Citation: CPT Pharmacometrics Syst. Pharmacol. (2019) 8, 230–239; doi:10.1002/psp4.12384

ARTICLE

# Modeling Exposure-Driven Adverse Event Time Courses in Oncology Exemplified by Afatinib

Ronald Niebecker[1,*], Hugo Maas[1], Alexander Staab[1], Matthias Freiwald[1,†] and Mats O. Karlsson[2,†]

**Models were developed to characterize the relationship between afatinib exposure and diarrhea and rash/acne adverse event (AE) trajectories, and their predictive ability was assessed. Based on pooled data from seven phase II/III clinical studies including 998 patients, mixed-effects models for ordered categorical data were applied to describe daily AE severity. Clinical trial simulation aided by trial execution models was used for internal and external model evaluation. The final exposure-safety model consisted of longitudinal logistic regression models with first-order Markov elements for both AEs. Drug exposure was included as daily area under the concentration-time curve (AUC), and drug effects on the AEs were correlated. Clinical trial simulation allowed adequate prediction of maximum AE grades and AE severity time courses but overestimated the proportion of AE-dependent dose reductions and discontinuations. Both diarrhea and rash/acne were correlated with afatinib exposure. The developed modeling framework allows a prospective comparison of dosing strategies and study designs with respect to safety.**

## Study Highlights

**WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?**

☑ As expected from epidermal growth factor receptor inhibition, afatinib is associated with diarrhea and rash/acne adverse events (AEs). These have been shown to be dose dependent and manageable by dose reductions, suggesting a relationship between exposure and response.

**WHAT QUESTION DID THIS STUDY ADDRESS?**

☑ This study developed mixed-effects models to formally characterize the relationship between exposure and AEs and the assessment of their predictive ability for data not included in the model development process.

**WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?**

☑ Longitudinal logistic regression models with first-order Markov elements can describe the AE severity time course. Exposure-response is subject to time-dependent effects, including an attenuated drug effect on AEs during the course of treatment and a delayed drug effect on rash/acne onset. Clinical trial simulations allowed the prediction of maximum AE grades but overpredicted AE-dependent dose reductions and discontinuations.

**HOW MIGHT THIS CHANGE DRUG DISCOVERY, DEVELOPMENT, AND/OR THERAPEUTICS?**

☑ The developed modeling and simulation framework can be used for the prospective comparison of dosing strategies and study designs with respect to safety, thereby supporting future quantitative benefit–risk assessments.

Afatinib is a potent irreversible inhibitor of the ErbB family membrane receptors, including the epidermal growth factor receptor (EGFR). It is in clinical development for various solid tumor indications and has recently been approved for treatment of patients with the EGFR-mutation–positive advanced or metastatic non-small cell lung cancer.[1]

Afatinib's most common treatment-related adverse events (AEs) include diarrhea and rash/acne. These AEs have been shown to be dose dependent[2] and manageable, combining preventive and active treatment options, including a defined and effective dose-reduction scheme validated in clinical trials,[3–5] thereby resulting in only a few patients discontinuing as a result of these AEs.

Standard exposure-response analysis commonly focuses on the highest AE grades observed during the course of treatment within a patient and correlate this with some measure of exposure, e.g., observed trough concentrations.[6–10] However, such an approach neglects the time course of exposure and AE development, thereby losing information on, for example, toxicity-driven dose adjustments as

†Both authors contributed equally.

[1]Translational Medicine and Clinical Pharmacology, Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach, Germany; [2]Department of Pharmaceutical Biosciences, Uppsala University, Uppsala, Sweden. *Correspondence: Ronald Niebecker (Ronald.Niebecker@boehringer-ingelheim.com)

Received: August 27, 2018; accepted: January 2, 2019. doi:10.1002/psp4.12384

**Table 1 Observed data stratified by study**

| | Model building | | | | | | | ME |
|---|---|---|---|---|---|---|---|---|
| **Study** | **A**[16] | **B**[15] | **C**[2] | **D**[17] | **E**[19] | **F**[18] | **Total** | **G**[3] |
| Patients | 50 | 41 | 129 | 390 | 97 | 62 | 769 | 229 |
| W/o PK data (%) | 0 (0) | 1 (2.4) | 3 (2.3) | 37 (9.5) | 24 (25) | 2 (3.2) | 67 (8.7) | 17 (7.4) |
| Max. diarrhea (%) | | | | | | | | |
| 0 | 4 (8.0) | 4 (9.8) | 7 (5.4) | 50 (13) | 26 (27) | 0 (0) | 91 (12) | 10 (4.4) |
| 1 | 10 (20) | 18 (44) | 54 (42) | 148 (38) | 37 (38) | 12 (19) | 279 (36) | 96 (42) |
| 2 | 16 (32) | 9 (22) | 42 (33) | 125 (32) | 19 (20) | 27 (44) | 238 (31) | 88 (38) |
| ≥ 3 | 20 (40) | 10 (24) | 26 (20) | 67 (17) | 15 (15) | 23 (37) | 161 (21) | 35 (15) |
| Max. rash/acne (%) | | | | | | | | |
| 0 | 13 (26) | 6 (15) | 8 (6.2) | 83 (21) | 26 (27) | 5 (8.1) | 141 (18) | 23 (10) |
| 1 | 12 (24) | 9 (22) | 30 (23) | 118 (30) | 22 (23) | 8 (13) | 199 (26) | 68 (30) |
| 2 | 20 (40) | 21 (51) | 61 (47) | 133 (34) | 29 (30) | 32 (52) | 296 (38) | 101 (44) |
| ≥ 3 | 5 (10) | 5 (12) | 30 (23) | 56 (14) | 20 (21) | 17 (27) | 133 (17) | 37 (16) |
| Median days observation time (censored[a]) | 74.5 (16.0) | 92.0 (37.0) | 393 (169) | 113 (70.0) | 99.0 (65.0) | 148 (65.5) | 124 (70) | 364 |
| Days with AE status observations (censored[a]) | 4,375 (2,140) | 5,175 (2,993) | 66,645 (39,075) | 62,272 (42,678) | 12,570 (8,708) | 11,560 (7,294) | 162,597 (102,888) | 106,435 |
| Dose reductions (%) | | | | | | | | |
| Associated with diarrhea and/or rash/acne[b] | 19 (76) | 16 (59) | 78 (68) | 138 (72) | 22 (56) | 51 (78) | 324 (70) | 95 (52) |
| Because of other AEs | 6 (24) | 11 (41) | 37 (32) | 55 (28) | 17 (44) | 14 (22) | 140 (30) | 89 (48) |
| Discontinuation (%) | | | | | | | | |
| Associated with diarrhea and/or rash/acne[b] | 18 (36) | 6 (15) | 1 (0.78) | 26 (6.7) | 6 (6.2) | 8 (13) | 65 (8.5) | 6 (2.6) |
| Progression | 31 (62) | 30 (73) | 101 (78) | 322 (83) | 55 (57) | 44 (71) | 583 (76) | 180 (79) |
| Other reasons | 1 (2) | 5 (12) | 21 (16) | 42 (11) | 36 (37) | 10 (16) | 115 (15) | 28 (12) |
| Censored | 0 (0) | 0 (0) | 6 (4.7) | 0 (0) | 0 (0) | 0 (0) | 6 (0.78) | 15 (6.6) |

