonymous and coded. No informed consent statement was obtained from patients, the study being retrospective.[34, 35]

In addition to these oncology studies, an additional data set was utilized to demonstrate the predictive performance of the new minimal CTMM (mCTMM) methodology (Paper IV). The dataset included 22492 Likert pain score (ranging from 0 to 10) data collected in 231 placebo-treated subjects with painful distal diabetic neuropathy involved in three phase III studies.[72] Median follow-up duration was 124.5 days (range, 2.5-125.5). Intake of rescue medicine (acetaminophen) was recorded daily.

Table 4. *Overview of study designs and data*

Paper	Study design	Patients	Treatment	Dosing Schedule	Measurements
I	Single arm, open-label, multicenter phase II study[73]	64 Japanese cytokine-refractory mRCC	Axitinib	Starting dose: 5 mg b.i.d. Dose escalation up to 10 mg b.i.d. or reduction to 2 mg b.i.d. based on tolerability (blood pressure, non-hematologic adverse effects)	VEGF, sVEGFR-1, -2, -3, sKIT Cycle 1: pre-dosing, Cycle 2-7: day 1 EoT/discontinuation SLD Cycle 1: pre-dosing Subsequent odd no. cycles: day 1 EoT/discontinuation dBP[a] Cycle 1: pre-dosing, day 8, 15, 22 Cycle 2: day 1 OS
II	Open-label, multicenter, dose escalation, phase I/II study[29]	66 imatinib-resistant or –intolerant GIST	Sunitinib	Weeks on/off: starting daily dose 2/1: 50 mg q.d. 2/2: 25, 50, 75 mg q.d. 4/2: 50 mg q.d.	SUV_{max} and SUV_{mean}: for up to 6 lesions 2/1: cycle 1: day 0, 7, 21; cycle 8: day 14 2/2: cycle 1: day 0, 7, 28; cycle 4: day 28 4/2: cycle 1: day 0, 7, 42; cycle 4: day 28 SLD 2/1: cycle 1: day 0; cycle 4, 8, 12: day 14 2/2: cycle 1: day 0; cycle 2 and every other cycle: day 28 4/2: cycle 1: 0; cycle 2 and every other cycle: 28 VEGF, sVEGFR-2, sKIT (N=36) OS

Paper	Study design	Patients	Treatment	Dosing Schedule	Measurements
III	Retrospective non-interventional study[22, 23]	77 GIST	Imatinib	Starting dose: 400 or 800 mg q.d. Dose escalation or reduction based on safety and efficacy	MTD, V_{actual}[b], $V_{ellipsoid}$, density Baseline and at least once during treatment (at ~ 3, 6, and 12 months) for 1-2 liver lesions Median follow-up duration: 360 days (range, 82-495) OS: median of 4.5 years (range, 0.79-13) PFS: median 3.4 years (range, 0.25-13)
IV	Four phase I-III studies[22]	303 imatinib-resistant GIST	Sunitinib or placebo	Weeks on/off: starting daily dose 2/1: 50 mg q.d. 2/2: 25, 50, 75 mg q.d. 4/2: 25, 50, 75 mg q.d. Continuous: 37.5 mg q.d.	Fatigue: daily (n=55027, N=303), median follow-up duration of 157 (range, 7-687) days HFS: daily (n=39294, N=251), median follow-up duration of 147 (range, 10-540) days
V	Randomized, open-label, international pivotal phase III study[74]	HER2-positive, unresectable LABC or mBC previously treated with trastuzumab and a taxane	T-DM1 (N=484) vs Capecitabine-plus-lapatinib (N=478)	TDM: 3.6 mg/kg q3w Capecitabine: 1000 mg/m^2 b.i.d., days 1-14, q3w Lapatinib: 1250 mg q.d.	FACT-B: Cycle 1: day 1 (baseline) Odd cycles: day 1 TDM1 arm: n=2655 Capecitabine-plus-lapatinib : n=2192

N: number of patients; mRCC: metastatic renal cell carcinoma; b.i.d.: twice a day; VEGF, vascular endothelial growth factor; sVEGFR-1, -2, -3, soluble VEGF receptor 1, 2, 3; sKIT, soluble stem cell factor receptor; EoT, end of treatment; SLD: sum of longest diameters; dBP: diastolic blood pressure; OS: overall survival; GIST: gastrointestinal stromal tumor; q.d.: once daily; FDG-PET: [^{18}F]-fluorodeoxyglucose positron emission tomography; SUV: standardized uptake value; MTD: maximum transaxial diameter; V_{actual}: actual volume; $V_{ellipsoid}$: ellipsoidal volume; PFS: progression-free survival; HFS: hand-foot syndrome; HER2: human epidermal growth factor receptor 2; LABC: locally advanced breast cancer; MBC: metastatic breast cancer; q3w: every three weeks; FACT-B: functional assessment of cancer therapy-breast. [a] Only dBP data from the first month were used to prevent confounding effect of antihypertensive treatment administered to patients with dBP elevation. [b] calculated with the semi-automated segmentation by 'syngo CT Oncology' software version 2009E, Siemens Medical Solutions, Inc.

Pharmacokinetic metrics

Daily AUC ($AUC_{daily}=Dose_{daily}/(CL/F)$) was calculated for axitinib (Paper I) and sunitinib (Papers II, IV) using the EBEs of apparent clearance (*CL/F*), or its population value when no sunitinib PK data were available; EBEs were obtained from the respective published population PK models.[75, 76] No data were collected for sunitinib equipotent metabolite SU1266. Given the axitinib typical elimination half-life of 3h in Japanese patients, AUC_{daily} was assumed to be zero on days off-therapy. PK samples were not collected in the retrospective study in imatinib-treated GIST (Paper III).

In Paper V, for 331 out of 484 T-DM1-treated patients with available PK data, cycle 1 area under the concentration-time curve ($AUC_{cycle\ 1}$) and cycle 1 minimum concentration ($C_{min,cycle\ 1}$) at a nominal time of 504 hours were obtained from a published population PK model.[77] For T-DM1 patients missing PK data (n=153), $AUC_{cycle\ 1}$ and $C_{min,cycle\ 1}$ were calculated using the typical PK parameters corrected for individual covariates. No PK data were collected in the capecitabine-plus-lapatinib arm.

Models for circulating biomarkers

In Paper I, indirect response (IDR) models where axitinib inhibits VEGF degradation, and sVEGFR-1, 2, 3 and sKIT production, were investigated. AUC_{daily} was evaluated as a driver of drug effect in linear, maximal effect (E_{max}) and sigmoidal E_{max} models. Linear disease progression functions were tested to describe potential drug-independent biomarker changes.

In sunitinib-treated GIST patients (Papers II and IV), VEGF, sVEGFR-2 and -3, and sKIT time-courses were predicted using published models.[78]

Tumor models

Tumor size: uni- and three-dimensional measurements

Tumor size time-courses (*S(t)*), i.e. SLD (Papers I-II), MTD, actual volume (V_{actual}) and ellipsoidal volume ($V_{ellipsoid}$) (Paper III), were modelled using ordinary differential equations with the general form[51]:

$$\frac{dS}{dt} = net\_growth - drug\_induced\_decay \quad (9)$$

where *net_growth* represents the difference between growth and natural death. Various tumor growth models were explored, including zero-order, exponential, Gompertz, logistic and Weibull functions. To describe the drug-induced decay (*drug_induced_decay*), linear, power and E_{max} functions of drug dose or AUC_{daily} were compared. Moreover, in Paper I individual mod-

el-predicted changes in the different biomarkers (absolute value, or absolute or relative change from baseline) were tested as indirect drivers of axitinib effect on SLD, alone and in combination, using a PPP&D-like approach. To explain tumor regrowth during treatment, as observed in some patients, a time-dependent mono-exponential decay in the drug effect was included. Finally, the addition of an effect compartment to describe an effect delay was also tested when appropriate.

Tumor metabolic activity and density

IDR models with inhibition of the production or stimulation of the loss of response were evaluated to characterize the time-course of individual lesion SUV_{max}, which typically decreases in sunitinib-treated GIST patients (Paper II). Linear, power and E_{max} drug-effect relationships were considered, and an effect compartment with an equilibration half-life ($T_{1/2,ke0}$) fixed to 50 h was included to account for sunitinib accumulation and long elimination half-life ($T_{1/2,el}$~50 h)[79]. Linear and non-linear time-dependent disease progression functions applied on the production rate were tested to explain potential drug-independent increase in SUV_{max} during treatment. Additionally, a mono-exponential time-dependent decay in drug effect was evaluated. Since SUV_{max} and SUV_{mean} were highly correlated (r^2=0.96), the best SUV_{max} model was applied to SUV_{mean} data, re-estimating model parameters.

Similar IDR models were developed to describe tumor density, which tends to decrease during imatinib therapy in GIST patients (Paper III).

Dropout models

Dropout from tumor assessment is likely related to the measured SLD, i.e. it does not occur completely at random. Logistic dropout models were therefore developed to mimic the varying follow-up duration across patients in the SLD simulations (Papers I-II). In axitinib-treated mRCC (Paper I), the time since treatment initiation, the AUC_{daily}, the observed baseline SLD, the model-predicted SLD at the time of evaluation, and the occurrence of 20% increase in SLD from nadir (progressive disease, yes/no) were investigated as predictors. In sunitinib-treated GIST, a previously developed dropout model was implemented[23], and predictors were reevaluated for significance. For both drugs, dosing records were imputed based on the last observed dosing schedule, until the time of last observed SLD assessment in each study.

Correlations between variables

In Papers I-III, each biomarker or tumor model was developed separately. Models were then combined to investigate correlations between model pa-

rameters at the individual level, and the model structure was simplified when appropriate. Specifically, the following correlations were assessed:

- Between VEGF, sVEGFR-1, 2 and 3 (Paper I)
- Between SUV_{max} and SLD, and between SUV_{max} and biomarkers (VEGF, sVEGFR-2 and sKIT) (Paper II)
- Between MTD, V_{actual}, $V_{ellipsoid}$ and tumor density (Paper III)

Models for time-to-event outcomes

Parametric TTE models were developed to characterize OS (Paper I-III) and PFS (Paper III). Table 5 summarizes the investigated baseline hazard functions and predictors, which were evaluated alone and in combination.

Table 5. *Summary of investigated baseline hazard functions and predictors*

Paper	Baseline hazard	Constant predictors	Time-varying predictors[a]
I	Constant Gompertz Weibull Log-normal Log-logistic	• ECOG • Age, sex, body weight • Biomarkers, dBP and SLD at baseline[a]	• $Dose_{daily}$ • AUC_{daily} • X(t), ΔX(t) and X_{rel}(t) of biomarkers[b] and SLD • SLD(t) derivative • X_{rel} at week 2 of dBP, week 4 (dBP, biomarkers[b]), and week 8 (SLD) • Maximum dBP during cycle 1 • dBP > 90 mmHg during cycle 1 (yes/no)
II	Constant Weibull	• ECOG • Number of FDG-PET positive lesions • SLD and ΣSUV_{max} at baseline[a]	• AUC_{daily} • X(t) and X_{rel}(t) of SLD and ΣSUV_{max} • X(t) of SUV_{max} for the most active lesion • SLD(t) • ΣSUV_{max}(t) • SLD(t) • ΣSUV_{mean}(t) • X_{rel} at week 1 or 2 of ΣSUV_{max} and SUV_{max} for the lesion that responded the best
III	Constant Weibull Gompertz Log-normal Log-logistic	• Age, sex • (Log-transformed) MTD, V_{actual}, $V_{ellipsoid}$ and density at baseline[a]	• (Log-transformed) X(t), ΔX(t) and X_{rel}(t) of MTD, V_{actual} and $V_{ellipsoid}$ • X_{rel} up until 1.5, 3, 6, and 12 months of MTD, V_{actual}, $V_{ellipsoid}$ and density [c]

ECOG: Eastern Cooperative Oncology Group functional status; X(t): model-predicted time-course; ΔX(t): model-predicted change from baseline over time; X_{rel}(t): model-predicted relative change from baseline over time; X_{rel}: model-predicted relative change from baseline at a fixed time-point.[a] Model-predicted. [b] Biomarkers: VEGFR, sVEGFR-1, 2 and 3. [c] X_{rel} at 6 and 12 months were not considered in PFS analysis to minimize potential confounding between the predictors and PFS.

In Paper III, size-related predictors could be combined with density metrics, but not with other size-related predictors. Moreover, in the presence of several lesions, MTD, V_{actual} and $V_{ellipsoid}$ were summed across lesions, whereas density was averaged over lesions. For each endpoint, a competing TTE model described censoring, defined as loss of follow-up or non-occurrence of the event. Sunitinib and axitinib doses were extrapolated until death or censoring based on the last recorded dosing schedule, since protocols supported treatment continuation and no alternative treatment was available.

Models for adverse effects

Diastolic blood pressure

IDR models were investigated to characterize the time-course of dBP, which tends to increase during axitinib therapy (Paper I). Axitinib stimulatory effect on dBP was evaluated through linear, power and (sigmoidal) E_{max} functions of $Dose_{daily}$ or AUC_{daily}.

Fatigue and hand-foot syndrome

In Paper IV, the minimal continuous-time Markov model (mCTMM) was developed to describe fatigue and HFS data in sunitinib-treated GIST. The mCTMM is a simplification of the CTMM (Equations 4-7, Figure 4), where the equilibration time between any two neighboring states is governed by a single estimated parameter, the mean equilibration time (*MET*). *MET* is therefore the same for all transfers.

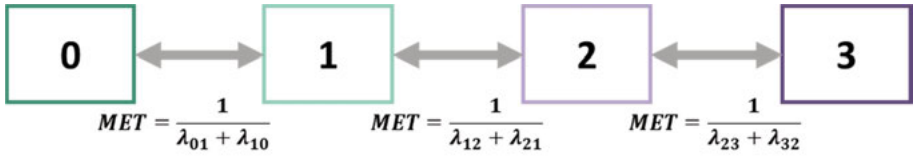

Figure 6. Schematic representation of a mCTMM for a four-state variable.

The transfer rate constants λ (e.g. Equations 10-12 for transitions between 0 and 1 (λ_{01}), and between 1 and 0 (λ_{10})) are expressed as functions of the *MET* and a set of probabilities $P_{ss,k}$, which represent the probability of scoring exactly k at steady-state and arise from a PO model (see Equations in Table 3). The effect of covariates and drug exposure can be implemented on *MET* and/or on parameters defining $P_{ss,k}$.

$$MET = \frac{1}{\lambda_{01} + \lambda_{10}} = \frac{1}{\lambda_{12} + \lambda_{21}} = \frac{1}{\lambda_{23} + \lambda_{32}} \quad (10)$$

$$\lambda_{01} = \frac{1}{MET \cdot \left(1 + \frac{P_{ss,0}}{P_{ss,1}}\right)} \quad (11)$$

$$\lambda_{10} = \lambda_{01} \cdot \frac{P_{ss,0}}{P_{ss,1}} \quad (12)$$

The response probability for the first observation of each individual can be estimated to $P_{ss,k}$ without any assumption regarding the previous score. The predictive performance of the mCTMM was compared to PO models and previously published DTMM[78]. The same predictors identified in DTMM were applied to PO models and mCTMM.

