Journal of Pharmacokinetics and Pharmacodynamics (2019) 46:473–484
https://doi.org/10.1007/s10928-019-09653-4

ORIGINAL PAPER

Modeling and simulation of the modified Rankin Scale and National Institutes of Health Stroke Scale neurological endpoints in intracerebral hemorrhage

Rik Schoemaker[1] · Satyaprakash Nayak[2] · Lutz O. Harnisch[3] · Mats O. Karlsson[4] on behalf of VISTA-ICH Collaboration

Received: 8 January 2019 / Accepted: 18 August 2019 / Published online: 29 August 2019

Abstract

Intracerebral hemorrhage (ICH) is a form of stroke characterized by uncontrolled bleeding into the parenchyma of the brain. There is no approved therapy for ICH and it is associated with very poor neurological outcomes with around half of subjects dying within 1 month and most subjects showing complete or partial disability. A key challenge is to identify subjects who could benefit from intervention using characteristics such as baseline hemorrhage volume and the increase in hemorrhage volume in the first few hours, which have been correlated with final outcomes in ICH. Combined longitudinal models were developed to describe stroke scales using categorical data (Modified Rankin Scale, mRS), continuous bounded data (National Institutes of Health Stroke Scale, NIHSS), and time to death. Covariate effects for baseline hematoma volume and maximum increase in hematoma volume were incorporated to assess the improvement in outcome when hematoma volume increase would be reduced by a potential treatment. The combined model provided an adequate description of stroke scales, with patients split into a Non-survival and a High-survival sub-population, and dropout due to death was well described by a constant hazard survival model. Models were compared indicating that the combined mRS/NIHSS model provided the most information, followed by the NIHSS-only model, and the mRS-only model, and finally the traditional statistical analysis on dichotomized response at 90 days. Simulations showed that substantial reductions in hematoma volume increase were required to increase the probability of a favorable outcome.

Keywords Mixed effect models · Model based drug development · Disease progression modeling · Pharmacometrics · Neuroscience

Introduction

Spontaneous intracerebral hemorrhage (ICH) is a major cause of death and disability accounting for approximately 10–15% of stroke cases worldwide [1, 2]. It is associated with poor clinical outcomes; approximately 35–52% of patients die within 1 month, and many patients remain functionally dependent after 6 months [1]. The current management of ICH is largely supportive, and the lack of effective treatment for acute and severe bleeds remains a critical and unmet need. Current attempts at a pharmacological intervention in ICH focus on reducing the increase in hematoma volume after the hemorrhage through compounds that induce blood coagulation in the brain, an effect likely induced by recombinant activated factor VII (rFVIIa) [3].

The VISTA-ICH Steering Committee are listed in "Acknowledgements".

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10928-019-09653-4) contains supplementary material, which is available to authorized users.

✉ Lutz O. Harnisch
Lutz.O.Harnisch@pfizer.com

[1] Occams, Amstelveen, The Netherlands

[2] Pfizer Inc., Cambridge, MA, USA

[3] Pfizer ltd., Sandwich, UK

[4] Pharmetheus, Uppsala, Sweden

We aimed to aid the design and analysis of ICH trials by developing models to describe the time course of two commonly assessed neurological endpoints in stroke trials: the modified Rankin Scale (mRS) and the National Institutes of Health Stroke Scale (NIHSS) scores. Trial outcome is usually measured at day 90 from the day of stroke onset. One of the most important challenges in the design of an ICH trial is to identify subjects which could benefit from a pharmacological intervention aimed at reducing the hemorrhage volume. Indeed, the results from the post hoc analysis of the Phase 3 rFVIIa study showed that subjects with age ≤ 70 years, hemorrhage volume < 60 mL and time from onset-to-treatment ≤ 2.5 h showed improved effect. Change in the hemorrhage volume at 24 h has also been identified to have an important predictive effect on the overall outcome [3]. A combined mRS/NIHSS model was constructed using the placebo data from the Virtual International Stroke Trials Archive-Intracerebral Hemorrhage (VISTA-ICH) database [4], incorporating dropout due to death information. The core concept regarding the mechanism of action of a candidate drug is that an increase in ICH hematoma volume (IHV) is associated with a poor outcome, and that reducing this increase is expected to prevent further deterioration of the stroke outcome. Consequently, the change in IHV is considered the driver of the longitudinal NIHSS and mRS models, under the assumption that the difference in outcome between placebo patients with a large IHV increase and patients with a small IHV increase provides an indication of the potential gain due to pharmacological intervention. The aim of this analysis it to help design an ICH study by identifying characteristics of subjects which are most susceptible to worsening in the absence of drug, and thus most likely to benefit from therapeutic intervention by modeling the neurological outcomes of interest (mRS and NIHSS) using characteristics such as baseline IHV and change in IHV over the first 24 h. In addition, we have also linked the expected outcomes by modeling the decrease in IHV and its effect on the neurological scores at 90 days.