AE, adverse event; ME, model evaluation set; PK, pharmacokinetic; w/o, without.
[a]For details on censoring, see Materials and Methods. [b]For details on definition of association, see Materials and Methods. Superscript numbers refer to literature references

**Table 2 Protocol specifications concerning treatment, dose reductions, and discontinuations because of undue toxicity**

| Study | Treatment | Criteria for dose reduction | Criteria for discontinuation |
|---|---|---|---|
| A–F | Starting dose: 50 mg q.d. 40 mg q.d. after protocol amendment in study C<br>No dose escalation<br>Max. 2 dose reduction in 10 mg increments | Any drug-related adverse event of CTCAE grade ≥ 3<br>CTCAE grade ≥ 3 diarrhea or grade ≥ 2 diarrhea for ≥ 7 consecutive days, despite antidiarrheal medication | AE qualifying for dose reduction after two previous dose reductions<br>No recovery to CTCAE grade 1/no AE within 14 days treatment interruption after AE qualifying for dose reduction |
| G | Starting dose: 40 mg q.d.<br>Dose escalation to 50 mg q.d., in case of no/only mild toxicity during initial 21 days of treatment<br>Dose reduction in 10 mg increments, down to 20 mg q.d. | Any drug-related adverse event of CTCAE grade ≥ 3<br>CTCAE grade ≥ 3 diarrhea or grade ≥ 2 diarrhea for ≥ 48 consecutive hours, despite antidiarrheal medication | AE qualifying for dose reduction at 20 mg q.d. dose<br>No recovery to CTCAE grade 1/no AE within 14 days treatment interruption after AE qualifying for dose reduction, and no indication of obvious clinical benefit allowing for prolonged recovery period |

AE, adverse event; CTCAE, common terminology criteria for adverse events; q.d., once daily.

well as recurrent AE episodes. Longitudinal logistic regression models offer the possibility to integrate these dynamics in a quantitative way and further allow the analysis of all observed AE grades, e.g., by treating these as ordered categorical data.[11–14] Hence, the objectives of this study were to (i) characterize the relationship between afatinib exposure and diarrhea and rash/acne AEs and (ii) explore the applicability of the developed models to predict the AE

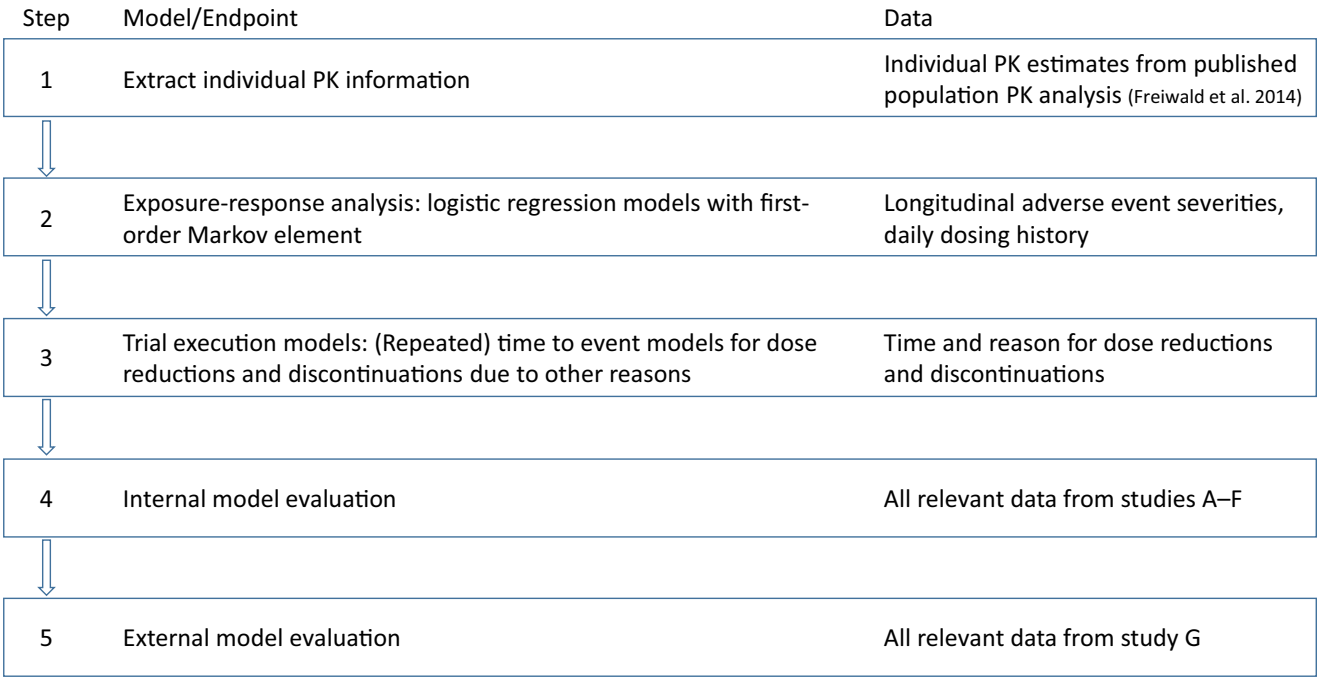


**Figure 1** Visualization of the model building und model evaluation steps, including the required data. PK, pharmacokinetic.

occurrence, severity, and resulting treatment decisions in typical phase II/phase III clinical trials.

## MATERIALS AND METHODS

### Patients and study design

This analysis was based on data from seven monotherapy phase II/III clinical studies,[2,3,15–19] including a total of 998 patients with various solid tumors (**Table 1**). All of the studies were conducted in accordance with the Declaration of Helsinki and the International Conference on Harmonization (ICH) guidelines for Good Clinical Practice and approved by relevant regulatory and independent ethics committees.