IRT model for FACT-B

The IRT model for FACT-B data was developed in three steps.

In step 1, the base IRT model linked the probability of each score k for each FACT-B item j to the latent well-being ($W_{i,l}$), specific to patient i and FACT-B subscale l. Item characteristic curves (ICCs) were used to illustrate this relationship. Each item was modeled using a graded response model (Equations 13-14):

$$P(Y_{ij} \geq k) = \frac{1}{1 + e^{-a_j(W_{i,l} - b_{j,k})}} \quad (13)$$

$$P(Y_{ij} = k) = P(Y_{ij} \geq k) - P(Y_{ij} \geq k + 1) \quad (14)$$

where a_j and $b_{j,k}$ (with $b_{j,k+1} \geq b_{j,k}$) are fixed effect item parameters representing the discrimination and difficulty parameters for item j. $W_{i,l}$ is modeled as a random variable, with higher values denoting better well-being, and assuming a normal distribution with mean of 0 and variance of 1 at baseline in the T-DM1 arm, and estimated mean and variance at later observations. Correlations between $W_{i,l}$ on the different subscales were estimated. In addition, to test the hypothesis that items in the BCS only shared their specificity to the disease, but each of them likely belongs to another subscale (physical, social/family, emotional, functional), each item was reassigned to one of the other four subscales, based on goodness-of-fit criteria. In this step, data from each patient and visit were treated as independent, to facilitate parameter estimation without assuming any shape of the well-being time-course.

In step 2, the longitudinal well-being model was built, using the EBEs of well-being obtained from step 1 as dependent variables, and reconciling data from each patient. Linear and non-linear functions of time were investigated to characterize well-being time-course. In addition, a covariate analysis evaluated the effect of baseline factors related to demographics, disease status and prior therapies, and T-DM1 exposure on model parameters (Table 6).

Table 6. *Covariates evaluated on longitudinal well-being model parameters*

Demographics	Disease status and prior therapies	TDM-1 exposure[a]
Age	Treatment line	AUC_{cycle1}
Ethnicity	ECOG performance status	$C_{min,cycle1}$
Race	Tumor burden category	
	Site of disease involvement	
	Number of disease sites	
	Hormone receptor status	
	Presence of liver, bone, lung and brain metastases	

[a] Not evaluated on baseline parameters.

Finally, in step 3, the final models from steps 1 and 2 were combined into a longitudinal IRT model, without re-estimating parameters. Model diagnostics at the item and subscale levels, and typical individual predictions, were generated.

Model development and evaluation

Software

All pharmacometric analyses were performed using the NLME software NONMEM version 7.3. FOCE-I method was used for continuous data, and the Laplacian estimation method for categorical and time-to-event data (OS, PFS, dropout).

Data management, graphical and statistical exploratory data analyses, model diagnostics and data post-processing were assisted by R software version 2.15.3 or above, the R package Xpose4, Perl-speaks-NONMEM (PsN) toolkit version 4, and Pirana software version 2.9.0 or above.[80]

Model discrimination and evaluation

Model selection was based on clinical plausibility, goodness-of-fit plots and the OFV (-2·log-likelihood). Upon addition of one parameter (1 degree of freedom), an OFV decrease of at least 6.63 at $p<0.01$, or 3.84 at $p<0.05$, was considered as statistically significant.

The predictive performance of continuous models was assessed using (prediction-corrected) visual predictive checks ((pc)VPCs)[81], where 95% confidence intervals (CI) derived from 500-1000 simulated datasets were compared to the observed data. TTE model performance was evaluated using Kaplan-Meier VPCs, where the 95% CI derived from 200 simulations was compared to the observed TTE data.

For the longitudinal IRT model (Paper V), VPCs based on 200 simulations were generated for the average score and the proportion of each score for each item, as well as the sum of scores for each subscale. Furthermore, to diagnose the relationship between well-being and the response in the IRT model, the estimated cumulative probability curve of each score versus well-being was plotted and compared to the fit of a generalized additive model to the data versus the well-being EBEs, using a cross-validated cubic spline as a smoothing function.[58]

Finally, in Paper IV simulation-based diagnostics comparing the frequency of the maximum achieved score across individuals in the observed and simulated data sets were used. Moreover, the number of transitions in each simulated dataset averaged by the number of individuals was calculated and its distribution was numerically compared to the corresponding observed value. VPCs comparing the observed proportion of each score and the corresponding 95% CIs obtained from 100 simulations were generated.

Additionally, relative standard errors (RSE) of parameter estimates were obtained from the NONMEM Sandwich matrix for continuous models and from the R matrix for TTE and categorical models.

Exponential and additive IIV were evaluated as appropriate. Semi-parametric distributions were tested when indicated graphically.[82] When information on several tumor lesions within an individual were available (Papers II-III), an inter-lesion variability (ILV) term κ_{ij} was implement in addition to the IIV term η_i, as shown in Equation 15 for the parameter θ_i of the ith individual with L lesions:

$$\theta_i = \begin{cases} \theta \cdot \exp(\eta_i + \kappa_{i1}) & \text{if lesion 1} \\ \vdots & \\ \theta \cdot \exp(\eta_i + \kappa_{iL}) & \text{if lesion L} \end{cases} \quad (15)$$

where θ is the typical value of the parameter in the population. η_i and κ_{ij} are assumed to be normally distributed with mean 0 and variance ω^2 and π_j^2 ($\pi_1^2 = \ldots = \pi_L^2$), respectively.

The RUV was evaluated for all continuous data models using additive, proportional or combined error models. Log-transformation of the data and predictions was performed when indicated by graphical diagnostics, to satisfy the assumption that residuals are normally distributed. Moreover, NONMEM level-2 data item was used to group together observations from lesions assessed on the same scan, or biomarkers sampled simultaneously.

Covariate analysis

The covariate analysis in Paper V was performed using the stepwise covariate model building procedure (SCM) implemented in PsN. A significance level of $p<0.01$ was used in the forward selection, and $p<0.001$ in the backward elimination. All continuous covariates were tested on parameters using linear relationships, and piece-wise linear relationships were evaluated upon inclusion of a linear relationship by the SCM. The effects of categorical covariates were implemented as a change in the typical parameter value in relation to the reference category (i.e. most common).

Results

Models for circulating biomarkers

Biomarker time-courses in mRCC patients (Paper I) were well described by IDR models where axitinib inhibits VEGF degradation (through an effect on the first-order degradation rate constant k_{out}, Equation 16) and sVEGFR-1, 2 and 3 production (through an effect on the zero-order production rate constant R_{in}, Equation 17). No axitinib drug effect was identified on sKIT. VEGF levels increased linearly with time (slope α) in a drug-independent manner. Inhibitory E_{max} (VEGF, sVEGFR-1 and 3) and sigmoidal E_{max} functions (sVEGFR-2) characterized the drug effect, assuming a maximum inhibitory effect I_{max} of 1. Estimates of typical AUC_{50} (the AUC_{daily} leading to half I_{max}) were in the range of observed AUC_{daily} (31.95-1861 µg·h/L), with VEGF being the most sensitive to axitinib (lowest AUC_{50}). A common typical AUC_{50} was estimated for sVEGFR-2 and 3. Moreover, since the IIV on AUC_{50} in sVEGFR-1, 2 and 3 models were highly correlated (80-99%) and of similar magnitude, a common IIV term was used. VEGF and sVEGFR-1 turnover time ($MRT=1/k_{out}$) was typically shorter than for sVEGFR-2 and -3.

$$\frac{dA}{dt} = R_{in} \cdot (1 + \alpha \cdot t) - k_{out} \cdot \left(1 - \frac{I_{max} \cdot AUC_{daily}^{\gamma}}{AUC_{50}^{\gamma} + AUC_{daily}^{\gamma}}\right) \cdot A(t) \quad (16)$$

$$\frac{dA}{dt} = R_{in} \cdot \left(1 - \frac{I_{max} \cdot AUC_{daily}^{\gamma}}{AUC_{50}^{\gamma} + AUC_{daily}^{\gamma}}\right) - k_{out} \cdot A(t) \quad (17)$$

$$R_{in} = Base \cdot k_{out} \quad (18)$$

Parameters were estimated with reasonable uncertainty, except MRT in sVEGFR-1 model (Table 7). pcVPCs show a good predictive ability of the joint biomarker model (Figure 7).

Table 7. *Parameter estimates and their uncertainty in the joint biomarker model in axitinib-treated mRCC (Typical value (RSE%)[IIV CV% (RSE%)])*

	VEGF	sVEGFR-1	sVEGFR-2	sVEGFR-3
Base (pg/mL)	65.0 (7.8) [43 (12)]	83.5 (2.9) [17 (12)]	8850 (2.8) [15 (12)]	19500 (6.5) [49 (15)]
MRT (days)	0.722 (25) [-]	0.624 (69)[a] [-]	19.7 (17) [75 (22)]	5.76 (12) [-]
I_{max}	1 FIX [-]	1 FIX [-]	1 FIX [-]	1 FIX [-]
AUC_{50} (µg·h/L)	354 (13) [39 (34)]	1380 (13) [45 (17)] [b]	717 (8.6) [c] 45 (17)] [b]	717 (8.6) [c] [45 (17)] [b]
γ	1 FIX [-]	1 FIX [-]	0.733 (16) [-]	1 FIX [-]
α ($year^{-1}$)	0.650 (28) [87 (22)]	[-]	[-]	[-]
RUV [d]	0.376 (5.9) [-]	0.193 (5.3) [-]	0.162 (14) [-]	0.263 (6.5) [-]
Common RUV [d]	0.0593 (26) [e] [-]	0.0593 (26) [e] [-]	0.0593 (26) [e] [-]	0.0593 (26) [e] [-]

[a] The 95% CI obtained from sampling importance resampling was 0.0444-1.58 days; [b] A common IIV term for AUC_{50} in sVEGFR-1, 2 and 3 models was quantified; [c] Common AUC_{50} parameter for sVEGFR-2 and 3; [d] Expressed as standard deviation on log-scale; [e] Common RUV for all four biomarkers. CV: coefficient of variation

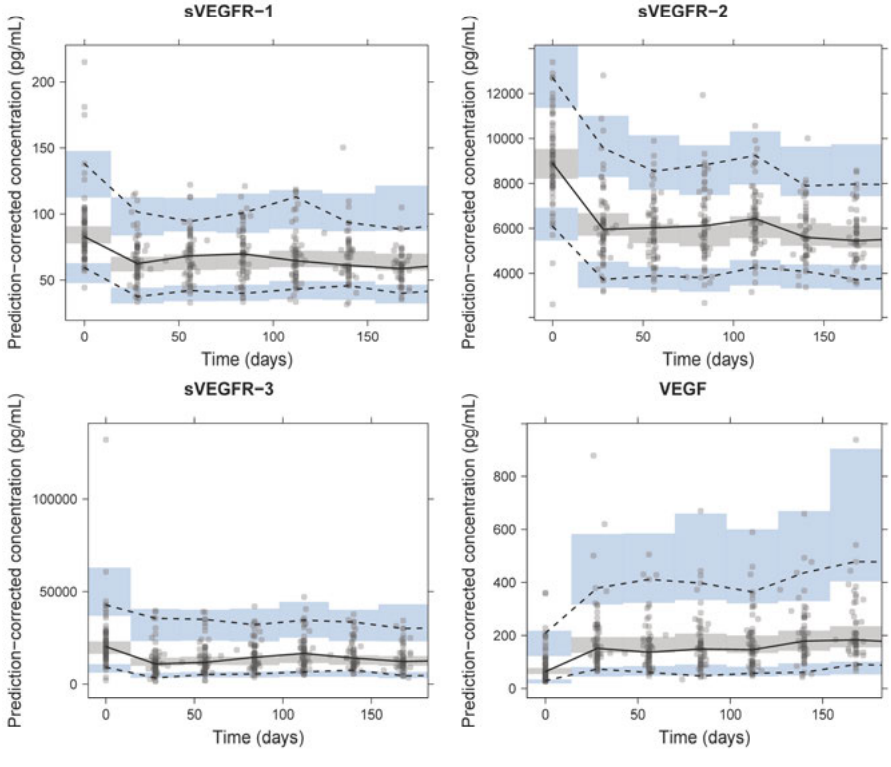

Figure 7. Prediction-corrected VPCs of the joint biomarker models. Median (solid line), 5th, and 95th percentiles (dashed lines) of the observed data (solid circles) are compared to the 95% CI (shaded areas) for the median, 5th, and 95th percentiles of the simulated data (500 simulations).

Tumor models

Tumor size: uni- and three-dimensional measurements

The *net_growth* and *drug_induced_decay* functions and their parameter estimates (Equation 9) are summarized in Table 8 for the developed models. Exponential growth with first-order growth rate constant K_G best characterized SLD data (Papers I-II), while logistic growth models involving an additional carrying capacity parameter S_{max} (i.e. the maximum size above which the lesion cannot grow) fitted MTD, V_{actual} and $V_{ellipsoid}$ lesion data better (Paper III). K_G was associated with substantial IIV (60-170 %) in all three data sets. Since information on K_G was limited in the mRCC data set (Paper I), a published K_G value obtained from a simplified tumor growth inhibition model in mRCC was used as informative prior.[83] Baseline lesion size (S_0) for all metrics in Paper III exhibited a bimodal distribution, which was handled by a mixture model with two subpopulations: subpopulation 1 (estimated probability $P_{pop\ 1}$ of 0.348) had typically larger baseline tumor size than subpopulation 2. Total S_0 was 3.4, 41 and 42 times larger in subpopulation 1 than in subpopulation 2 for MTD, V_{actual} and $V_{ellipsoid}$, respectively, for a patient with two liver metastases. In addition, ILV was identified in S_0, specific to each metrics and subpopulation. Furthermore, S_{max} typical value was specific to each lesion, and metrics.

In Paper I, the model-predicted relative change in sVEGFR-3 from baseline over time, $sVEGFR3_{rel}(t)$, was used to indirectly drive axitinib effect on SLD. In Paper II, sunitinib effect on SLD was driven by AUC_{daily} through an effect compartment (with hypothetical amount C_e, Equation 19).

$$\frac{dC_e}{dt} = k_{e0} \cdot (AUC_{daily} - C_e) \quad (19)$$

All tumor size models included an exponential time-dependent decrease in the drug effect, with rate constant λ, which was typically estimated to be two times slower in imatinib-treated GIST, and five times slower in sunitinib-treated GIST, compared to axitinib-treated mRCC. Parameter uncertainties were within acceptable range (≤47% RSE) except for $S_{max,lesion1}$ in $V_{ellipsoid}$ model (56 %RSE).