Methods

We accessed anonymized data on 687 placebo-treated subjects from five trials with ICH. In addition to demographic variables, the database contains IHV information, clinical scales like NIHSS and mRS, and time of death. Assessments of IHV were available in the VISTA database at baseline, 1 and 3 days. Assessments of mRS were available at either 7, 30, and 90 days, or at 15 and 90 days, and NIHSS assessments were available at either baseline, 1 h, and 1, 3, 15 and 90 days, or at baseline, 1, 3, 7 and 90 days.

Initially, separate models were developed to provide a description of the NIHSS and mRS time profiles, incorporating dropout due to death. These two models were ultimately merged into a single model. The full code is provided in the supplementary material.

NIHSS model

The NIHSS scale is a 0–42 point neurological function categorical scale [5, 6] where 0 corresponds to the absence of symptoms and 42 to the most severe disability. Model development started by examining individual NIHSS curves over time suggesting that there were three broad categories (Supplementary Fig. S-1):

- patients that showed an increase (deterioration) in their NIHSS scores and then died
- patients that showed an initial increase in NIHSS scores followed by a decrease (improvement)
- patients that showed a decrease (improvement) in their NIHSS scores

Several models were evaluated to describe these profiles, and a stable description was provided using a mixture for two populations with corresponding separate equations:

- Non-survival: patients that showed an increase in their NIHSS scores and then died
- High-survival: patients that were associated with a much larger probability of survival, where both initial increase and initial decrease in NIHSS scores was possible

The proportion of patients belonging to either population was determined by the estimation algorithm.

To ensure that NIHSS scores were restricted within their bounds, a logit transform was used as part of the model. Modeling was performed using a continuous data-type analysis. The link between the modeled logit values and the observed results on the original scale was provided by:

$$NIHSS_{i,j} = \frac{42.4999 \cdot e^{Logit(NIHSS_{i,j}) + \varepsilon_{NIHSS_{i,j}}} - 0.4999}{1 + e^{Logit(NIHSS_{i,j}) + \varepsilon_{NIHSS_{i,j}}}} \quad (1)$$

where $NIHSS_{i,j}$ provides the observed NIHSS value for subject i at time j and $\varepsilon_{NIHSSi,j}$ is the additive residual error on the logit scale. Scaling values of 42.4999 and $-$ 0.4999 are used to allow predicting NIHSS values that, once rounded, will allow representations of either 42 or 0.

The time profiles for the Non-survival group were described by a linear change over time on the logit scale, where the individual slopes of the relationship were all positive due to the exponential power parametrization:

$$Slope_{NIHSS\,Non-survival,i} = Slope_{NIHSS\,Non-survival} \cdot e^{\eta Slope_{NIHSS-NS,i}} \quad (2)$$

$$Logit\left(NIHSS_{Non-survival\,i,j}\right) = Base_{NIHSS} + \eta Base_{NIHSS,i} + Slope_{NIHSS\,Non-survival,i} \cdot Time_j \quad (3)$$

where $Slope_{NIHSS\ Non\text{-}survival}$ is the typical slope for the Non-survival population, $\eta Slope_{NIHSS\text{-}NS,i}$ describes the inter-individual variability (IIV) contribution for subject i, and $Slope_{NIHSS\ Non\text{-}survival,i}$ is the individual slope over time. $Base_{NIHSS}$ is the typical baseline value for NIHSS on the logit scale, $\eta Base_{NIHSS,i}$ is the associated additive IIV, and $Logit(NIHSS_{Non\text{-}survival\ i,j})$ provides the NIHSS model predictions in the Non-survival population.