The data from six of the studies (A–F) were used for model building (starting dose: almost exclusively 50 mg q.d., exceeding the approved standard starting dose of afatinib of 40 mg q.d.) and the data of the remaining study (G) for model evaluation. In all studies, afatinib was administered continuously until the occurrence of disease progression or undue toxicity. **Table 2** provides a summary of the protocol specifications in each study concerning the starting dose, dose reductions, and if applicable dose escalations as well as regulations for the management of AEs, in particular those to be expected from EGFR inhibition. **Figure 1** provides a visualization of the model building and model evaluation steps, including the required data.

### Data analysis

Model fitting was carried out in NONMEM 7.2/7.3 (ICON Development Solutions, Ellicott City, MD),[20] aided by Perl-speaks-NONMEM (PsN).[21] For the exposure-AE models and the trial execution models, the Laplacian method and the exact likelihood were used for parameter estimation, respectively. Parameter uncertainty was determined from the variance-covariance matrix provided by NONMEM. R 3.0[22] was used for pre- and postprocessing.

The AE data set contained daily AE severity graded according to the Common Terminology Criteria for Adverse Events,[23] with grades 0, 1, 2, and ≥ 3 based on onset, end date, and severity grade of AE episodes recorded by investigators from the start of treatment until up to 28 days after the last dose. All diarrhea and rash/acne episodes with an onset date within this time period, i.e., treatment-emergent AEs, were considered independent of whether they were classified as treatment related. In case of (partial) missing onset or end dates of AE episodes, these were imputed in the source database using a conservative algorithm that maximized treatment-related episodes.

Actual patient-specific daily dosing history was implemented, and individual daily afatinib exposure of patients in terms of area under the plasma concentration-time curve was derived using empirical Bayes estimates from a previous population pharmacokinetic analysis.[24] For patients without evaluable pharmacokinetic observations population predictions were used.

### AE modeling

Longitudinal logistic regression models were applied to describe the patients' AE grades over time for both types of AEs together. The AE data were treated as ordered categorical data, and through a first-order Markov element, the probability of having a certain AE grade each day was dependent on the AE grade the preceding day.[11] The AE category of the day before baseline was assumed to be the same as the AE category on the first day of treatment. Different exposure-response models with or without temporal delays between exposure and AE grade were investigated to

characterize the drug effect, as was the necessity to account for the development of apparent tolerance.

As some of the recorded AE trajectories were atypical, incomplete, or not in accordance with protocol specifications, the following censoring rules were developed based on study protocol specifications and clinical plausibility:

- No dose reduction despite AE qualifying for dose reduction: periods with > 3 days delay in indicated treatment interruption because of AEs censored
- Spontaneous remission from AE qualifying for dose reduction without treatment interruption: periods with remission of grade 3 AE/≥ 7 consecutive days of grade 2 diarrhea censored at day before recovery
- Treatment interruption skipped: periods with dose reduction and no recorded treatment interruption censored at day prior to dose reduction
- Treatment resumed despite nonrecovery: censoring at start of treatment interruption
- AE episodes of implausibly long duration: censoring after > 7 consecutive days of grade 3 diarrhea, > 10 consecutive days of grade 2 diarrhea, or > 14 consecutive days of grade 3 rash/acne.

In addition, the observed data were censored for grade 3 AEs with imputed onset date at start of afatinib treatment, which could result from the conservative imputation algorithm.

To minimize potential impact on parameter estimation and model stability, the final stage of model development was based on censored data only. Simulation-based diagnostics were, however, based on entire data set.

### Model discrimination and evaluation

Model selection was performed based on the comparison of the objective function value (OFV) provided by NONMEM for nested models and the Akaike information criterion for nonnested models. The difference in OFV was assumed to be approximately $\chi^2$ distributed, and a significance level of $P = 0.001$ was used for the addition of extra parameters. Additional criteria for model discrimination were parameter uncertainty and model stability.

Model evaluation mainly relied on simulation-based diagnostics. Simulations were performed based on individual exposure estimates for the subjects included in this analysis. The criteria used in the assessment were both technical (e.g., number and type of transitions between AE grades) and clinically oriented and included but were not limited to maximum AE grades and frequency and reasons of dose reductions and discontinuations. Kaplan–Meier type visual predictive checks (KM-VPCs) for first grade 2/3 AE, first and second dose reduction, and discontinuation were used as additional model evaluation tools.

To allow for meaningful diagnostics, trial execution models had to be developed to account for (i) dropout as a result of disease progression, (ii) dropout as a result of other reasons, e.g., patients who were noncompliant with protocol, lost to follow-up, experienced an AE other than diarrhea and rash/acne, or refused continued medication, and (iii) dose reductions because of AEs other than diarrhea or rash/acne.

In this context, a dose reduction or treatment discontinuation was regarded as "associated with diarrhea and/or rash/acne" if there was a diarrhea and/or rash/acne episode qualifying for dose reduction (see **Table 2**) recorded within an interval of −8 days to +3 days counted from the recorded treatment stop date. As a result of different data cutoff points (entailed by the nature of most oncological trials) and different rules to determine the association used for this analysis, the herein reported total number of dose reductions and discontinuations as well as the number of events associated with diarrhea and/or rash/acne may slightly differ when compared with already reported results.

## RESULTS

### Data

A description of the data sets is provided in **Table 1**. Within the model building subset (starting dose: almost exclusively 50 mg q.d.), diarrhea and rash/acne were observed in 88% and 82% of patients, respectively. Of the dose reductions, 70% were associated with diarrhea and/or rash/acne. In 8.5% of patients, discontinuation was recorded to be the result of an AE, and the treatment stop dates were associated with diarrhea and/or rash/acne. In the model evaluation subset, diarrhea and rash/acne AEs were observed with similar frequency and were associated with 52% of dose reductions and 2.6% of discontinuations (starting dose: 40 mg q.d.).

### AE models

**Diarrhea.** The final exposure diarrhea model was a longitudinal logistic regression model with a first-order Markov element and interindividual variability (IIV) on the baseline probabilities; the parameter estimates are provided in **Table S1**. Accounting for the dependency between neighboring observations significantly improved the model in terms of OFV as well as predicted number of transitions between AE grades. A maximal efficacy ($E_{max}$) model best described the drug effect (for parameterization, see Eqs. 1–5, **Supplemental Material**), with exponential IIV included on $E_{max}$. The magnitude of the drug effect was further dependent on the current AE state, with different drug effects estimated for no AE, grade 1 diarrhea, and grade 2/3 diarrhea. There was no apparent delay in onset or washout of the afatinib effect on diarrhea that could not be explained by the waxing and waning of daily exposure.