In axitinib-treated mRCC (Paper I), the probability of dropping out during one day (μ) was estimated to increase with occurrence of progressive disease (θ_{PD}=1.22), higher SLD at the time of evaluation (θ_{SLD}=0.00282 mm^{-1}), increasing time since start of study (θ_{Time}=0.00371 day^{-1}) and decreasing AUC_{daily} (θ_{AUC}=-0.00529 $L \cdot h^{-1} \cdot \mu g^{-1}$) (Equation 20). μ was scaled to the time since the last observation ($\Delta Time$) to derive the probability of dropping out between two visits (P) (Equation 21).

$$\mu = \frac{1}{1 + e^{-(\theta_0 + \theta_{PD} \cdot PD + \theta_{SLD} \cdot SLD + \theta_{Time} \cdot t + \theta_{AUC} \cdot AUC_{daily})}} \quad (20)$$

$$P = 1 - (1 - \mu)^{\Delta Time} \quad (21)$$

A similar dropout model was used in SLD simulations for sunitinib-treated GIST (Paper II), including the observed SLD at dropout and occurrence of progressive disease as predictors. The VPCs of the final tumor size models (Figure 8) demonstrated a good predictive performance of the model.

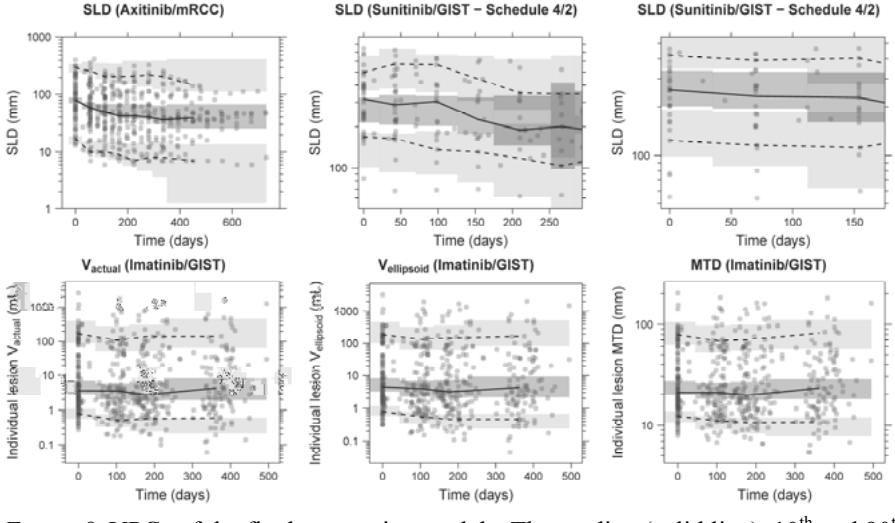

Figure 8. VPCs of the final tumor size models. The median (solid line), 10th and 90th percentiles (or 5th and 95th for SLD in axitinib-treated mRCC, dashed lines) of the observed data are compared to the 95% confidence intervals (shaded areas) for the corresponding median (dark grey) and percentiles (light grey) of the simulated data (500 samples for SLD, 1000 samples for V_{actual}, $V_{ellipsoid}$ and MTD).

Table 8. *Parameter estimates and their uncertainty in the tumor size models (Typical value (RSE%)[IIV CV% (RSE%)])*

	Paper I Axitinib/mRCC	Paper II Sunitinib/GIST	Paper III Imatinib/GIST		
Size metrics	SLD	SLD	MTD	V_{actual}	$V_{ellipsoid}$
net_growth	$K_G \cdot S(t)$	$K_G \cdot S(t)$	$K_G \cdot S(t) \cdot \left(1 - \frac{S(t)}{S_{\max}}\right)$	$K_G \cdot S(t) \cdot \left(1 - \frac{S(t)}{S_{\max}}\right)$	$K_G \cdot S(t) \cdot \left(1 - \frac{S(t)}{S_{\max}}\right)$
drug_induced_decay	$\theta \cdot sVEGFR3_{rel}(t) \cdot e^{-\lambda \cdot t} \cdot S(t)$	$\theta \cdot C_e \cdot e^{-\lambda \cdot t} \cdot S(t)$	$\theta \cdot Dose/400 \cdot e^{-\lambda \cdot t} \cdot S(t)$	$\theta \cdot Dose/400 \cdot e^{-\lambda \cdot t} \cdot S(t)$	$\theta \cdot Dose/400 \cdot e^{-\lambda \cdot t} \cdot S(t)$
S_0 (mm or mL) [a]	- [b]	263 (6.8) [54 (9.6)]	*Subpopulation 1* lesion 1: 76.6 (12) lesion 2: 41.9 (15) [47 (11)] [c] *Subpopulation 2* lesion 1: 20.9 (6.3) lesion 2: 14.2 (6.7) [29 (13)] [c]	*Subpopulation 1* lesion 1: 161 (34) lesion 2: 29.7 (46) [140 (12)] [c] *Subpopulation 2* lesion 1: 3.45 (16) lesion 2: 1.21 (18) [76 (13)] [c]	*Subpopulation 1* lesion 1: 187 (35) lesion 2: 33.4 (46) [140 (12)] [c] *Subpopulation 2* lesion 1: 3.93 (17) lesion 2: 1.27 (19) [78 (12)] [c]
K_G (week^{-1})	0.00361 (1.8) [160 (20)]	0.0105 (24) [60 (38)]	0.00176 (47) [170 (18)]	0.00861 (38) [135 (20)]	0.00882 (43) [140 (27)]
S_{max} (mm or mL) [a]	-	-	lesion 1: 76.6 (12) [-] lesion 2: 41.9 (15) [-]	lesion 1: 76.6 (12) [-] lesion 2: 41.9 (15) [-]	lesion 1: 76.6 (12) [-] lesion 2: 41.9 (15) [-]
θ (week^{-1} or mg^{-1}.L.h^{-2}) [d]	-0.174 (15) [-]	0.0166 (21) [63 (37)]	0.0124 (36) [77 (20)]	0.0547 (21) [56 (29)]	0.0610 (24) [42 (36)]
$P_{pop\ 1}$	-	-	0.348 (18) [-] [e]	0.348 (18) [-] [e]	0.348 (18) [-] [e]
λ (week^{-1})	0.101 (18) [72 (16)]	0.0201 (45) [-]	0.0475 (35) [-]	0.0469 (28) [-]	0.0508 (37) [-]
RUV (%)	10.5 (8.2) [35 (21)]	6.68 (12) [-]	14.0 (6.6) [-]	36.8 (8.0) [-]	43.3 (7.5) [-]

[a] mm for SLD and MTD, mL for V_{actual} and $V_{ellipsoid.}$ [b] B2 method was used, i.e. observed baseline SLD was included as covariate, acknowledging the same RUV as for post-baseline observations. [c] IIV variance shared between lesions, within each subpopulation and model. [d] week^{-1} when *drug_induced_decay* is driven by $sVEGFR3_{rel}(t)$ or Dose/400; mg^{-1}.L.h^{-2} when *drug_induced_decay* is driven by $AUC_{daily.}$ [e] Common parameter to MTD, V_{actual} and $V_{ellipsoid}$ models.

Tumor metabolic activity and density

IDR models with stimulation of the production of response (Equations 22-23, Table 9) adequately characterized lesion-level data of SUV_{max} in sunitinib-treated GIST (Paper II) and of liver metastases density in imatinib-treated GIST (Paper III).

$$\frac{dA}{dt} = R_{in} - k_{out} \cdot Effect \cdot A(t) \quad (22)$$

$$R_{in} = Base \cdot k_{out} \quad (23)$$

Drug effect (*Effect*) on SUV_{max} was driven by sunitinib AUC_{daily} through an effect compartment (with hypothetical amount C_e, Equation 19), while the effect on density was driven by the imatinib daily dose, normalized by the median dose of 400 mg. The turnover time ($MRT=1/k_{out}$) was 1.8 weeks for SUV_{max}, and 10.7 weeks for tumor density. No disease progression component or time-dependent decay in drug effect was included in the models. For the baseline (A_0) and the drug effect parameter (θ) in both models, IIV was estimated to be larger than ILV.

Table 9. *Parameter estimates and their uncertainty in the SUV_{max} and density models*

	Paper II Sunitinib/GIST			Paper III Imatinib/GIST		
Size metrics	SUV_{max}			Density		
Effect	$1 + \theta \cdot C_e$			$1 + \theta \cdot Dose/400$		
	TV (RSE%)	IIV CV% (RSE%)	ILV CV% (RSE%)	TV (RSE%)	IIV CV% (RSE%)	ILV CV% (RSE%)
A_0 (unitless or HU) [a]	7.59 (5.9)	32 (16)	23 (16)	59.0 (5.7)	30 (10)	18 (19)
Box-Cox A_0	-	-	-	-1.06 (47) [b]	-	-
k_{out} ($week^{-1}$)	0.556 (29)	-	-	0.0935 (32)	-	-
θ ($mg^{-1}.L.h^{-1}$ or unitless) [c]	0.946 (15)	74 (26)	57 (20)	0.154 (29)	120 (17)	53 (38)
RUV (%)	41.7 (20)	-	-	20.6 (6.8)	-	

[a] Unitless for SUV_{max}, Hounsfield (HU) for density. [b] The sum of IIV and ILV terms was Box-Cox transformed to account for the skewed random effect distributions. The 95% CI from log-likelihood profiling was (-2.01;-0.397). [c] $mg^{-1}.L.h^{-1}$ for SUV_{max}, unitless for density.

VPCs showed satisfactory predictive performance of both models (Figure 9). Additionally, the developed individual lesion SUV_{max} model could adequately describe SUV_{mean} data from the same lesions (not shown), and the drug effect parameter estimates were similar for both metrics (0.920 $mg^{-1}.L.h^{-1}$, 77% CV IIV and 57% CV ILV in SUV_{mean} model).

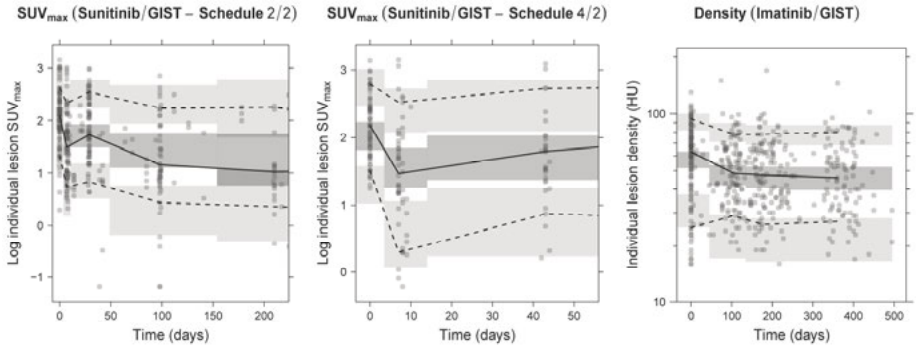

Figure 9. VPCs of the final SUV_{max} and density models. The median (solid line), 10^{th} and 90^{th} percentiles (dashed lines) of the observed data are compared to the 95% CI (shaded areas) for the median (dark grey), 10^{th} and 90^{th} percentiles (light grey) of the simulated data (500 samples for SUV_{max}, 1000 samples for density).

Correlations between variables and typical predictions

In Paper II, the correlation between drug effects on SUV_{max} and SLD at the individual level was estimated to 85%. Conversely, no statistically significant correlation existed between parameters in the SUV_{max} and biomarker models. In Paper III, correlations between density model parameters and MTD, V_{actual} or $V_{ellipsoid}$ model parameters were estimated to <20%.

Figure 10 depicts the typical predictions of the ratio to baseline for the different tumor metrics.

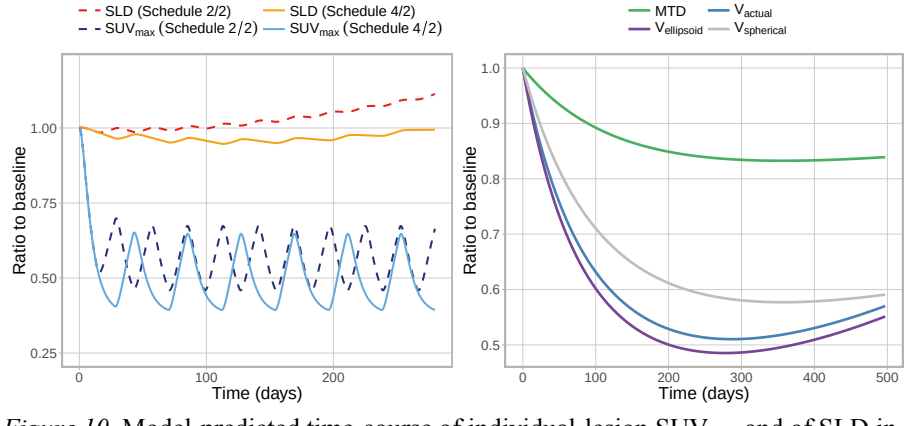

Figure 10. Model-predicted time-course of individual-lesion SUV_{max} and of SLD in a typical GIST patient treated with 50 mg q.d. of sunitinib according to two dosing schedules (left), and of four tumor size metrics in a typical GIST patient treated with 400 mg imatinib (right). Spherical volume was calculated as $V_{spherical}=\pi/6 \cdot MTD^3$.

Models for time-to-event outcomes

Parametric TTE models (Equation 24) were developed to describe OS and PFS data (Papers I-III). Baseline hazard functions ($h_0(t)$) that best characterized the data are summarized in Table 10, together with predictors included in the final models and their corresponding parameter estimates (θ_1, θ_2).

$$h(t) = h_0(t) \cdot e^{\theta_1 \cdot Predictor_1 + \theta_2 \cdot Predictor_2} \quad (24)$$

Table 10. *Functions, parameter estimates and their uncertainty for the final OS and PFS models.*

	Paper I Axitinib/mRCC	Paper II Sunitinib/GIST	Paper III Imatinib/GIST	
Endpoint	OS	OS	OS	PFS
$h_0(t)$	Log-logistic[a] $\frac{(e^{-\beta_0})^{\frac{1}{\gamma}} \cdot t^{\frac{1}{\gamma}-1}}{\gamma \cdot \left(1 + (e^{-\beta_0} \cdot t)^{\frac{1}{\gamma}}\right)}$	Constant[b] λ	Log-normal[a] $\frac{1}{t \cdot \sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{\log(t)-\mu}{\sigma}\right)^2} \Big/ 1 - \Phi\left(\frac{\log(t)-\mu}{\sigma}\right)$	
Predictor 1	SLD(t)	$RCFB_{max,wk\ 1}$	$\log(V_{actual}(t))$	$V_{actual,rel,3m}$
Predictor 2	-	-	-	$\log(V_{actual,0})$
β_0	7.09 (3.2)	-	-	-
γ	0.298 (22)	-	-	-
λ (week^{-1})	-	0.0191 (3.1)	-	-
μ	-	-	1.42 (15)	1.19 (12)
σ	-	-	8.46 (5.0)	7.79 (3.4)
θ_1	0.0115 (17) [mm^{-1}]	5.36 (8.5)	0.190 (36)	0.836 (19)
θ_2	-	-	-	0.239 (29)

[a] time in days. [b] time in hours.