The time profiles for the High-survival group were described by the sum of two time-varying functions: (i) a linear change over time on the logit scale and (ii) a non-linear relationship with time approaching a constant value (Emax) as time increases. Neither the linear slope nor the non-linear Emax were constrained for sign, allowing both increases and decreases in NIHSS with time:

$$Slope_{NIHSS\,High-survival,i} = Slope_{NIHSS\,High-survival} + \eta Slope_{NIHSS-HS,i} \quad (4)$$

$$Emax_{NIHSS,i} = Emax_{NIHSS} + \eta Emax_{NIHSS,i} \quad (5)$$

$$Logit\left(NIHSS_{High-survival\,i,j}\right) = Base_{NIHSS} + \eta Base_{NIHSS,i} + Slope_{NIHSS\,High-survival,i} \cdot Time_j + \frac{Time_j \cdot Emax_{NIHSS,i}}{Time_j + ET50_{NIHSS}} \quad (6)$$

where $Slope_{NIHSS\ High\text{-}survival}$ is the typical slope for the High-survival population, $\eta Slope_{NIHSS\text{-}HS,i}$ is the IIV, and $Slope_{NIHSS\ High\text{-}survival,i}$ is the individual slope over time. $Emax_{NIHSS}$ is the typical maximum change for the High-survival population, $\eta Emax_{NIHSS,i}$ is the IIV, and $Emax_{NIHSS,i}$ is the individual Emax. $ET50_{NIHSS}$ provides the time point at which 50% of the maximum change occurred. $Logit(NIHSS_{High\text{-}survival\ i,j})$ provides the NIHSS model prediction in the High-survival population.

Finally, a single parameter ($D90_{NIHSS\ High\text{-}survival}$) was used to quantify the change to the score at day 90 by reparametrizing $Slope_{NIHSS\ High\text{-}survival,i}$:

$$Slope_{NIHSS\,High-survival,i} = \left(D90_{NIHSS\,High-survival} + \eta D90_{NIHSS,i} - \frac{90 * Emax_{NIHSS,i}}{90 + ET50_{NIHSS}}\right)/90 \quad (7)$$

where $D90_{NIHSS\ High\text{-}survival}$ is the population average logit change from Baseline to day 90 for the High-survival population, and $\eta D90_{NIHSS,i}$ is the IIV. Illustrations are provided in Supplementary Fig. S-2.

Dropout due to death model

Individual time to death outcomes were recorded in the database, and dropout due to death was modeled using a constant hazard survival model, where the probability of survival is given by:

$$S(Time_j) = e^{-BHAZ \cdot Time_j} \quad (8)$$

The hazard rate BHAZ has separate values for the High-survival and the Non-survival population.

mRS model

The mRS scale [7, 8] is a seven point scale with integer values ranging from 0 to 6, associated with increasing severity of symptoms, with death covered by the mRS of 6. Since death is already covered by the dropout due to death model component, mRS scores of 6 in the database were recoded to missing during analysis. The remaining mRS ordered categorical data outcomes (0–5) were described using the proportional odds model [9], where the population probabilities over time were modeled using cumulative scores for mRS = 0, mRS $\leq$ 1, mRS $\leq$ 2, mRS $\leq$ 3, and mRS $\leq$ 4. The probabilities (P) were transformed to logits which changes the scale from 0 *to* 1 to $-\infty$ *to* ∞:

$$Logit(x) = \log\left(\frac{P(x)}{1 - P(x)}\right) \quad (9)$$

The underlying time profile for the mRS scores was described using a simplified version of the NIHSS continuous time-dependent equation because mRS profiles provided less information than NIHSS profiles.