The inclusion of an apparent tolerance component additive on the half maximal effective concentration ($EC_{50}$) parameter of the drug effect submodel markedly improved the model ($\Delta$OFV = 730 for 2 df, based on the backward exclusion of this term from the final model). Tolerance itself was best described by an $E_{max}$ model driven by the cumulative number of days on treatment. The addition of two separate penalty terms on the $E_{max}$ parameter of drug effect becoming effective after the first day of grade 2/3 diarrhea AE and rash/acne AE, respectively, resulted in further improvement (see Eqs. 2–4, **Supplemental Material**). Tolerance models driven by drug exposure or cumulative drug exposure were inferior.

**Rash/Acne.** The final exposure rash/acne model (see **Table S1**) was structurally similar to the model for diarrhea, with several exceptions. The magnitude of the drug effect only distinguished different effects for no AE vs. any grade rash/acne. Furthermore, there was a delay in the onset of the drug effect, which was accounted for by including a transit compartment. The washout of effect during the off-treatment periods was faster than the onset. The inclusion of apparent tolerance also improved the model for rash/acne. In contrast to diarrhea, tolerance was best described with an $E_{max}$ model driven by afatinib exposure, with the resensitization of patients during off-treatment periods characterized by a tolerance half-life (see Eqs. 6–12, **Supplemental Material**).

**Joint diarrhea and rash/acne model.** The correlation between the random effects on the $E_{max}$ parameters of the two AE types was positive, with an estimated correlation coefficient of 0.75 and significantly improved the model ($\Delta$OFV = 25 for 1 df, based on the backward exclusion of this parameter from the final model).

### Trial execution models

Dropout as a result of disease progression was described by a time-to-event model with a step function for the hazard describing a decreased risk to dropout within the first 28 days of treatment. Different hazards were estimated for each study.

Dropout as a result of other reasons was described by a separate time-to-event model with a constant hazard and no covariates other than study. A joint hazard rate was estimated for studies A and B, as the number of events in these studies was too low to estimate separate hazard rates.

Dose reductions as a result of AEs other than diarrhea or rash/acne were described by a repeated time-to-event model with constant hazards specific for each study. The patients were not at risk to experience such a dose reduction in case they already had two dose reductions, the previous dose reduction was < 7 days ago, and in case they were in an off-treatment period following a dose reduction as a result of a diarrhea or rash/acne AE. The length of the resulting off-treatment periods was fixed to 7 days, corresponding to the mean duration of such off-treatment periods in the analyzed studies used for model building.

The parameter estimates for all trial execution models are presented in **Table S2**. Model evaluation of the trial execution models with KM-VPCs showed that the selected models were able to mirror the general trend of the observed data (**Figure S1**).

### Model evaluation

**Internal model evaluation.** Based on all subjects included in the model-building data set, simulations of 100 studies were performed. Daily AE severity as well as changes in dose were predicted for a time course of 65 weeks, which corresponded to the time when 90% of the patients had discontinued treatment. All patients were assumed to start treatment with no AEs. A starting dose of 50 mg afatinib was used, and the patients were dose reduced and discontinued according to protocol specifications (see **Table 2**).

**Table 3 Internal model evaluation: maximum adverse event grades, discontinuations, and dose reductions**

| | Observed | Simulated | Range |
|---|---|---|---|
| | Count (%) | Mean (%) | |
| Max CTCAE grades | | | |
| Diarrhea | | | |
| No occurrence | 91 (12) | 107 (14) | 85–132 |
| Grade 1 | 281 (37) | 289 (38) | 249–318 |
| Grade 2 | 237 (31) | 218 (28) | 183–245 |
| Grade 3 | 160 (21) | 154 (20) | 133–185 |
| Rash/acne | | | |
| No occurrence | 141 (18) | 115 (15) | 91–147 |
| Grade 1 | 203 (26) | 229 (30) | 203–257 |
| Grade 2 | 294 (38) | 307 (40) | 272–349 |
| Grade 3 | 131 (17) | 118 (15) | 95–150 |
| Discontinuation[a] | | | |
| Progression | 527 (69) | 495 (64) | 466–528 |
| Other reasons | 109 (14) | 96.8 (13) | 75–117 |
| Diarrhea and/or rash/acne AE | 64 (8.3) | 133 (17) | 108–158 |
| Censored | 69 (9.0) | 44.3 (5.8) | 29–58 |
| Dose reductions | | | |
| No reduction | 425 (55) | 441 (58) | 409–476 |
| Reduction | 344 (45) | 329 (43) | 293–360 |
| 1 reduction | 230 (30) | 215 (28) | 185–250 |
| 2 reductions | 114 (15) | 113 (15) | 79–140 |
| Reason for reduction | | | |
| Because of diarrhea and/or rash/acne AE | 323 (71) | 328 (74) | 279–365 |
| Because of other AE | 135 (29) | 115 (26) | 93–138 |

AE, adverse event; CTCAE, common terminology criteria for adverse events.
[a]At a cutoff of 65 weeks.

Concerning the more clinically oriented model evaluation criteria, the observed within-individual maximum AE severity grades for both diarrhea and rash/acne were within the range of the simulated AE severity grades (**Table 3**). The shapes of the simulated AE severity profiles over time resembled the observed pattern: after the start of treatment, there was a steep increase in diarrhea of all grades, with a lower proportion of grade 2/3 diarrhea observed at later times. Rash/acne evolved more slowly and persisted longer, with a considerable proportion of patients having grades 1 or 2 rash/acne throughout treatment (**Figure 2**).