Figure 11 shows the relation between the OS hazard ratio (HR) and V_{actual}.

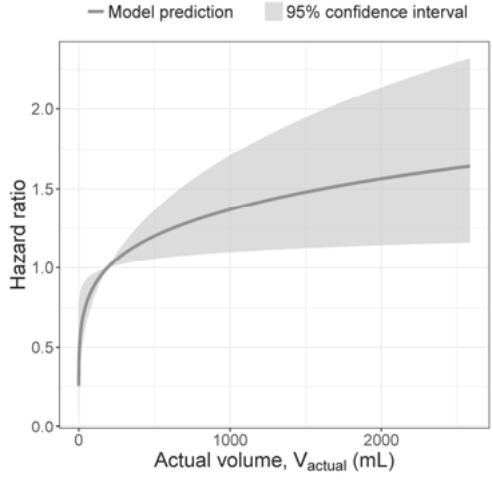

Figure 11. Model-predicted HR from the final OS model in imatinib-treated GIST patients versus the actual volume (V_{actual}). The 95% CI is based on parameter uncertainty.

In axitinib-treated mRCC, SLD time-course (SLD(t)) best predicted OS; other predictors related to SLD, VEGF, sVEGFR-1 and 2, and dBP led to statistically significant OFV drops but relied on single individuals. In sunitinib-treated GIST, the relative change from baseline in SUV_{max} at week 1 for the lesion that responded the best ($RCFB_{max,wk1}$) predicted OS better than any SLD-related metrics. In imatinib-treated GIST, log-transformed V_{actual} time-course was included in the final OS model, while log-transformed V_{actual} baseline combined with the relative change in V_{actual} at three months best described PFS data. None of the density-related variables could predict OS or PFS.

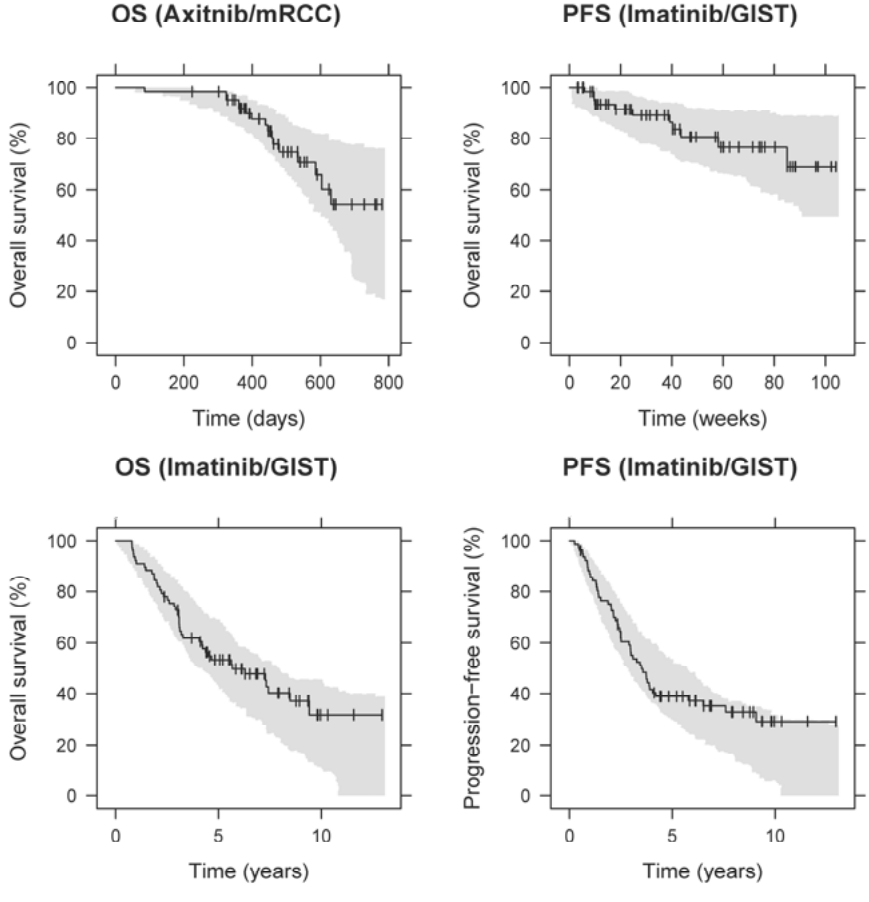

Figure 12. VPCs of time-to-event models. The observed Kaplan-Meier curve (black line) is compared to the 95% confidence interval (shaded area) derived from model simulations (200 samples). Vertical ticks represent censored observations.

Models for adverse effects

Diastolic blood pressure

In mRCC (Paper I), the overall increase in dBP was best described by an IDR model where axitinib stimulated dBP (effect on the production rate constant R_{in}) through an E_{max} model driven by AUC_{daily} (Equation 25). The latter was parameterized in terms of a maximal proportional increase in effect E_{max} (0.197, 14% RSE) and a slope parameter ($S_0=E_{max}/AUC_{50}$ of 0.00127 $L.h^{-1}.\mu g^{-1}$, 50% RSE), leading to a derived AUC_{50} of 155 µg·h/L. No IIV was identified on drug effect parameters.

$$\frac{ddBP}{dt} = R_{in} \cdot \left(1 + \frac{E_{max} \cdot S_0 \cdot AUC_{daily}}{E_{max} + S_0 \cdot AUC_{daily}}\right) - k_{out} \cdot dBP(t) \qquad (25)$$

Baseline dBP typical value was 78.9 mmHg (1.4% RSE), and its IIV (6.7% CV, 12% RSE) was Box-Cox transformed, with an estimated shape parameter of -5.42 (42% RSE). The mean turnover time MRT ($1/k_{out}$) was estimated to 4.92 days (19% RSE). RUV was described by an additive model (σ=6.13, 7.0% RSE). The pcVPCs (Figure 13) showed that the model well predicted dBP data during the first month of treatment.

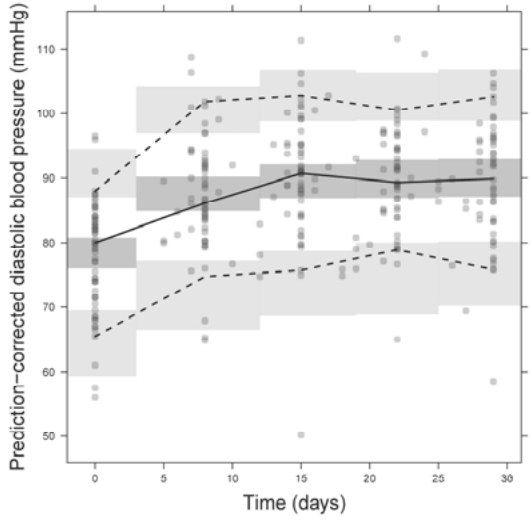

Figure 13. VPCs of the final dBP model. The median (solid line), 5th and 95th percentiles (dashed lines) of the observed data are compared to the 95% CI (shaded areas) for the median (dark grey), 5th and 95th percentiles (light grey) of the simulated data (500 samples).

mCTMM for fatigue and hand-foot syndrome

mCTMM were successfully fitted to Likert pain scores (not shown), fatigue and HFS scores. For both fatigue and HFS, the steady-state cumulative probabilities in the mCTMM were expressed as:

$$P_{ss}(Y_{ij} \geq k) = \frac{1}{1 + e^{-(\alpha_k + \theta_{sVEGFR3} \cdot sVEGFR3_{rel}(t))}} \quad (26)$$

$\theta_{sVEGFR3}$ is the slope of the effect of the model-predicted relative change in sVEGFR-3 from baseline over time ($sVEGFR3_{rel}(t)$) (linear on logit scale). Consistent with previous findings with DTMM, larger decreases in sVEGFR-3 levels were associated with larger steady-state probabilities of developing fatigue or HFS. *MET* estimates reflected a longer equilibration time for fatigue (37.9 days) than for HFS (16.6 days). α_k is as described in Table 3 (Introduction section). The large IIV on α_1 denotes a large variability in score distributions in the absence of drug. The uncertainty was below 30% RSE for all parameters except for α_1 and its IIV in the HFS example (Table 11).

DTMM and mCTMM better fitted the data (lower OFV) and had better simulation properties than PO models, which do not account for Markov properties and substantially overpredict the number of transitions in each individual (Table 12, Figure 14, VPCs not shown). mCTMM could not describe fatigue and HFS data as well as DTMM, but were more parsimonious. Moreover, both DTMM and mCTMM well predicted the average number of transitions per individual (despite a slight overprediction by the mCTMM in the fatigue example) (Table 12). Both models provided similar description of the proportion of patients with maximum achieved scores (Figure 15), with some slight tendency of mCTMM to overpredict the proportion of patients with a maximum HFS score of 0, and to underpredict the proportion of patients with maximum HFS scores of 2.

Table 11. *Parameter estimates and their uncertainty (%RSE) in the mCTMM*

Parameter	Fatigue	Hand-foot syndrome
α_1	-1.85 (7.5)	-17.5 (45)
b_2	-1.08 (7.9)	-1.99 (9.1)
b_3	-1.92 (10)	-2.66 (12)
MET (days)	37.9 (11)	16.6 (7.3)
$\theta_{sVEGFR3}$	-1.87 (15)	-20.5 (8.5)
$\omega_{\alpha 1}$ [a]	0.770 (19)	11.2 (90)
$\omega_{\theta sVEGFR3}$ [b]	1.07 (7.7)	0.475 (28)

α_1: intercept parameter on the logit scale for score k=1; b_k: parameter for score k such that $\alpha_k = \alpha_{k-1} + b_k$; ω_P: standard deviation of the IIV associated with parameter P. [a] Incorporated as additive. [b] Incorporated as exponential.

Table 12. *Summary of the data and ordered categorical model characteristics*

	Observed	Proportional odds	mCTMM	DTMM
Fatigue				
OFV	-	55900	8382/7999 [a]	6766
No. of parameters (fixed/random effects)	-	5 (4/1)	7 (5/2)	20 (16/4)
Average no. of transitions per ID [b]	1.90	34.7	3.23	1.91
Hand-foot syndrome				
OFV	-	24286	4784/4471 [a]	3802
No. of parameters (fixed/random effects)	-	5 (4/1)	7 (5/2)	19 (16/3)
Average no. of transitions per ID [b]	1.35	11.9	1.13	1.13

ID: individual. [a] The first OFV value corresponds to a mCTMM where the probability of the first score in each ID is estimated to $P_{ss,k}$, and can be compared to the OFV of the PO model. The second OFV value corresponds to a mCTMM, where the unobserved score prior to the first observation is assumed to be the same as the score at the first observation, and can be compared to the OFV of the DTMM. [b] For all models, calculated based on 100 simulated data sets. Table adapted from Schindler and Karlsson (2017).

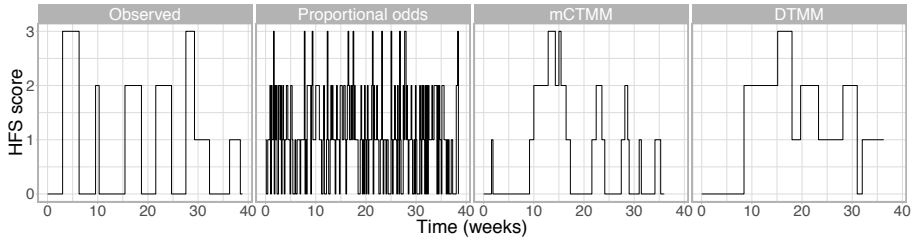

Figure 14. Examples of individual observed and simulated HFS score profiles.

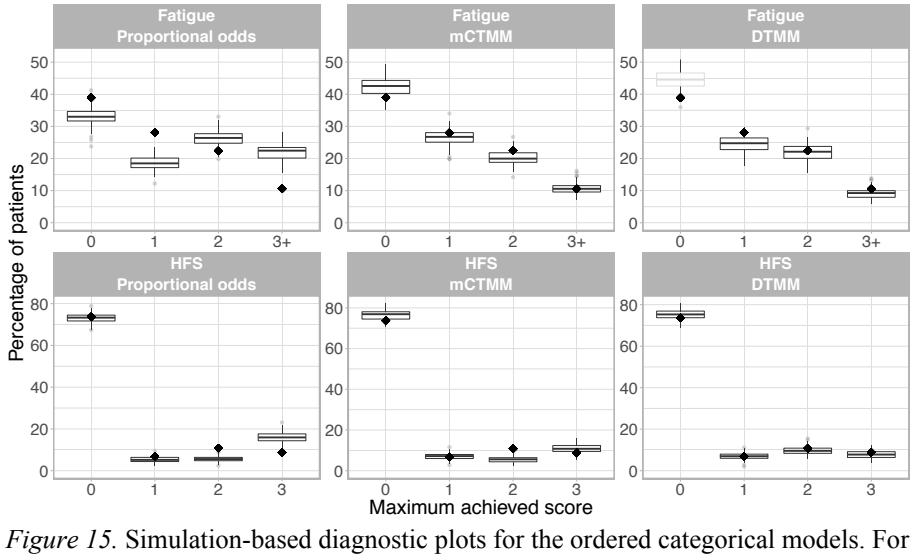

Figure 15. Simulation-based diagnostic plots for the ordered categorical models. For each score, the percentage of patients achieving this score as a maximum in the observed data set (diamonds) and in the simulated data sets (box plots based on 100 samples) are compared.

IRT model

In the base IRT model (step 1), four latent well-being variables ($W_{physical}$, W_{social}, $W_{emotional}$ and $W_{functional}$) were used to describe FACT-B item data. Reassignment of each BCS item to one of the four other subscales resulted in a more parsimonious model and improved model fit (OFV lower by 1138 points) compared to a model with five latent variables (including a BCS-specific latent well-being). The illustrate the relationship between the outcome of each item and the well-being ICCs (Figure 16, with reassigned BCS items marked with '*'). Items with flatter ICCs are less sensitive to changes in well-being. Interestingly, for some items (e.g. "Emotional support from family") the ICCs are located towards low well-being values; these items do not enable to differentiate among individuals with better well-being.