The time profiles for the Non-survival group were described by a linear change over time on the logit scale; individual slopes of the relationship were constrained to be negative due to the exponential parametrization:

$$Slope_{mRS\,Non-survival,i} = Slope_{mRS} \cdot e^{\eta Slope_{mRS-NS,i}} \quad (10)$$

$$Logit\left(mRS_{0,Non-survival\,i,j}\right) = Base_{mRS} + \eta Base_{mRS,i} + Slope_{mRS\,Non-survival,i} \cdot Time_j \quad (11)$$

where $Slope_{mRS\ Non\text{-}survival}$ is the typical slope for the Non-survival population, $\eta Slope_{mRS\text{-}NS,i}$ is the IIV, and $Slope_{mRS\ Non\text{-}survival,i}$ is the individual slope over time. $Base_{mRS}$ is the typical baseline value for P(mRS = 0) on the logit scale, $\eta Base_{mRS,i}$ is the IIV, and $Logit(mRS_{0,Non\text{-}survival\ i,j})$ provides the model prediction for mRS = 0 in the non-survival population. The standard proportional odds model assumes that the scores on the logit scale are

parallel. The score that describes the probability over time of an mRS score of zero (the best possible outcome) should go down over time for the Non-survival patients, and the higher scores are parallel to this profile and hence go down as well.

The time profiles for the High-survival group only supported the Emax model component to describe change over time:

$$Emax_{mRS,i} = Emax_{mRS} \cdot e^{\eta Emax_{mRS,i}} \quad (12)$$

$$Logit\left(mRS_{0,High-survival\,i,j}\right) = Base_{mRS} + \eta Base_{mRS,i} + \frac{Time_j \cdot Emax_{mRS,i}}{Time_j + ET50_{mRS}} \quad (13)$$

where $Emax_{mRS}$ is the typical maximum effect for the High-survival population, $\eta Emax_{mRS,i}$ is the IIV, and $Emax_{mRS,i}$ is the individual Emax. Due to the exponential IIV model, individual Emax values can only be positive. $ET50_{mRS}$ provides the time point at which 50% of the maximum change occurred. Finally, $Logit(mRS_{0,High-survival\ i,j})$ provides the model predictions for mRS = 0 in the High-survival population.

The logits for the cumulative mRS scores were described as:

$$Logit\left(mRS_{\leq X,i,j}\right) = Logit\left(mRS_{\leq (X-1),i,j}\right) + BX \quad (14)$$

where X is the cumulative mRS score and BX is the parallel shift for score X. The logits can be back-transformed into cumulative probabilities. Illustrations are provided in Supplementary Fig. S-3.

Combined mRS/NIHSS model

The NIHSS, mRS, and dropout due to death models were combined to allow simultaneous estimation of both mRS and NIHSS profiles. By combining the models it is possible to estimate the correlation between mRS and NIHSS baseline parameters. The combined model also imposed a restriction: assignment to one of the mixture populations was assumed to be consistent within individuals for the three previously independent model components of dropout due to death, mRS outcome, and NIHSS outcome.

Hematoma volume covariate effects

The effect of the baseline hematoma volume (ICH0) was implemented on the baseline parameters for mRS and NIHSS using the following equations:

$$Base_{mRS,i} = Base_{mRS} + \eta Base_{mRS,i} + EHVB_{mRS} * \log(ICH0_i/14.57) \quad (15)$$

$$Base_{NIHSS,i} = Base_{NIHSS} + \eta Base_{NIHSS,i} + EHVB_{NIHSS} * \log(ICH0_i/14.57) \quad (16)$$

where 14.57 is the median observed ICH0, $EHVB_{mRS}$ describes the influence of ICH0 on mRS baseline parameter, and $EHVB_{NIHSS}$ describes the influence of ICH0 on NIHSS baseline parameter.

The effect of the increase in IHV from ICH0 to the maximum post-baseline value (DICHMAX) was implemented on the mixture fraction (P1):

$$Cov_{P1,i} = (DICHMAX_i - 8.64) \cdot DICHEFF \quad (17)$$

$$\Psi = \log\left(\frac{P1}{1 - P1}\right) \quad (18)$$

$$P1_i = \frac{e^{\Psi + Cov_{P1,i}}}{1 + e^{\Psi + Cov_{P1,i}}} \quad (19)$$

where 8.64 is the mean observed DICHMAX, and DICHEFF describes the influence of DICHMAX on P1. The Ψ-step ensures that individual probabilities $P1_i$ remain inside the 0–1 interval.