The simulations predicted a higher-than-observed frequency of diarrhea and/or rash/acne AEs that required treatment interruption and subsequently resulted in dose reductions and discontinuations (**Table 3**). Although the total number of simulated dose reductions (mean 442, range 384–493) was consistent with the observed number of dose reductions (458), the proportion of dose reductions attributed to diarrhea and/or rash/acne tended to be overpredicted. Furthermore, the number of patients who discontinued because of prolonged and/or repeated diarrhea and/or rash/acne AEs was overpredicted (observed, 8.3%

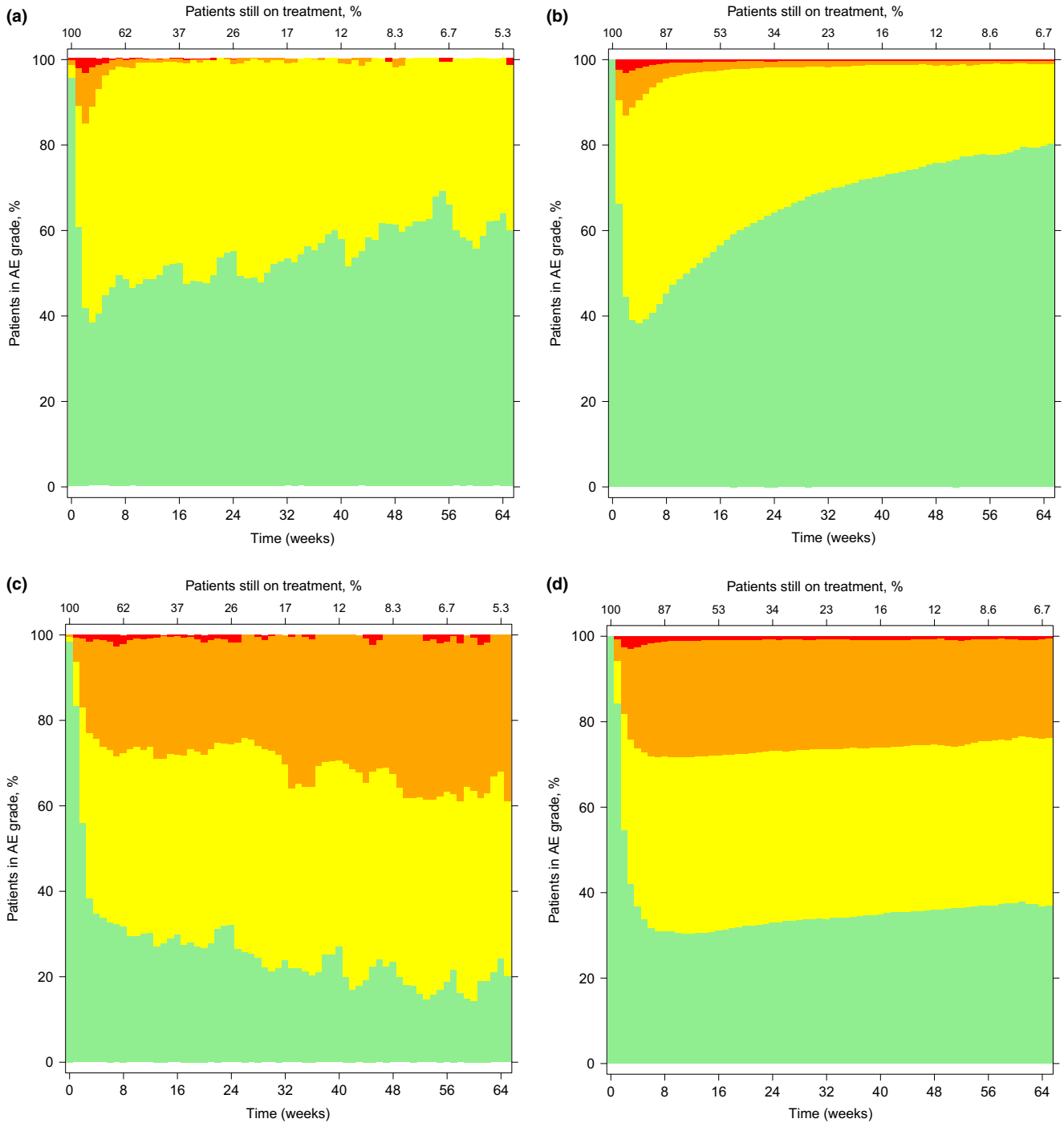


**Figure 2** Time course of (**a**) = observerd diarrhea (**b**) = simulated diarrhea (**c**) = observed rash/acne (**d**) = simulated rash/acne severity for model-building data set and internal evaluation: proportion of patients with no adverse event (green) or concurrent adverse event of grade 1 (yellow), grade 2 (orange), and grade 3 (red). Proportion of patients still on treatment indicated on top *x*-axis. AE, adverse event.

of patients; mean (range) of simulations, 17% (14–21%)), in particular for patients on the starting dose level (observed, 4.3%; mean (range) of simulations, 10% (7.4–13%)), i.e., with no previous dose reduction. As a consequence, the proportion of patients discontinuing because of disease progression or as a result of other reasons was underpredicted, as was the proportion of patients who were predicted to still be on treatment at the end of the simulation period.

In addition, a technical evaluation of the frequencies of transitions between different AE stages was performed (**Table S3**, **Figure S4**). For the majority of transitions, the mean frequencies of the simulations were well in accordance with the observed transition frequencies. A notable