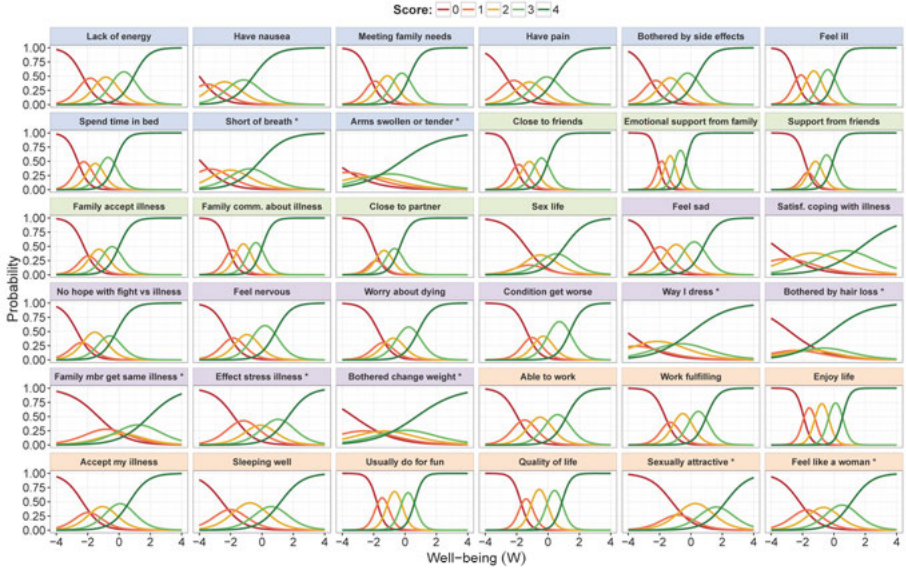

Figure 16. Item characteristic curves (ICCs) obtained from the base IRT model for FACT-B. Subscales are differentiated by color: blue for physical, green for social/family, purple for emotional and orange for functional. Higher scores indicate better outcome). * BCS item.

In the longitudinal well-being model (step 2), well-being time-course ($W_l(t)$) for each subscale l was best described by an asymptotic function of time:

$$W_l(t) = W_{0,l} + W_{ss,l} \cdot \left(1 - e^{-\frac{ln(2)}{T_{1/2,l}} \cdot t}\right) \quad (27)$$

where $W_{0,l}$ is the baseline well-being, $W_{ss,l}$ the steady-state well-being and $T_{1/2,l}$ the progression half-life. $W_{ss,l}$ estimate denoted whether patients improved (positive), remained stable (zero) or worsened (negative) over time.

Since individual $W_{ss,l}$ were highly correlated between subscales (62-96%), an IIV term common to all subscales was estimated together with an inter-subscale variability term. This allows patients to progress differently on each subscale. In both arms, emotional well-being typically improved over time, whereas functional well-being typically stayed stable. Typical $W_{ss,physical}$ was not significantly different from zero (i.e. stayed stable) in the T-DM1 arm, but was estimated to -0.251 in the capecitabine-plus-lapatinib arm (i.e. worsens by 0.251 standard deviation from baseline). $W_{ss,social/family}$ typically worsened in both arms, and progression was more pronounced in T-DM1 arm compared to capecitabine-plus-lapatinib arm. $T_{1/2}$ estimates were in the range of 1.5-2.3 times the treatment cycle length for all subscales except social/family well-being (5.6 times). Correlations between individual W_0 on different subscales (44-81%) and between the IIV in W_0 and the IIV in W_{ss} (-23 to -33%) were estimated.

In the covariate analysis, Asian patients were identified to have typically worse baseline social/family and functional well-being than non-Asian patients. Similarly, patients with ECOG of 1 had typically worse physical and functional well-being than those with ECOG of 0. None of the investigated relations between T-DM1 exposure and W_{ss} was statistically significant.

All parameters were estimated with reasonable uncertainty (Table 13).

Table 13. *Parameter estimates and their uncertainty obtained from the longitudinal well-being model for FACT-B*

Parameter	Typical value (%RSE)[a]	IIV[a,b] (%RSE)
$W_{0,physical}$ (unitless)	0 fixed	0.76 (2.5)
$W_{0,social/family}$ (unitless)	-0.0901 (23)	0.69 (2.5)
$W_{0,emotional}$ (unitless)	0 fixed	0.81 (2.6)
$W_{0,functional}$ (unitless)	0 fixed	0.79 (2.5)
$W_{ss,physical}$ (unitless)	TDM1: 0 fixed C+L: -0.251 (11)	IIV[c]: 0.59 (4.9) ISV[c]: 0.16 (14)
$W_{ss,social/family}$ (unitless)	TDM1: -0.244 (15) C+L: -0.137 (32)	
$W_{ss,emotional}$ (unitless)	0.295 (6.8)	
$W_{ss,functional}$ (unitless)	0 fixed	
$T_{1/2,physical}$ (days)	30.7 (9.0)	121 (4.2)[d]
$T_{1/2,social/family}$ (days)	117 (18)	
$T_{1/2,emotional}$ (days)	35.1 (12)	
$T_{1/2,functional}$ (days)	48.9 (7.4)	
Asian on $W_{0,social/family}$	-0.441 (12)	-
Asian on $W_{0,functional}$	-0.181 (21)	-
ECOG 1 on $W_{0,physical}$ and $W_{0,functional}$	-0.268 (11)	-

[a] Common to T-DM1 and capecitabine-plus-lapatinib (C+L) arms, unless stated otherwise. [b] Reported as standard deviation for W_{ss} and CV% for $T_{1/2}$. [c] common to all W_{ss}. [d] Common to all $T_{1/2}$. ISV: inter-subscale variability.

The base IRT and longitudinal well-being models were combined into the longitudinal IRT model (step 3). The VPCs demonstrate that the combined model was able to satisfactorily predict item-level FACT-B data (Figure 17, focused on T-DM1 arm), despite a slight overprediction of the average baseline score for some physical items at baseline, and underprediction of "Having pain" item score in the capecitabine-plus-lapatinib arm. In addition, subscale-level VPCs (not shown) indicate that the model can predict the median trend and the variability in both treatment arms, although at some early time points the upper percentile tends to be overpredicted in the physical and breast cancer subscale.

The differences in typical steady-state probabilities and expected scores between the two arms are illustrated in Figure 18.

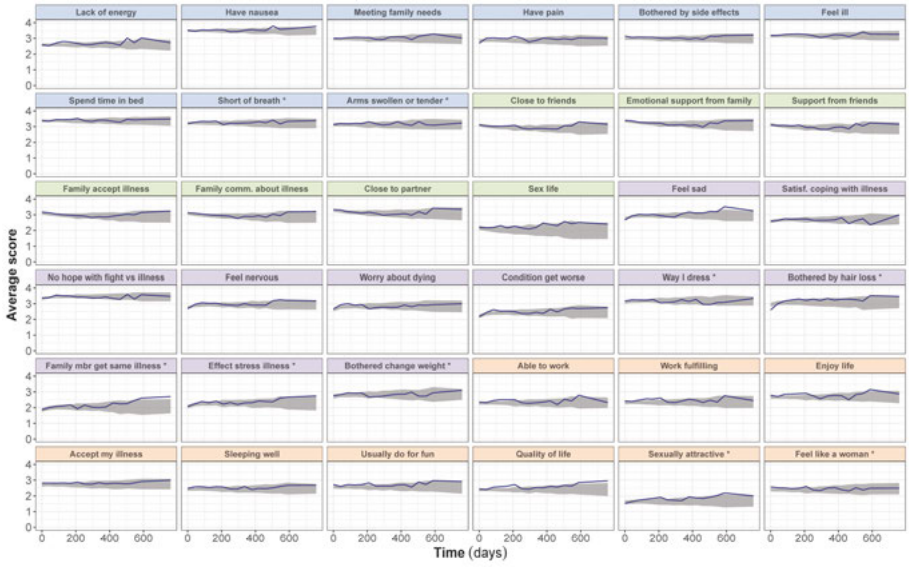

Figure 17. Item-level VPCs of the longitudinal IRT model (T-DM1 arm).

Figure 18. Typical steady-state probabilities (circles) and expected scores ($\sum_{k=0}^{4} P(Y = k) \cdot k$, cross symbols) for each FACT-B item, as predicted by the longitudinal IRT model.

Discussion

Integrated modeling frameworks in TKI-treated cancers

To identify predictors of clinical outcome in cancer patients treated with TKIs, three modeling frameworks were established linking drug exposure, soluble and/or imaging biomarkers, dBP (when applicable) and OS. These may be used to further explore alternative dosing regimen and potential dose individualization.

In cytokine refractory mRCC patients, axitinib inhibited VEGF degradation and sVEGFR-1, 2 and 3 production; a more potent effect was identified on sVEGFR-2 and 3 (lower AUC_{50}) than on sVEGFR-1, and all were associated with significant IIV (45% CV). Evaluation of genotype-parameter relationships in axitinib-treated mRCC[84] could explain (part of) the IIV, and potential differences in clinical outcome, as reported in sunitinib-treated mRCC and metastatic colorectal cancer patients where single nucleotide polymorphisms predicted lower sVEGFR-2 intrinsic activity.[85] In addition, larger sVEGFR-3 decreases were associated with more pronounced tumor shrinkage. Similar biomarker-tumor relationships were reported in sunitinib-treated GIST[23] and hepatocellular carcinoma[86]. Hence, our results reinforce the utility of sVEGFR dynamics in predicting tumor response during TKI treatment.

Early dBP elevation after the start of therapy has been proposed as an easy-to measure marker of effective VEGF inhibition and as a potential predictor of OS.[75] In our analysis, dBP was not predictive of OS, whereas in a statistical analysis of long-term OS data in mRCC, larger dBP during the first 8 weeks of axitinib treatment was associated with longer OS[87]. The discrepancies may be due to shorter follow-up periods in our study (max. 112 versus 285 weeks).

OS in axitinib-treated mRCC was best predicted by SLD time-course, with an estimated hazard ratio of 1.12 for every 10 mm increase in SLD. This differs from findings in sunitinib-treated GIST where sVEGFR-3 dynamics and baseline SLD were best associated with OS[23], whereas in sunitinib-treated mRCC sVEGFR-2 and 3 dynamics (absolute time-courses and relative change from baseline over time) were not predictive of PFS[85]. The reasons for such differences are yet to be elucidated, but may partly be due to differences in the tumor dynamics: stable disease was more frequent

in sunitinib-treated GIST[88] while most patients in the present axitinib-treated mRCC population had a partial response[73].

Tumor metabolic activity (SUV_{max}) time-course in GIST patients was described by a model where sunitinib stimulated the loss of SUV_{max} response. A deprivation of tumor accessibility to glucose and a potential reduction of glucose transporter 1 expression have been suggested as potential mechanisms behind these changes. Alternative models to the one presented here may be evaluated as more mechanistic knowledge becomes available. Variability in SUV_{max} response was estimated to be large between patients and between lesions within a patient. To explain the latter, the effect of potential explanatory factors such as tumor localization or mutation status may be evaluated in the future, if such data become available. Interestingly, the drug effect on SUV_{max} correlated with that on SLD, but not with those on circulating biomarkers, most likely because different pathways are involved. Our model also demonstrated that SUV_{max} dynamics is sensitive to sunitinib intermittent schedules, with a rapid typical decrease during treatment and a recovery toward baseline during off-treatment periods (Figure 10). This contrasts with SLD changes which are much slower and of smaller magnitude, and could explain why SUV_{max} but not SLD was predictive of OS. A HR of 0.59 was predicted by our model for every 10% decrease in SUV_{max} for the lesion that responds best after one week of treatment. Similar findings were reported from a modeling analysis in non-small cell lung cancer treated with the epidermal growth factor receptor inhibitor erlotinib (estimated HR of 0.84 for 10% decrease in peak SUV).[89] In contrast, FDG-PET response at week 4 did not correlate with clinical outcome in sunitinib-treated mRCC[90]. While differences between tumor types might explain such differences, FDG-PET assessment times 2-5 days after sunitinib last dose (i.e. when response may be returning to baseline) may have prevented the identification of a relationship. Similarly, Hansson et al. characterized the sensitivity in biomarker dynamics (VEGF, sVEGFR-2 and 3, and sKIT) to dosing schedule in sunitinib-treated GIST[23]. It is therefore crucial to acknowledge schedule dependency in the design of clinical trials and their analysis. Altogether, our results indicate that early FDG-PET changes can be used as a marker of GIST response to sunitinib and may predict clinical outcome.

In GIST patients, the longitudinal responses of liver metastases to first-line imatinib treatment were modelled, including three size metrics (MTD, V_{actual} and $V_{ellipsoid}$) and tumor density. All size metrics were described by the same structural model, where a logistic growth model evidenced that the tumor growth rate decreases as the tumor size approaches the carrying capacity (maximum tumor size). The latter was estimated to a value similar to the liver volume for a typical patient with two lesions.[91, 92] Typical doubling times of 7.4 years for MTD and 1.5 years for V_{actual} and $V_{ellipsoid}$ (similar to

1.0 year reported for untreated GIST before surgical resection)[93] were derived from K_G. K_G reflects the exponential tumor growth rate when tumors are small relative to the carrying capacity. Volume time-courses were predicted to be typically more sensitive to detect imatinib-induced size changes than MTD (Figure 10). Additionally, $V_{ellipsoid}$ but not spherical volume ($V_{spherical}$, based on MTD) adequately approximated V_{actual}, thereby confirming previous findings.[35] For tumors with easily identifiable borders, $V_{ellipsoid}$ therefore offers a convenient alternative to V_{actual}, which may be impractical due to restricted software availability, staff training and time resources. Tumor density changes (typical decrease) was only weakly correlated to tumor size metrics, indicating that structural and size changes may occur independently and provide distinct information on tumor response.

Overall, volume metrics better predicted OS and PFS data than MTD, although differences in goodness-of-fit criteria were marginal. Tumor density had no predictive value (alone or combined with size metrics). Log-transformed V_{actual} time-course best predicted long-term OS (data available for up to 13 years), indicating a non-linear relationship between the HR and V_{actual} (Figure 11). V_{actual} relative change from baseline up to 3 months after therapy initiation together with baseline log-transformed V_{actual} best predicted PFS.

These results encourage the incorporation of volume measurements in prospective clinical trials with larger study populations, to evaluate their ability to predict clinical outcomes versus RECIST in patients with GIST and other tumors known to exhibit non-uniform changes. Moreover, the collection of PK data in future trials is highly encouraged, since PK could likely explain part of the IIV identified in tumor response. Moreover, collection of FDG-PET data[94] would be of value to compare the predictive ability of tumor metabolic activity and volume measurements for OS. In addition, only one or two liver metastases were evaluated in imatinib-treated GIST patients, and the volume-OS relationship should be confirmed based on all lesions.

The developed modeling frameworks in mRCC and GIST patients increase our understanding of the quantitative relationships between drug exposure, circulating or imaging biomarkers, tumor response and clinical outcome. Noteworthy, the use of simultaneous, PPP&D and IPPSE approaches was not always possible due to due model instability which prevented likelihood ratio tests to be performed.

Interestingly, off-treatment periods in sunitinib-treated patients were associated with a flare in symptoms[95], and clinical trials are still ongoing to assess the efficacy and safety of alternative sunitinib dosing schedule (e.g. individualized schedule).[96, 97] The models presented here and those integrating biomarkers, tumor size, adverse effects and OS[22, 23] may be used for simulations to evaluate dosing schedules that would best balance benefit and risks, and guide early clinical decisions to improve clinical benefit. Im-

portantly, dropout models should be included to allow for realistic SLD simulations at a population level.

As demonstrated for FDG-PET, tumor size and density metrics, the developed methodology to account for ILV makes it possible to evaluate the response of individual lesions on OS and is likely to provide more accurate predictions of tumor dynamics. Additionally, analyzing data from all lesions (target and non-target) would best reflect overall tumor burden.

Finally, information on tumor growth rate is often lacking as patients are generally withdrawn from treatment upon progression. In axitinib-treated mRCC, a literature prior[83] was used to facilitate parameter estimation. Since prior data are often not available, collection of pre-treatment data, which has proved feasible in phase I studies at minor additional costs[98], should be encouraged for a better prediction of tumor size dynamics.

Minimal CTMM for ordered categorical adverse effect data

The mCTMM is a simplification of the CTMM that accounts for Markov elements with a reduced number of estimated parameters. The *MET* informs about the equilibration time of the system; the transit time from the lowest to the highest of n categories can be calculated as $MET \cdot n-1$. The effect of time and covariates (e.g. patient characteristics, drug exposure) can be investigated on the steady-state probabilities, or on the *MET* to explain that some transition may become more/less frequent over time (e.g. in Likert example).

Although DTMM provided better fit to the fatigue and HFS data in sunitinib-treated GIST, mCTMM well predicted the average number of transitions per individual and the maximum achieved scores, while requiring much fewer estimated parameters than DTMM. Unlike in DTMM, the steady-state probability at any time (including the first observation) can be derived for each score in the mCTMM, regardless of the previous score. The mCTMM did not appropriately predict the proportion of HFS scores of 1 and 3 (not shown), whereas DTMM provided a good description of these scores. This discrepancy may partly be explained by the single $sVEGFR\text{-}3_{rel}$ effect applied on the steady state probabilities in mCTMM, whereas in DTMM $sVEGFR\text{-}3_{rel}$ effects were conditional on the previous observed score,[22] conferring more flexibility to DTMM but being less easy to interpret.

The NCI-CTCAE defined grades for more than 700 adverse effects, some of which are often closely monitored during oncology clinical trials. Serial correlations between measurements may occur for adverse effects that are frequently assessed during the course of treatment. The use of mCTMM is encouraged when Markov features are either evident from the data or suspected. If the estimated *MET* is in the range or larger than the observation

frequency, Markov elements should not be neglected, and PO models should not be used. Indeed, PO models may largely inflate the number of transitions when used for simulations and may lead to misspecification of the exposure-response relationships. mCTMM is more appropriate than DTMM when the time intervals between observations are not uniform within a patient (e.g. due to missing observations or study protocol). mCTMM may not perform well in cases where transitions of different magnitudes are observed within the same patient. By estimating different *MET* between the different states, mCTMM can be extended to mimic the CTMM. Finally, possible extensions to the mCTMM include the addition of a dropout/death state from which patient cannot transit back.

IRT pharmacometric analysis of PROs

An IRT pharmacometric approach describing PRO data collected in a cancer clinical study was developed for the first time. The developed model appropriately described the both item-level and subscale-level longitudinal FACT-B data in T-DM1 and capecitabine-plus-lapatinib patients. A three-step approach is proposed to facilitate model building and parameter estimation. The use of four correlated latent well-being variables described the multi-dimensional nature of FACT-B. BCS items are heterogeneous,[48, 99, 100, 101] and each of them was consequently reassigned to one of the other subscales. The ICCs suggested that FACT-B items contained varying information on the latent well-being; in such cases IRT is considered advantageous over sum of score approaches.[102] Furthermore, in IRT models covariate effects applied on the latent well-being variable affect all items in a subscale to an extent that depends on their relationship to the latent variable. This approach is more parsimonious than analyzing each item separately.

Although IRT has been shown to have superior power to detect drug effect,[58, 62] none of the evaluated relationships between T-DM1 exposure and the latent well-being variables was statistically significant. The limited range of exposure and the use of cycle 1 metrics may have limited the ability to detect a relationship. Contrasting results on exposure-response relationships for T-DM1 safety and efficacy outcomes have been published. Li *et al.* reported no meaningful exposure-response relationship for any categorical safety endpoints (grade ≥3 treatment-related adverse event, thrombocytopenia and hepatotoxicity) [103], while a modeling analysis of platelet count and liver enzyme time-courses in T-DM1-treated mBC patients evidenced exposure-response trends[104]. Thrombocytopenia and hepatotoxicity are generally asymptomatic, and they may therefore not impact patients' well-being. Additionally, longer median OS and PFS were reported for patients with higher T-DM1 exposure ($C_{min,cycle1}$)[103, 105], but these relations were not consistent across exposure metrics and were likely confounded by baseline risk

factors[103]. Our results suggests that patients with lower T-DM1 exposure do not exhibit worse well-being than others and that well-being is not expected to worsen if doses were increased in this patient subpopulation.

A less favorable baseline well-being was found in Asian patients (physical and social/family subscales) and in patients with ECOG of 1 (physical and functional). As all FACT-B versions underwent rigorous linguistic validation, differences in socio-economic status are more likely to explain the racial differences.[48, 99, 100, 106, 107, 108] Moreover, correlations between ECOG and well-being and the lack of correlation between well-being and disease-related factors are in accordance with previously-published results.[48, 109, 110]

Physical well-being typically stayed stable in the T-DM1 arm while it worsened in the active control arm. A more favorable outcome in the "bothered by side effects" item was shown for T-DM1 patients in a previous statistical analysis.[74] Noteworthy, the effect of an open-label design on patient reporting is not clear.[111, 112] The large IIV in progression on all subscales could not be explained by any of the baseline covariates.

Possible extensions of this IRT framework include the use of methods to model non-ignorable missingness (e.g. informative dropout [60, 113], missing answers[114]) if indicated by the data. The IRT may also be combined with a mCTMM if serial correlations are suspected. In addition, similar IRT models may be applied to other PRO data, and used to elucidate which of the items are the most informative.[58, 59]

Conclusions

This thesis demonstrates how data from PK, efficacy, safety and time-to-event clinical outcomes in oncology can be integrated in a unified structure to investigate how they correlate with each other. All continuous variables such as biomarker levels, tumor size and dBP, were analyzed without categorization, and their full time-courses were taken into account, thereby leveraging the data collected during clinical trials or clinical practice. Clinical decisions are likely best informed by a combination of several markers, and the developed modeling frameworks may serve as a platform for performing simulations to investigate new dosing strategies that would best balance benefits and risks and therefore improve clinical benefit of anti-cancer therapies. Such models also have application in drug development to accelerate the development of new compounds with similar mechanism of action by serving as proof-of-concept in early phases of development, predicting which cancer types may best respond to the drug and guiding the selection of doses and dosing schedules to be used in later phases. This thesis also introduces new modeling methodologies for efficient and thorough analyses of categorical data with Markov properties and of PRO questionnaire data.

Specific findings are as follow:

- Changes in sVEGFR-3 were linked to SLD dynamics, which could in turn predict OS better than biomarker- or hypertension-related predictors in advanced RCC patients treated with axitinib.

- The schedule-dependent metabolic response of GIST treated with sunitnib was characterized and substantial inter-lesion variability was quantified. Larger decreases in tumor metabolism for the lesion that best responds after one week of treatment were predictive of longer OS.

- Tumor volume better detected size changes and were slightly better at predicting OS than uni-dimensional measurements in imatinib-treated GIST. Density measurements were not predictive of OS or PFS.

- The mCTTM was developed to facilitate the analysis of ordered categorical adverse effect data with Markovian features, when existing ap-

proaches are not appropriate (non-uniform assessment intervals) or not easily implemented (variables with large number of categories).

- The developed IRT pharmacometric framework provides a thorough description of FACT-B longitudinal data in breast cancer patients, acknowledging the multi-dimensional nature of the questionnaire and allowing covariate and exposure effects to be evaluated on the latent variables.

Future perspectives include:

- Integrating all models of efficacy and safety into a unified framework and performing simulations balancing benefit and risk of various dosing regimen

- Identifying covariates explaining inter-lesion variability in response (e.g. location)

- Further evaluating the value of FDG-PET and volumetric changes in larger populations

- Accounting for the non-target lesions and appearance of new lesions in the quantification of overall tumor burden

- Further investigating when IRT models are the most advantageous over summary-score based approaches

- Combining IRT and mCTMM for the description of composite scores with serial correlations.

Acknowledgements

The work presented in this thesis was carried out at the *Department of Pharmaceutical Biosciences*, Uppsala University, Uppsala, Sweden.

I am grateful for financial support from DDMoRe initiative, the Swedish Cancer Society and Genentech Inc., South San Francisco, California, USA. I would also like to thank Smålands Nation and Apotekarsocieten for the travelling grants to attend and present my work at various international conferences and workshops.

I would like thank all the people who have directly or indirectly contributed to this thesis during the past six years:

My main supervisor, *Professor Lena Friberg*, for your invaluable guidance, support and kindness since day one Thank you for thoroughly reviewing every part of this thesis and contributing with your expertise to all steps of the projects, and for always keeping your door open. It has been truly uplifting and inspiring to work with you! Thank you also for giving me the freedom to broaden my experience across the road and across the ocean.

My co-supervisor, *Professor Mats Karlsson*, for your constant enthusiasm for innovative projects, for understanding my "I don't understand" and taking time to explain things again (and again ...), and for taking me out of my "comfort zone". You have enormously contributed to help me grow as a scientist.

My coauthor *Michael Amantea* from Pfizer for your valuable input on sunitinib and axitinib projects

My coauthors *Gaia Schiavon*, *Ron Mathijssen* and *Alessandro Ruggiero* for your open-mindedness and clinical expertise on imatinib project. Special thanks to *Sreenath Krishnan*: your ability to work efficiently and independtly and be very pro-active have impressed me!

My coauthors *Bert Lum*, *Bei Wang*, *Angelica Quartino*, *Chunze Li*, *Sandhya Girish* and *Jin Jin* for bringing a great project on patient-reported outcomes. Thank you *Jin* for your enthusiasm in this project, *Bert* for always being so available and putting the project into perspective and *Bei* for wel-

coming me in San Francisco. Thanks to *Jonathan Squire* for the data preparation and to *Sebastian Ueckert* for your help with IRT.

I am thankful to my half-time opponents *Marylore Chenel*, *Emilie Hénin* and *Elodie Plan* for your valuable feedback and the lively discussion.

I would also like to thank *Marie Sandström* for the great opportunity to work at Pharmetheus, and *Arnaud*, *Niclas* and *Elodie* for this enriching experience.

I would also like to thank *Professor Margareta Hammarlund-Undenaes* for providing a great work environment, *Professor Ulrika Simonsson* for your supervision when we worked together, *Andy Hooker* for your scientific input and help with Xpose4, *Kajsa Harling* and *Rikard Nordgren* for developing PsN so efficiently and always being ready to help, *Emma Hansson* for helping me take my first steps in pharmacometrics, *Kristin* for sharing your expertise in TTE, *Siv* for asking "naïve" questions before I do and for the good spirit you bring to the department, *Mia* for encouraging outstanding science to be shared on Fridays, *Bettan* for sharing your knowledge. I am also thankful to *Magnus*, *Tobias* and *Jerker* for IT support, *Jörgen* for giving me the opportunity to teach, *Marina*, *Elisabeth*, *Ulrica* and *Karin* for administrative help, and *Agneta* for making our lives at the department easier. Thanks also to *Johan Wallin* for attending my seminars and for interesting discussions.

To the ones who were here when I started and who have set the bar very high: *Akash* for your pragmatism, *Elin* for the French and not-so-Swedish nights, *Elodie* for the IRT meetings and nice discussions around a coffee, *Jocke* for being able to teach complex things in a simple manner, *Martin* for the Michelin-quality dinners, *Ron* for always being ready to share your knowledge and for helping me get started with Pirana, *Sebastian* for interesting conversations may they be work-related or not, *Vijay* for pushing me to read, read, read and for the nice trips, *Waqas* for accepting me in the lunch club and the good times outside work, *Åsa* for your contagious laughter.

The "oncology group" members: *Brendan* for all your crazy ideas, *Alexandre*, *Camille* and *Philippe* for bringing a French touch to those meetings, and *Ida* for your great analytical skills.

My past and present office mates: *Sofia*, *Oskar A.* and *Oskar C.* for sharing the "old fika room", *Irena* for the inspiring conversations, *Thomas* for the great working mood, *Gunnar* for the mysterious count down on the white board, *Sreenath* for sharing the office during the most stressful period, *Chunli* for being such a relaxed roommate, *Anders K.* for whistling loud enough from the office next door so that I can hear.

Julia for all the fun moments in India, *Henrik* for countless puns, *Anders T.* and *Moustafa* for your company on the Dingo tour, *Nebojsa* for making the parties last, *Yasunori* for hosting D&D nights, *Ari* for bringing your good

mood to the fika, *Paolo* for your great mood, *Trine* for visiting us here, *Elke* for bringing so much life to the group, *João* for being am easy-going travel buddy in India, *Steve* for your humor and great pictures, *Rasmus* for the warm welcome in Copenhagen, *Stein* for amazing song writing, *Gustaf* for your acting skills, *Marina* for the long discussions about opponents, *Lénaïg* and *Estelle* for bringing some new French influence to the group, *Eva* for brainstorming about mCTMM and IRT, *Siti* for keeping me company in Greece, *Margreke* for the nice weekend activities with the families, *Sebastian W.*, *Winn*, *Eirini*, *Ami*, *Erik*, *Chenyan*, *Shijun*, *Robin*, *Erik*, *Ronald*, *Yang*, *Jessica*, *Yevgen*, *Viktor*, *Annika*, *Brigitte*, and all others who contributed to the amazing atmosphere in the department!

Ana for the late nights at BMC and for being a trustworthy friend, *Anna* for the good tiiimes and Sunday fika in Sysslomansgatan, *Anne-Gaëlle* for convincing me to start this crazy adventure and for being there since our Göteborgs time, *Aziz* for being such a positive and creative person, *Chayan* and *Lani* for making me travel to the most beautiful places in the world, *David* for your friendship and delicious BBQs, *Gopi* for the great dinners with not-very-spicy food, *Irena* for being much more than an officemate, *Salim* for both fun and serious conversations.

Last but not least, I wish to thank my family and friends from France. *Christine*, *Arnaud* et *Guillaume* pour m'avoir accueillie dans votre famille. *Ma famille au grand complet*, parce que vous avez grandement contribué à ce que je devienne la personne que je suis aujourd'hui. *Papa* pour avoir toujours été fier de moi et m'avoir encouragée dans mes efforts. *Maman* pour être un pilier dans ma vie, et pour m'avoir soutenue dans tous mes projets même si ceux-ci m'ont un peu éloignée géographiquement.

Mon *Benjamin*, pour ton immense soutien jour après jour, pour m'avoir suivi dans mes aventures au bout du monde, avoir apporté un peu de ta créativité à cette thèse, pour nos rires, nos rêves et nos moments de bonheur au quotidien. Ma petite *Alice*, parce que le bonheur que tu apportes dans ma vie ne se mesure pas.