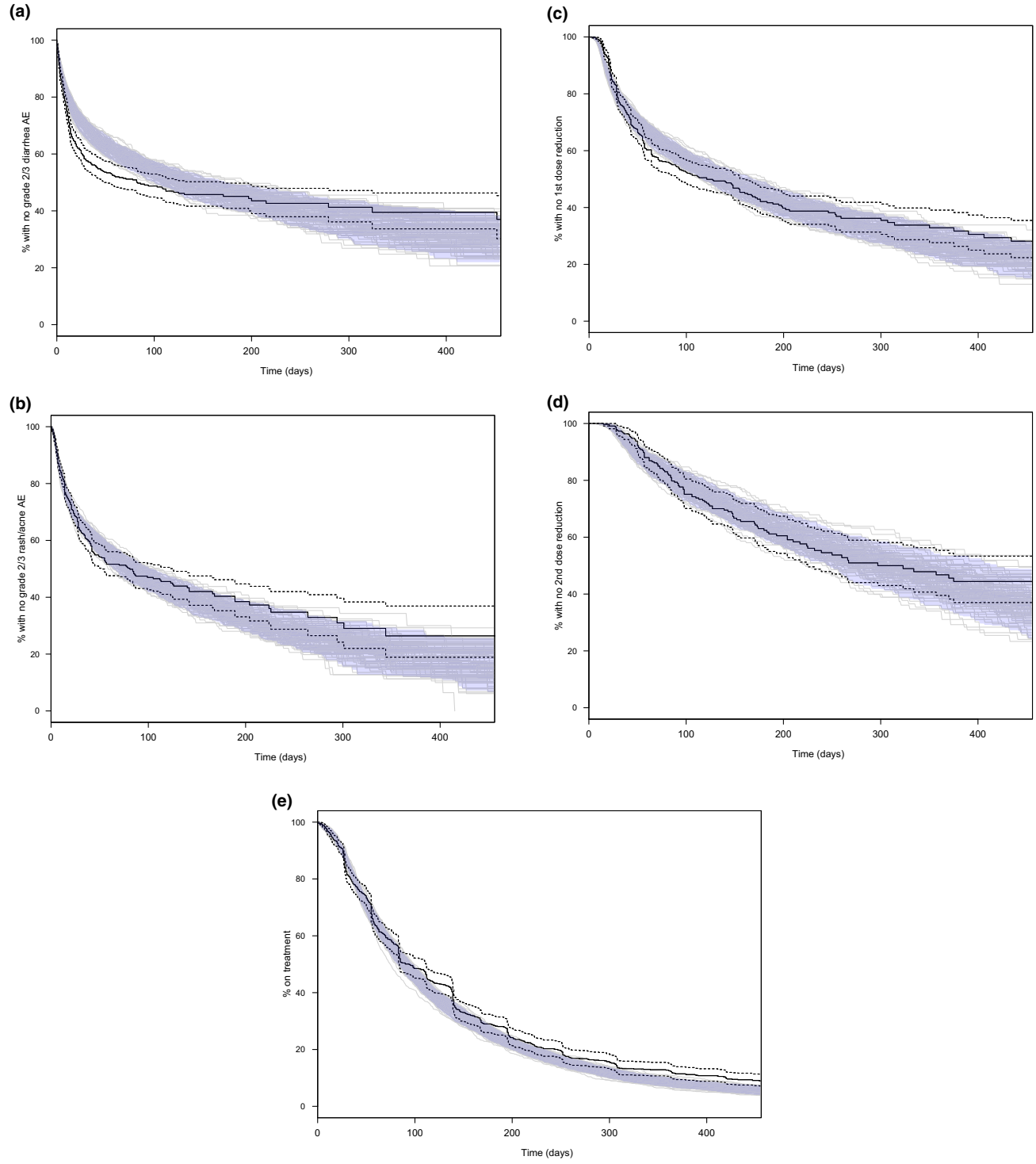


**Figure 3** Kaplan–Meier-type visual predictive checks of first diarrhea (**a**) and first rash/acne grade 2/3 adverse events (**b**; both censored at time of first dose reduction/discontinuation) and first (**c**) and second (**d**) dose reduction and discontinuation (**e**): observed Kaplan–Meier curve (black solid line) with standard errors (black dashed lines), superimposed with simulated Kaplan–Meier curves (gray lines); shaded area corresponds to 95% confidence interval of the simulations. AE, adverse event.

exception concerned the probability to still have grade 2 diarrhea on the following day during off-treatment periods, 83% of the observed transitions vs. 87% in average for the simulations. Furthermore, as the simulations were driven by the protocol stipulation that treatment should be interrupted in case a patient experienced a grade 3 AE, the simulations did not provide any transitions from grade 3 diarrhea or rash/acne during on-treatment periods.

**Table 4 External model evaluation: maximum adverse event grades, discontinuations, and dose reductions**

| | Observed | Simulated | |
|---|---|---|---|
| | Count (%) | Mean (%) | Range |
| Max CTCAE grades | | | |
| Diarrhea | | | |
| No (grade 0) | 10 (4.4) | 31.1 (14) | 17–42 |
| Grade 1 | 98 (43) | 76.3 (33) | 59–92 |
| Grade 2 | 86 (38) | 73.7 (32) | 54–90 |
| Grade 3 | 35 (15) | 47.9 (21) | 32–66 |
| Rash/acne | | | |
| No (grade 0) | 23 (10) | 18.9 (8.3) | 8–29 |
| Grade 1 | 68 (30) | 52.4 (23) | 39–72 |
| Grade 2 | 101 (44) | 103 (45) | 78–123 |
| Grade 3 | 37 (16) | 55.1 (24) | 38–69 |
| Discontinuation[a] | | | |
| Progression | 150 (66) | 119 (52) | 98–135 |
| Other reasons | 26 (11) | 26.8 (12) | 16–36 |
| Diarrhea and/or rash/acne AE | 6 (2.6) | 46.8 (20) | 32–59 |
| Censored | 47 (21) | 36.0 (16) | 23–48 |
| Dose escalation | 16 (7.0) | 15.7 (6.9) | 9–23 |
| Dose reductions | | | |
| No reduction | 95 (41) | 86.6 (38) | 71–105 |
| Reduction | 134 (59) | 142 (62) | 124–158 |
| 1 reduction | 87 (38) | 71.8 (31) | 56–90 |
| 2 reductions | 47 (21) | 70.7 (31) | 56–92 |
| Reason for reduction | | | |
| Because of diarrhea and/or rash/acne AE | 95 (52) | 165 (77) | 138–186 |
| Because of other AE | 86 (48) | 48.7 (23) | 31–66 |

AE, adverse event; CTCAE, common terminology criteria for adverse events.
[a]At a cutoff of 105 weeks.

KM-VPCs (**Figure 3**) of first grade 2/3 diarrhea and rash/acne AEs revealed a mismatch between the observed and model-predicted shape of the hazard function, especially with regard to diarrhea. Concerning first and second dose reductions, the observed Kaplan–Meier curve was generally contained within the prediction intervals, with the exception of an overprediction of first dose reductions occurring at a late stage of treatment. The KM-VPC for discontinuation visualizes the overprediction in the number of patients who discontinued and further illustrated that the clinical trial simulations (CTS) did not account for increased discontinuation rates at specified visits.

**External model evaluation.** To complement model evaluation, a prediction of the study not included in the model building subset (G)[3] was performed. Concerning the trial execution models, the parameter estimates from the study most similar with regard to studied indication and inclusion criteria were used, i.e., study C.[2] The dosing regimen in the study to be predicted started on 40 mg and allowed a single dose escalation step (see **Table 2**), which was different from the studies included in the model-building data set. Hence, an empirical trial execution model was set up according to which 10% (= observed proportion) of the patients fulfilling the criteria of no or only mild toxicity during the initial 21 days of treatment were dose escalated. Simulations were performed for a 105-week period corresponding to the maximum treatment time of 80% of the patients included in the study to be predicted.

The external model evaluation focused on those criteria with clinical relevance. In principle, the outcome of the evaluation followed the same trends as for the internal evaluation (see **Table 4**). Because the proportion of patients who discontinued because of diarrhea and/or rash/acne AEs was very low in the study to be predicted, the deviation between the observation and simulation results were even more pronounced with regard to dose reductions and discontinuation.

## DISCUSSION

The current analysis established afatinib dose-exposure-response relationships, taking into account both the dynamics and the categorical nature of the most frequently observed AEs associated with afatinib treatment, i.e., diarrhea and rash/acne. The chosen modeling framework consisting of longitudinal logistic regression with first-order Markov elements accompanied by trial execution models allowed full CTS, i.e., the prediction of individual AE time courses and their consequences on treatment decisions (dose reduction, treatment interruption, discontinuation).

Higher afatinib exposure increased the probability to experience diarrhea and rash/acne AEs up to a maximum effect. The drug effect on both AEs inversely depended on the current AE grade: the contribution of drug effect to the odds for an increased AE were higher if a patient was in “no AE” state, and lowest in case a patient already had diarrhea or rash/acne. This finding might reflect the effectiveness of the selected AE management strategies foreseen in clinical study protocols. As the vast majority of patients started treatment with no AE, the contribution of drug effect on the transitions from “no AE” to AEs of any grade was estimated to be higher than on transitions between different AE grades.

Further characteristics of the drug effects were identified. For diarrhea, there were no indications of a delay in the drug effect, which is consistent with the reported early onset of diarrhea following afatinib treatment.[4] In contrast, the effect on the onset of rash/acne was delayed, with the wash-out of the effect during off-treatment periods occurring at a faster rate than the onset, potentially reflecting successful comedication. The development of apparent tolerance significantly improved both exposure-response models. For diarrhea, tolerance driven by exposure was inferior to the final model, with tolerance driven by the cumulative days on treatment, suggesting a contribution of localized organ-specific effects not correlated to systemic exposure. However, the development of the tolerance models was data driven, which limits the interpretation of the mechanisms. Furthermore, no covariate analysis was performed (e.g., pretreatment with tyrosin kinase inhibitors, which may have an influence on

tolerance development) as the focus was on the methodological aspects.

The combined modeling framework of exposure-response and trial execution models was able to predict reasonably well maximum AE grades occurring during the course of the studies in the model building as well as the model evaluation subset. However, CTS also showed that AE-guided treatment decisions could not be predicted correctly; in particular, too many patients in the simulations had prolonged AE episodes, i.e., no recovery within a 14-day off-treatment period. Misspecification of apparent tolerance was most visible by the overestimation of patients who discontinued because of diarrhea or rash/acne AEs after two previous dose reductions as well as in the KM-VPCs of first grade 2/3 diarrhea episodes and first dose reductions.

There are several potential reasons that could explain the disparity between observed and predicted data. First, the developed model could from the structure be inadequate to predict longitudinal AE time courses. As stated before, however, others have successfully applied models with a similar structure to characterize the time course of graded AEs in relation to drug exposure.[11–14] In none of these examples, though, were the developed models evaluated based on a comparison between predicted and observed AE-guided treatment decisions, which is in contrast to the current analysis.

Logistic regression models are empirical in nature and offer limited understanding of the occurrence of AEs only, but mechanistic models are lacking. When the simulations resulted in too many patients with long-persisting grade 2 diarrhea eventually resulting in discontinuation according to protocol stipulations, the question on the adequacy of a first-order Markov model arose. In such a model, the prediction for the next timepoint solely depends on the current state, essentially disregarding previous history. Although parts of the current model, including tolerance as well as delays in appearance and resolution of AEs, contain information on such history, parameters accounting for the stage time during off-treatment periods could, however, not be informed based on the available data.

Another cause of bias presumably originated from the fact that the observed data exhibited implausible AE trajectories resulting from imputation, recording errors, and/or deviations from per protocol treatment. The current analysis required accurate recording of daily AE severity, both for periods of progression toward higher AE grades as well as recovery toward no AE grade. However, AE grading in the original clinical studies was done in retrospect, without patient diaries or similar recording aids. Although this was appropriate for the original study objectives, it might have limited applicability for the chosen modeling framework. Furthermore, in an oncology setting, deviations from protocol specifications such as treatment continuation despite AEs qualifying for dose reduction may occur at the discretion of the physician. They are, however, difficult to mirror in simulations that assume treatment decisions to follow a limited number of predefined objective criteria rather than numerous undefined or hidden rules (e.g., experience of investigators), perfect adherence to protocol specifications, and accurate recording. Censoring according to defined criteria was performed to handle both causes of potential bias but proved to be a too simplistic solution. Clearly censoring did not occur at random, i.e., was only required in case patients experienced grade 2/3 AEs.

Additional limitations of this study include that the required trial execution models for competing risks were not optimized, for instance, as the dropout model did not account for increased discontinuation rates following scheduled visits with imaging assessment for disease progression and that no dedicated exposure-efficacy model was developed. The effects of covariates other than exposure and time were not investigated. No data on comedication were available; therefore, comedication effects could be not assessed explicitly within this analysis. Furthermore, this analysis was based on pooled data across multiple studies. The differences between indications or between studied populations were not investigated.

In conclusion, this study characterized the relationships between afatinib exposure and diarrhea and rash/acne AEs. The inherent capability of the modeling framework to predict individual AE time courses and their consequences on treatment decisions (dose reduction, treatment interruption, discontinuation) allows, in principle, prospective comparisons of dosing strategies and study designs with respect to safety. It should work best if there were a limited number of objective prespecified rules that guide treatment decisions. As the authors believe in the importance of quantitative benefit–risk assessment, further studies are encouraged to investigate modeling approaches to characterize exposure-response relationships and enable CTS for the prospective comparisons of dosing strategies.

**Supporting Information.** Supplementary information accompanies this paper on the *CPT: Pharmacometrics & Systems Pharmacology* website (www.psp-journal.com).

**Figure S1.** Kaplan–Meier-type visual predictive checks for trial execution models.
**Figure S2.** Time course of observed diarrhea and rash/acne severity, for model building data set with and without censoring.
**Figure S3.** Time course of observed and simulated diarrhea and rash/acne severity, for model evaluation data set.
**Figure S4.** Observed and simulated transitions between diarrhea and rash/acne adverse event grades, transition box-and-whisker plots.
**Table S1.** Final parameter estimates, exposure-adverse event models.
**Table S2.** Final parameter estimates, trial execution models.
**Table S3.** Observed and simulated transitions between diarrhea and rash/acne adverse event grades, transition matrices.
**Supplementary Material S1.** Pseudo-code parameterization of the drug effect models for diarrhea and rash/acne.
**Data S1.** Mock data set for exposure-adverse event analysis.
**Data S2.** NONMEM model code for exposure-adverse event analysis.