References

1. Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J. Clinical development success rates for investigational drugs. *Nature biotechnology* 2014, **32**(1)**:** 40-51.
2. Venkatakrishnan K, Friberg LE, Ouellet D, Mettetal JT, Stein A, Troconiz IF*, et al.* Optimizing oncology therapeutics through quantitative translational and clinical pharmacology: challenges and opportunities. *Clinical pharmacology and therapeutics* 2015, **97**(1)**:** 37-54.
3. Kummar S, Gutierrez M, Doroshow JH, Murgo AJ. Drug development in oncology: classical cytotoxics and molecularly targeted agents. *British journal of clinical pharmacology* 2006, **62**(1)**:** 15-26.
4. Food and Drug Administration. Challenge and opportunity on the Critical Path to new medical products (https://www.fda.gov/downloads/ScienceResearch/SpecialTopics/CriticalPathInitiative/CriticalPathOpportunitiesReports/UCM113411.pdf). 2004
5. Manolis E, Rohou S, Hemmings R, Salmonson T, Karlsson M, Milligan PA. The Role of Modeling and Simulation in Development and Registration of Medicinal Products: Output From the EFPIA/EMA Modeling and Simulation Workshop. *CPT: pharmacometrics & systems pharmacology* 2013, **2:** e31.
6. Workgroup EM, Marshall SF, Burghaus R, Cosson V, Cheung SY, Chenel M*, et al.* Good Practices in Model-Informed Drug Discovery and Development: Practice, Application, and Documentation. *CPT: pharmacometrics & systems pharmacology* 2016, **5**(3)**:** 93-122.
7. Buil-Bruna N, Lopez-Picazo JM, Martin-Algarra S, Troconiz IF. Bringing Model-Based Prediction to Oncology Clinical Practice: A Review of Pharmacometrics Principles and Applications. *The oncologist* 2016, **21**(2)**:** 220-232.
8. Biomarkers Definitions Working G. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clinical pharmacology and therapeutics* 2001, **69**(3)**:** 89-95.
9. Food and Drug Administration. Guidance for Industry: Clinical trial endpoints for the approval of cancer drugs and biologics. [cited April 9th, 2015] Available from: http://www.fda.gov/downloads/Drugs/Guidances/ucm071590.pdf
10. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R*, et al.* New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 2009, **45**(2)**:** 228-247.
11. Wilson MK, Collyar D, Chingos DT, Friedlander M, Ho TW, Karakasis K*, et al.* Outcomes and endpoints in cancer trials: bridging the divide. *The lancet oncology* 2015, **16**(1)**:** e43-52.
12. National Cancer Institute. NCI Dictionary of Cancer Terms. [cited 2015-04-07]Available from: http://www.cancer.gov/dictionary

13. Sharma MR, Maitland ML, Ratain MJ. RECIST: no longer the sharpest tool in the oncology clinical trials toolbox---point. *Cancer research* 2012, **72**(20)**:** 5145-5149; discussion 5150.
14. Jain RK, Duda DG, Willett CG, Sahani DV, Zhu AX, Loeffler JS*, et al.* Biomarkers of response and resistance to antiangiogenic therapy. *Nature reviews Clinical oncology* 2009, **6**(6)**:** 327-338.
15. Waterton JC, Pylkkanen L. Qualification of imaging biomarkers for oncology drug development. *Eur J Cancer* 2012, **48**(4)**:** 409-415.
16. Henry NL, Hayes DF. Cancer biomarkers. *Molecular oncology* 2012, **6**(2)**:** 140-146.
17. Nishida N, Yano H, Nishida T, Kamura T, Kojiro M. Angiogenesis in cancer. *Vascular health and risk management* 2006, **2**(3)**:** 213-219.
18. Smith NR, Baker D, James NH, Ratcliffe K, Jenkins M, Ashton SE*, et al.* Vascular endothelial growth factor receptors VEGFR-2 and VEGFR-3 are localized primarily to the vasculature in human primary solid cancers. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2010, **16**(14)**:** 3548-3561.
19. Zurita AJ, Jonasch E, Wu HK, Tran HT, Heymach JV. Circulating biomarkers for vascular endothelial growth factor inhibitors in renal cell carcinoma. *Cancer* 2009, **115**(10 Suppl)**:** 2346-2354.
20. Duda DG. Molecular Biomarkers of Response to Antiangiogenic Therapy for Cancer. *ISRN cell biology* 2012, **2012**.
21. Deprimo SE, Huang X, Blackstein ME, Garrett CR, Harmon CS, Schoffski P*, et al.* Circulating levels of soluble KIT serve as a biomarker for clinical outcome in gastrointestinal stromal tumor patients receiving sunitinib following imatinib failure. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2009, **15**(18)**:** 5869-5877.
22. Hansson EK, Ma G, Amantea MA, French J, Milligan PA, Friberg LE*, et al.* PKPD Modeling of Predictors for Adverse Effects and Overall Survival in Sunitinib-Treated Patients With GIST. *CPT: pharmacometrics & systems pharmacology* 2013, **2:** e85.
23. Hansson EK, Amantea MA, Westwood P, Milligan PA, Houk BE, French J*, et al.* PKPD Modeling of VEGF, sVEGFR-2, sVEGFR-3, and sKIT as Predictors of Tumor Dynamics and Overall Survival Following Sunitinib Treatment in GIST. *CPT: pharmacometrics & systems pharmacology* 2013, **2:** e84.
24. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine* 2009, **50 Suppl 1:** 122S-150S.
25. Vanderhoek M, Perlman SB, Jeraj R. Impact of different standardized uptake value measures on PET-based quantification of treatment response. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine* 2013, **54**(8)**:** 1188-1194.
26. Pinker K, Riedl C, Weber WA. Evaluating tumor response with FDG PET: updates on PERCIST, comparison with EORTC criteria and clues to future developments. *European journal of nuclear medicine and molecular imaging* 2017, **44**(Suppl 1)**:** 55-66.

27. Young H, Baum R, Cremerius U, Herholz K, Hoekstra O, Lammertsma AA*, et al.* Measurement of clinical and subclinical tumour response using [18F]-fluorodeoxyglucose and positron emission tomography: review and 1999 EORTC recommendations. European Organization for Research and Treatment of Cancer (EORTC) PET Study Group. *Eur J Cancer* 1999, **35**(13)**:** 1773-1782.
28. Van den Abbeele AD. The lessons of GIST--PET and PET/CT: a new paradigm for imaging. *The oncologist* 2008, **13 Suppl 2:** 8-13.
29. Demetri GD, Heinrich MC, Fletcher JA, Fletcher CD, Van den Abbeele AD, Corless CL*, et al.* Molecular target modulation, imaging, and clinical evaluation of gastrointestinal stromal tumor patients treated with sunitinib malate after imatinib failure. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2009, **15**(18)**:** 5902-5909.
30. Dimitrakopoulou-Strauss A, Ronellenfitsch U, Cheng C, Pan L, Sachpekidis C, Hohenberger P*, et al.* Imaging therapy response of gastrointestinal stromal tumors (GIST) with FDG PET, CT and MRI: a systematic review. *Clinical and translational imaging* 2017, **5**(3)**:** 183-197.
31. Walz PC, Bush ML, Robinett Z, Kirsch CF, Welling DB. Three-dimensional segmented volumetric analysis of sporadic vestibular schwannomas: comparison of segmented and linear measurements. *Otolaryngology--head and neck surgery : official journal of American Academy of Otolaryngology-Head and Neck Surgery* 2012, **147**(4)**:** 737-743.
32. Welsh JL, Bodeker K, Fallon E, Bhatia SK, Buatti JM, Cullen JJ. Comparison of response evaluation criteria in solid tumors with volumetric measurements for estimation of tumor burden in pancreatic adenocarcinoma and hepatocellular carcinoma. *American journal of surgery* 2012, **204**(5)**:** 580-585.
33. Kuo CJ, Ke BH, Wu NY, Kuo J, Hsu HH. Prognostic value of tumor volume for patients with advanced lung cancer treated with chemotherapy. *Computer methods and programs in biomedicine* 2017, **144:** 165-177.
34. Schiavon G, Ruggiero A, Bekers DJ, Barry PA, Sleijfer S, Kloth J*, et al.* The effect of baseline morphology and its change during treatment on the accuracy of Response Evaluation Criteria in Solid Tumours in assessment of liver metastases. *Eur J Cancer* 2014, **50**(5)**:** 972-980.
35. Schiavon G, Ruggiero A, Schoffski P, van der Holt B, Bekers DJ, Eechoute K*, et al.* Tumor volume as an alternative response measurement for imatinib treated GIST patients. *PloS one* 2012, **7**(11)**:** e48372.
36. Choi H, Charnsangavej C, Faria SC, Macapinlac HA, Burgess MA, Patel SR*, et al.* Correlation of computed tomography and positron emission tomography in patients with metastatic gastrointestinal stromal tumor treated at a single institution with imatinib mesylate: proposal of new computed tomography response criteria. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2007, **25**(13)**:** 1753-1759.
37. Ronot M, Bouattour M, Wassermann J, Bruno O, Dreyer C, Larroque B*, et al.* Alternative Response Criteria (Choi, European association for the study of the liver, and modified Response Evaluation Criteria in Solid Tumors [RECIST]) Versus RECIST 1.1 in patients with advanced hepatocellular carcinoma treated with sorafenib. *The oncologist* 2014, **19**(4)**:** 394-402.
38. van der Veldt AA, Meijerink MR, van den Eertwegh AJ, Haanen JB, Boven E. Choi response criteria for early prediction of clinical outcome in patients with metastatic renal cell cancer treated with sunitinib. *British journal of cancer* 2010, **102**(5)**:** 803-809.

39. Abramson RG, Lakomkin N, Hainline A, Kang H, Hutson MS, Arteaga CL. The Attenuation Distribution Across the Long Axis of Breast Cancer Liver Metastases at CT: A Quantitative Biomarker for Predicting Overall Survival. *AJR American journal of roentgenology* 2017**:** W1-W7.

40. National Cancer Institute. Common Terminology Criteria for Adverse Events (CTCAE). [cited 2017/11/08]Available from: https://ctep.cancer.gov/protocoldevelopment/electronic_applications/ctc.htm
41. Izzedine H, Rixe O, Billemont B, Baumelou A, Deray G. Angiogenesis inhibitor therapies: focus on kidney toxicity and hypertension. *American journal of kidney diseases : the official journal of the National Kidney Foundation* 2007, **50**(2)**:** 203-218.
42. Li J, Gu J. Hand-foot skin reaction with vascular endothelial growth factor receptor tyrosine kinase inhibitors in cancer patients: A systematic review and meta-analysis. *Critical reviews in oncology/hematology* 2017, **119:** 50-58.
43. Rini BI, Cohen DP, Lu DR, Chen I, Hariharan S, Gore ME*, et al.* Hypertension as a biomarker of efficacy in patients with metastatic renal cell carcinoma treated with sunitinib. *Journal of the National Cancer Institute* 2011, **103**(9)**:** 763-773.
44. George S, Reichardt P, Lechner T, Li S, Cohen DP, Demetri GD. Hypertension as a potential biomarker of efficacy in patients with gastrointestinal stromal tumor treated with sunitinib. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO* 2012, **23**(12)**:** 3180-3187.
45. Basch E. The rationale for collecting patient-reported symptoms during routine chemotherapy. *American Society of Clinical Oncology educational book / ASCO American Society of Clinical Oncology Meeting* 2014**:** 161-165.
46. Basch E. Beyond the FDA PRO guidance: steps toward integrating meaningful patient-reported outcomes into regulatory trials and US drug labels. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research* 2012, **15**(3)**:** 401-403.
47. Patrick DL, Burke LB, Powers JH, Scott JA, Rock EP, Dawisha S*, et al.* Patient-reported outcomes to support medical product labeling claims: FDA perspective. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research* 2007, **10 Suppl 2:** S125-137.
48. Brady MJ, Cella DF, Mo F, Bonomi AE, Tulsky DS, Lloyd SR*, et al.* Reliability and validity of the Functional Assessment of Cancer Therapy-Breast quality-of-life instrument. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 1997, **15**(3)**:** 974-986.
49. Ette EI, Williams PJ. *Pharmacometrics : the science of quantitative pharmacology*. John Wiley & Sons: Hoboken, N.J., 2007.
50. Bender BC, Schindler E, Friberg LE. Population pharmacokinetic-pharmacodynamic modelling in oncology: a tool for predicting clinical response. *British journal of clinical pharmacology* 2015, **79**(1)**:** 56-71.
51. Ribba B, Holford NH, Magni P, Troconiz I, Gueorguieva I, Girard P*, et al.* A review of mixed-effects models of tumor growth and effects of anticancer drug treatment used in population analysis. *CPT: pharmacometrics & systems pharmacology* 2014, **3:** e113.
52. Mould DR, Walz AC, Lave T, Gibbs JP, Frame B. Developing Exposure/Response Models for Anticancer Drug Treatment: Special Considerations. *CPT: pharmacometrics & systems pharmacology* 2015, **4**(1)**:** n/a-n/a.

53. Bonate PL. *Pharmacokinetic-pharmacodynamic modeling and simulation*. Springer: New York, 2011.
54. Claret L, Girard P, Hoff PM, Van Cutsem E, Zuideveld KP, Jorga K*, et al.* Model-based prediction of phase III overall survival in colorectal cancer on the basis of phase II tumor dynamics. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2009, **27**(25)**:** 4103-4108.
55. Schindler E, Karlsson MO. A Minimal Continuous-Time Markov Pharmacometric Model. *The AAPS journal* 2017.
56. Nguyen TH, Han HR, Kim MT, Chan KS. An introduction to item response theory for patient-reported outcome measurement. *The patient* 2014, **7**(1)**:** 23-35.
57. Chang CH, Reeve BB. Item response theory and its applications to patient-reported outcomes measurement. *Evaluation & the health professions* 2005, **28**(3)**:** 264-282.
58. Ueckert S, Plan EL, Ito K, Karlsson MO, Corrigan B, Hooker AC*, et al.* Improved utilization of ADAS-cog assessment data through item response theory based pharmacometric modeling. *Pharmaceutical research* 2014, **31**(8)**:** 2152-2165.
59. Novakovic AM, Krekels EH, Munafo A, Ueckert S, Karlsson MO. Application of Item Response Theory to Modeling of Expanded Disability Status Scale in Multiple Sclerosis. *The AAPS journal* 2017, **19**(1)**:** 172-179.
60. Krekels E, Novakovic AM, Vermeulen AM, Friberg LE, Karlsson MO. Item Response Theory to Quantify Longitudinal Placebo and Paliperidone Effects on PANSS Scores in Schizophrenia. *CPT: pharmacometrics & systems pharmacology* 2017, **6**(8)**:** 543-551.
61. Gottipati G, Karlsson MO, Plan EL. Modeling a Composite Score in Parkinson's Disease Using Item Response Theory. *The AAPS journal* 2017, **19**(3)**:** 837-845.
62. Buatois S, Retout S, Frey N, Ueckert S. Item Response Theory as an Efficient Tool to Describe a Heterogeneous Clinical Rating Scale in De Novo Idiopathic Parkinson's Disease Patients. *Pharmaceutical research* 2017, **34**(10)**:** 2109-2118.
63. Desmee S, Mentre F, Veyrat-Follet C, Guedj J. Nonlinear Mixed-effect Models for Prostate-specific Antigen Kinetics and Link with Survival in the Context of Metastatic Prostate Cancer: A Comparison by Simulation of Two-stage and Joint Approaches. *The AAPS journal* 2015, **17**(3)**:** 691-699.
64. Zhang L, Beal SL, Sheiner LB. Simultaneous vs. sequential analysis for population PK/PD data I: best-case performance. *Journal of pharmacokinetics and pharmacodynamics* 2003, **30**(6)**:** 387-404.
65. Lacroix BD, Friberg LE, Karlsson MO. Evaluation of IPPSE, an alternative method for sequential population PKPD analysis. *Journal of pharmacokinetics and pharmacodynamics* 2012, **39**(2)**:** 177-193.
66. Bauer RJ, Guzy S, Ng C. A survey of population analysis methods and software for complex pharmacokinetic and pharmacodynamic models with examples. *The AAPS journal* 2007, **9**(1)**:** E60-83.
67. Iqbal N, Iqbal N. Imatinib: a breakthrough of targeted therapy in cancer. *Chemotherapy research and practice* 2014, **2014:** 357027.
68. Raymond E, Faivre S. Learning experiences with sunitinib continuous daily dosing in patients with pancreatic neuroendocrine tumours. *Current oncology* 2014, **21**(6)**:** 309-317.

69. Tzogani K, Skibeli V, Westgaard I, Dalhus M, Thoresen H, Slot KB*, et al.* The European Medicines Agency approval of axitinib (Inlyta) for the treatment of advanced renal cell carcinoma after failure of prior treatment with sunitinib or a cytokine: summary of the scientific assessment of the committee for medicinal products for human use. *The oncologist* 2015, **20**(2)**:** 196-201.
70. Ryan Q, Ibrahim A, Cohen MH, Johnson J, Ko CW, Sridhara R*, et al.* FDA drug approval summary: lapatinib in combination with capecitabine for previously treated metastatic breast cancer that overexpresses HER-2. *The oncologist* 2008, **13**(10)**:** 1114-1119.
71. Peddi PF, Hurvitz SA. Ado-trastuzumab emtansine (T-DM1) in human epidermal growth factor receptor 2 (HER2)-positive metastatic breast cancer: latest evidence and clinical potential. *Therapeutic advances in medical oncology* 2014, **6**(5)**:** 202-209.
72. Plan EL, Elshoff JP, Stockis A, Sargentini-Maier ML, Karlsson MO. Likert pain score modeling: a Markov integer model and an autoregressive continuous model. *Clinical pharmacology and therapeutics* 2012, **91**(5)**:** 820-828.
73. Tomita Y, Uemura H, Fujimoto H, Kanayama HO, Shinohara N, Nakazawa H*, et al.* Key predictive factors of axitinib (AG-013736)-induced proteinuria and efficacy: a phase II study in Japanese patients with cytokine-refractory metastatic renal cell Carcinoma. *Eur J Cancer* 2011, **47**(17)**:** 2592-2602.
74. Welslau M, Dieras V, Sohn JH, Hurvitz SA, Lalla D, Fang L*, et al.* Patient-reported outcomes from EMILIA, a randomized phase 3 study of trastuzumab emtansine (T-DM1) versus capecitabine and lapatinib in human epidermal growth factor receptor 2-positive locally advanced or metastatic breast cancer. *Cancer* 2014, **120**(5)**:** 642-651.
75. Rini BI, Garrett M, Poland B, Dutcher JP, Rixe O, Wilding G*, et al.* Axitinib in metastatic renal cell carcinoma: results of a pharmacokinetic and pharmacodynamic analysis. *Journal of clinical pharmacology* 2013, **53**(5)**:** 491-504.
76. Houk BE, Bello CL, Kang D, Amantea M. A population pharmacokinetic meta-analysis of sunitinib malate (SU11248) and its primary metabolite (SU12662) in healthy volunteers and oncology patients. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2009, **15**(7)**:** 2497-2506.
77. Lu D, Girish S, Gao Y, Wang B, Yi JH, Guardino E*, et al.* Population pharmacokinetics of trastuzumab emtansine (T-DM1), a HER2-targeted antibody-drug conjugate, in patients with HER2-positive metastatic breast cancer: clinical implications of the effect of covariates. *Cancer chemotherapy and pharmacology* 2014, **74**(2)**:** 399-410.
78. Hansson EK. Pharmacometric Models for Biomarkers, Side Effects and Efficacy in Anticancer Drug Therapy. Acta Universitatis Upsaliensis, Uppsala, 2012.
79. Papaetis GS, Syrigos KN. Sunitinib: a multitargeted receptor tyrosine kinase inhibitor in the era of molecular cancer therapies. *BioDrugs : clinical immunotherapeutics, biopharmaceuticals and gene therapy* 2009, **23**(6)**:** 377-389.
80. Keizer RJ, Karlsson MO, Hooker A. Modeling and Simulation Workbench for NONMEM: Tutorial on Pirana, PsN, and Xpose. *CPT: pharmacometrics & systems pharmacology* 2013, **2:** e50.
81. Bergstrand M, Hooker AC, Wallin JE, Karlsson MO. Prediction-corrected visual predictive checks for diagnosing nonlinear mixed-effects models. *The AAPS journal* 2011, **13**(2)**:** 143-151.

82. Petersson KJ, Hanze E, Savic RM, Karlsson MO. Semiparametric distributions with estimated shape parameters. *Pharmaceutical research* 2009, **26**(9)**:** 2174-2185.
83. Claret L, Mercier F, Houk BE, Milligan PA, Bruno R. Modeling and simulations relating overall survival to tumor growth inhibition in renal cell carcinoma patients. *Cancer chemotherapy and pharmacology* 2015, **76**(3)**:** 567-573.
84. Escudier B, Rini BI, Motzer RJ, Tarazi J, Kim S, Huang X*, et al.* Genotype Correlations With Blood Pressure and Efficacy From a Randomized Phase III Trial of Second-Line Axitinib Versus Sorafenib in Metastatic Renal Cell Carcinoma. *Clinical genitourinary cancer* 2015, **13**(4)**:** 328-337 e323.
85. Diekstra MH, Fritsch A, Kanefendt F, Swen JJ, Moes D, Sorgel F*, et al.* Population Modeling Integrating Pharmacokinetics, Pharmacodynamics, Pharmacogenetics, and Clinical Outcome in Patients With Sunitinib-Treated Cancer. *CPT: pharmacometrics & systems pharmacology* 2017, **6**(9)**:** 604-613.
86. Ait-Oudhia S, Mager DE, Pokuri V, Tomaszewski G, Groman A, Zagst P*, et al.* Bridging Sunitinib Exposure to Time-to-Tumor Progression in Hepatocellular Carcinoma Patients With Mathematical Modeling of an Angiogenic Biomarker. *CPT: pharmacometrics & systems pharmacology* 2016, **5**(6)**:** 297-304.
87. Eto M, Uemura H, Tomita Y, Kanayama H, Shinohara N, Kamei Y*, et al.* Overall survival and final efficacy and safety results from a Japanese phase II study of axitinib in cytokine-refractory metastatic renal cell carcinoma. *Cancer science* 2014, **105**(12)**:** 1576-1583.
88. Demetri GD, van Oosterom AT, Garrett CR, Blackstein ME, Shah MH, Verweij J*, et al.* Efficacy and safety of sunitinib in patients with advanced gastrointestinal stromal tumour after failure of imatinib: a randomised controlled trial. *Lancet* 2006, **368**(9544)**:** 1329-1338.
89. Suleiman AA, Frechen S, Scheffler M, Zander T, Kahraman D, Kobe C*, et al.* Modeling tumor dynamics and overall survival in advanced non-small-cell lung cancer treated with erlotinib. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer* 2015, **10**(1)**:** 84-92.
90. Kayani I, Avril N, Bomanji J, Chowdhury S, Rockall A, Sahdev A*, et al.* Sequential FDG-PET/CT as a biomarker of response to Sunitinib in metastatic clear cell renal cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2011, **17**(18)**:** 6021-6028.
91. Suzuki K, Kohlbrenner R, Epstein ML, Obajuluwa AM, Xu J, Hori M. Computer-aided measurement of liver volumes in CT by means of geodesic active contour segmentation coupled with level-set algorithms. *Medical physics* 2010, **37**(5)**:** 2159-2166.
92. Eechoute K, Fransson MN, Reyners AK, de Jong FA, Sparreboom A, van der Graaf WT*, et al.* A long-term prospective population pharmacokinetic study on imatinib plasma concentrations in GIST patients. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2012.
93. Choi JW, Choi D, Kim KM, Sohn TS, Lee JH, Kim HJ*, et al.* Small submucosal tumors of the stomach: differentiation of gastric schwannoma from gastrointestinal stromal tumor with CT. *Korean journal of radiology : official journal of the Korean Radiological Society* 2012, **13**(4)**:** 425-433.

94. Farag S, de Geus-Oei LF, Van der Graaf WT, van Coevorden F, Grunhagen DJ, Reyners AKL*, et al.* Early response evaluation by (18)F-FDG-PET influences management in gastrointestinal stromal tumor (GIST) patients treated with imatinib with neo-adjuvant intent. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine* 2017.
95. Desar IM, Mulder SF, Stillebroer AB, van Spronsen DJ, van der Graaf WT, Mulders PF*, et al.* The reverse side of the victory: flare up of symptoms after discontinuation of sunitinib or sorafenib in renal cell cancer patients. A report of three cases. *Acta Oncol* 2009, **48**(6)**:** 927-931.

96. George S, Blay JY, Casali PG, Le Cesne A, Stephenson P, Deprimo SE*, et al.* Clinical evaluation of continuous daily dosing of sunitinib malate in patients with advanced gastrointestinal stromal tumour after imatinib failure. *Eur J Cancer* 2009, **45**(11)**:** 1959-1968.
97. Bjarnason GA, Khalil B, Hudson JM, Williams R, Milot LM, Atri M*, et al.* Outcomes in patients with metastatic renal cell cancer treated with individualized sunitinib therapy: correlation with dynamic microbubble ultrasound data and review of the literature. *Urologic oncology* 2014, **32**(4)**:** 480-487.
98. Ferte C, Fernandez M, Hollebecque A, Koscielny S, Levy A, Massard C*, et al.* Tumor growth rate is an early indicator of antitumor drug activity in phase I clinical trials. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2014, **20**(1)**:** 246-252.
99. Yoo HJ, Ahn SH, Eremenco S, Kim H, Kim WK, Kim SB*, et al.* Korean translation and validation of the functional assessment of cancer therapy-breast (FACT-B) scale version 4. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 2005, **14**(6)**:** 1627-1632.
100. Pandey M, Thomas BC, Ramdas K, Eremenco S, Nair MK. Quality of life in breast cancer patients: validation of a FACT-B Malayalam version. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 2002, **11**(2)**:** 87-90.
101. Patoo M, Allahyari AA, Moradi AR, Payandeh M. Persian Version of Functional Assessment of Cancer Therapy- Breast (FACT-B) Scale: Confirmatory Factor Analysis and Psychometric Properties. *Asian Pacific journal of cancer prevention : APJCP* 2015, **16**(9)**:** 3799-3803.
102. Petrillo J, Cano SJ, McLeod LD, Coon CD. Using classical test theory, item response theory, and Rasch measurement theory to evaluate patient-reported outcome measures: a comparison of worked examples. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research* 2015, **18**(1)**:** 25-34.
103. Li C, Wang B, Chen SC, Wada R, Lu D, Wang X*, et al.* Exposure-response analyses of trastuzumab emtansine in patients with HER2-positive advanced breast cancer previously treated with trastuzumab and a taxane. *Cancer chemotherapy and pharmacology* 2017.
104. Bender BC, Schaedeli-Stark F, Koch R, Joshi A, Chu YW, Rugo H*, et al.* A population pharmacokinetic/pharmacodynamic model of thrombocytopenia characterizing the effect of trastuzumab emtansine (T-DM1) on platelet counts in patients with HER2-positive metastatic breast cancer. *Cancer chemotherapy and pharmacology* 2012.

105.Wang J, Song P, Schrieber S, Liu Q, Xu Q, Blumenthal G*, et al.* Exposure-response relationship of T-DM1: insight into dose optimization for patients with HER2-positive metastatic breast cancer. *Clinical pharmacology and therapeutics* 2014, **95**(5)**:** 558-564.
106.Glangkarn S, Promasatayaprot V, Porock D, Edgley A. Measuring quality of life in thai women with breast cancer. *Asian Pacific journal of cancer prevention : APJCP* 2011, **12**(3)**:** 637-644.
107.Ng R, Lee CF, Wong NS, Luo N, Yap YS, Lo SK*, et al.* Measurement properties of the English and Chinese versions of the Functional Assessment of Cancer Therapy-Breast (FACT-B) in Asian breast cancer patients. *Breast cancer research and treatment* 2012, **131**(2)**:** 619-625.
108.Bonomi AE, Cella DF, Hahn EA, Bjordal K, Sperner-Unterweger B, Gangeri L*, et al.* Multilingual translation of the Functional Assessment of Cancer Therapy (FACT) quality of life measurement system. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 1996, **5**(3)**:** 309-320.
109.Chee Chean D, Kuo Zang W, Lim M, Zulkefle N. Health Related Quality of Life (HRQoL) among Breast Cancer Patients Receiving Chemotherapy in Hospital Melaka: Single Centre Experience. *Asian Pacific journal of cancer prevention : APJCP* 2016, **17**(12)**:** 5121-5126.
110.Dehkordi A, Heydarnejad MS, Fatehi D. Quality of Life in Cancer Patients undergoing Chemotherapy. *Oman medical journal* 2009, **24**(3)**:** 204-207.
111.Gnanasakthy A, DeMuro C, Clark M, Haydysch E, Ma E, Bonthapally V. Patient-Reported Outcomes Labeling for Products Approved by the Office of Hematology and Oncology Products of the US Food and Drug Administration (2010-2014). *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2016, **34**(16)**:** 1928-1934.
112.Atkinson TM, Wagner JS, Basch E. Trustworthiness of Patient-Reported Outcomes in Unblinded Cancer Clinical Trials. *JAMA oncology* 2017, **3**(6)**:** 738-739.
113.Gastonguay MR, French JL, Heitjan DF, Rogers JA, Ahn JE, Ravva P. Missing data in model-based pharmacometric applications: points to consider. *Journal of clinical pharmacology* 2010, **50**(9 Suppl)**:** 63S-74S.
114.Holman R, Glas CA. Modelling non-ignorable missing-data mechanisms with item response theory models. *The British journal of mathematical and statistical psychology* 2005, **58**(Pt 1)**:** 1-17.

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Pharmacy 243

Editor: The Dean of the Faculty of Pharmacy

A doctoral dissertation from the Faculty of Pharmacy, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Pharmacy. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Pharmacy".)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-336420

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2018