**Acknowledgments.** Boehringer Ingelheim was responsible for the design and conduct of all clinical trials that provided data for this analysis and the collection and management of the data. Igor Varvodic is gratefully acknowledged for generation of the NONMEM data sets for this analysis.

**Funding.** This analysis was sponsored by Boehringer Ingelheim Pharma GmbH & Co. KG.

**Conflict of Interest.** Mats Karlsson has obtained consultancy fees from Boehringer Ingelheim Pharma GmbH & Co. KG. Ronald Niebecker, Hugo Maas, Alexander Staab, and Matthias Freiwald are employees of Boehringer Ingelheim Pharma GmbH & Co. KG.

**Author Contributions.** R.N., M.F., and M.O.K. wrote the manuscript; R.N., H.M., A.S., M.F., and M.O.K. designed the research; R.N., M.F., and M.O.K. performed the research; R.N., H.M., M.F., and M.O.K. analyzed the data.

1. Dungo, R.T. & Keating, G.M. Afatinib: first global approval. *Drugs* **73**, 1503–1515 (2013).
2. Yang, J.C. *et al.* Afatinib for patients with lung adenocarcinoma and epidermal growth factor receptor mutations (LUX-Lung 2): a phase 2 trial. *Lancet Oncol.* **13**, 539–548 (2012).
3. Sequist, L.V. *et al.* Phase III study of afatinib or cisplatin plus pemetrexed in patients with metastatic lung adenocarcinoma with EGFR mutations. *J. Clin. Oncol.* **31**, 3327–3334 (2013).
4. Yang, J.C. *et al.* Influence of dose adjustment on afatinib safety and efficacy in patients with advanced EGFR mutation-positive NSCLC. *Am. Soc. Clin. Oncol.* **27**, 2103–2110 (2016).
5. Okamoto, I. *et al.* Influence of dose adjustment on afatinib safety and efficacy in patients with advanced EGFR mutation-positive NSCLC. *Japan. Soc. Med. Oncol.* oral presentation at the 13th Annual Meeting of the Japanese Society of Medical Oncology (JSMO) in Sapporo, 16–18 July 2015.
6. Boudou-Rouquette, P. *et al.* Variability of sorafenib toxicity and exposure over time: a pharmacokinetic/pharmacodynamic analysis. *Oncologist* **17**, 1204–1212 (2012).
7. Fukudo, M. *et al.* Population pharmacokinetics/pharmacodynamics of erlotinib and pharmacogenomic analysis of plasma and cerebrospinal fluid drug concentrations in Japanese patients with non-small cell lung cancer. *Clin. Pharmacokinet.* **52**, 593–609 (2013).
8. Guilhot, F. *et al.* Plasma exposure of imatinib and its correlation with clinical response in the Tyrosine Kinase Inhibitor Optimization and Selectivity Trial. *Haematologica* **97**, 731–738 (2012).
9. Wind, S., Schmid, M., Erhardt, J., Goeldner, R.G. & Stopfer, P. Pharmacokinetics of afatinib, a selective irreversible ErbB family blocker, in patients with advanced solid tumors. *Clin. Pharmacokinet.* **52**, 1101–1109 (2013).
10. Zhao, Y.Y. *et al.* The relationship between drug exposure and clinical outcomes of non-small cell lung cancer patients treated with gefitinib. *Med. Oncol.* **28**, 697–702 (2011).
11. Zingmark, P.H., Kågedal, M. & Karlsson, M.O. Modelling a spontaneously reported side effect by use of a Markov mixed-effects model. *J. Pharmakokinet. Pharmacodyn.* **32**, 261–281 (2005).
12. Hénin, E. *et al.* A dynamic model of hand-and-foot syndrome in patients receiving capecitabine. *Clin. Pharmacol. Ther.* **85**, 418–425 (2009).
13. Keizer, R.J. *et al.* A model of hypertension and proteinuria in cancer patients treated with the anti-angiogenic drug E7080. *J. Pharmakokinet. Pharmacodyn.* **37**, 347–363 (2010).
14. Hansson, E.K. *et al.* PKPD modeling of predictors for adverse effects and overall survival in sunitinib-treated patients with GIST. *CPT Pharmacometrics Syst. Pharmacol.* **2**, e85 (2013).
15. Lin, N.U. *et al.* A phase II study of afatinib (BIBW 2992), an irreversible ErbB family blocker, in patients with HER2-positive metastatic breast cancer progressing after trastuzumab. *Breast Cancer Res. Treat.* **133**, 1057–1065 (2012).
16. Schuler, M. *et al.* A phase II trial to assess efficacy and safety of afatinib in extensively pretreated patients with HER2-negative metastatic breast cancer. *Breast Cancer Res. Treat.* **134**, 1149–1159 (2012).
17. Miller, V.A. *et al.* Afatinib versus placebo for patients with advanced, metastatic non-small-cell lung cancer after failure of erlotinib, gefitinib, or both, and one or two lines of chemotherapy (LUX-Lung 1): a phase 2b/3 randomised trial. *Lancet Oncol.* **13**, 528–538 (2012).
18. Katakami, N. *et al.* LUX-Lung 4: a phase II trial of afatinib in patients with advanced non-small-cell lung cancer who progressed during prior treatment with erlotinib, gefitinib, or both. *J. Clin. Oncol.* **31**, 3335–3341 (2013).
19. Seiwert, T.Y. *et al.* A randomized, phase II study of afatinib versus cetuximab in metastatic or recurrent squamous cell carcinoma of the head and neck. *Ann. Oncol.* **25**, 1813–1820 (2014).
20. Beal, S., Sheiner, L.B., Boeckmann, A. & Bauer, R.J. *NONMEM User's Guides (1989–2009)* (Icon Development Solutions, Ellicott City, MD, 2009).
21. Keizer, R.J., Karlsson, M.O. & Hooker, A. Modeling and simulation workbench for NONMEM: tutorial on Pirana, PsN and Xpose. *CPT Pharmacometrics Syst. Pharmacol.* **2**, e50 (2013).
22. R Core Team. R: A language and environment for statistical computing. <http://www.R-project.org> (2014).
23. Cancer Therapy Evaluation Program, National Cancer Institute(NCI) Division of Cancer Treatment and Diagnosis. Common terminology criteria for adverse events, version 3.0. <http://ctep.cancer.gov> (2003).
24. Freiwald, M., Schmid, U., Fleury, A., Wind, S., Stopfer, P. & Staab, A. Population pharmacokinetics of afatinib, an irreversible ErbB family blocker, in patients with various solid tumors. *Cancer Chemother. Pharmacol.* **73**, 759–770 (2014).