Software

Analyses were performed using NONMEM (version 7.3.0 [10]) with Laplacian estimation, supplemented with the PsN toolkit, version 4.4.8 [11]. R software (version 3.2.3 [12]) was used for general scripting, data management, goodness of fit analyses, simulation and model evaluation.

Model evaluation

Visual predictive checks (VPCs) were used for model assessment [13]; 1000 studies were simulated with a structure identical to the study analyzed and individual profiles were simulated using the individual parameters sampled from a multivariate normal distribution described using the population estimates.

Model discrimination was based on inspection of graphical diagnostics and changes in the objective function value (OFV) for nested/hierarchical models provided by NONMEM, where a p value of 0.05 was required for a significant model update.

Simulation of covariate influence

The impact of covariates on neurological outcome at day 90 was quantified using simulations. This was implemented using the medians of eight equal-sized quantiles of the ICH0 and DICHMAX values. These 8 by 8 combinations were used to simulate 10,000 subjects each, for the final combined mRS/NIHSS model. Simulated NIHSS and mRS

scores at day 90 were summarized by calculating the mean score for every combination, by counting the fraction of patients with an mRS score < 4 (a clinically important favorable outcome), the fraction of patients with an NIHSS score drop of > 4 points, and the fraction of patients with both an mRS score < 4 and an NIHSS drop of > 4 points.

The covariate model also allowed an investigation of the anticipated impact of reducing the observed IHV increase. Outcomes were simulated under an assumed placebo situation, and subsequent reductions in volume increase in 20% steps up to complete reduction (100%) of the volume increase.

Simulations for relative power assessments

The three longitudinal models (NIHSS-alone, mRS-alone and combined mRS/NIHSS), and the more common statistical approach of analyzing dichotomized mRS data at day 90 (mRS ≤ 4), were compared for two sampling strategies. The aim was to provide an indication of the potential gain of using one method or sampling strategy over another, in assessing treatment effects in ICH studies. Power and sample size calculations are usually intended to design future trials but in the absence of a likely effect size, these outcomes should only be considered as a comparison of the sensitivity to detect effects for different analysis methods and sampling strategies.

The comparisons were implemented by first simulating 10,000 subjects with the final combined mRS/NIHSS model using an enriched sampling schedule with mRS measurements at 7, 15, 30, 45 and 90 days, and NIHSS measurements at 0 and 1 h, and at 1, 3, 7, 15, 30, 45 and 90 days. An arbitrary treatment effect was implemented by increasing P1 by 0.075, and additionally by increasing both $D90_{NIHSS\ High\text{-}survival}$ and $Emax_{mRS}$ by 20%. From these simulated datasets, 500 subsets of 150 patients per arm were sampled and analyzed using the three longitudinal models, parameterized both with and without a treatment effect.

The same procedure was applied for a standard sampling scheme, with mRS measurements at 15 and 90 days, and NIHSS measurements at 0 and 1 h, and at 1, 3, 15, and 90 days (the sampling scheme used in 4 of the 5 VISTA-ICH trials).

The result was a set of 500 differences in OFV (dOFV) between the model with and the model without treatment effect. NONMEM had to report the number of significant digits for both models, and dOFV had to be positive not to include models converged to a local minimum. These dOFVs were then used to estimate the non-centrality parameter for a 2 degrees of freedom Chi square distribution given the two fixed effects parameters for the treatment effect; in the absence of a treatment effect, likelihood theory postulates that dOFVs are distributed according to a Chi square distribution, but in the presence of a treatment effect, the standard Chi square distribution becomes a non-central Chi square distribution and the non-centrality parameter is associated with the size of the treatment effects. The estimated non-central Chi square distribution parameters were subsequently used to obtain power curves for an entire range of sample sizes. This methodology corresponds to the Parametric Power Estimation procedure described in detail by Ueckert et al. [14].

The same 10,000 patient simulated dataset was used for deriving the relative power for different sample sizes using a standard statistical approach; simulated trials with different sample sizes were analyzed using a binary logit [3] model analyzed with binomial logistic regression, where the mRS scores were dichotomized for scores ≤ 4 at day 90. The procedure was repeated 2500 times for each combination of sample size and treatment effect and the number of significant trial outcomes was scored.

Results

The combined mRS/NIHSS model was constructed by linking NIHSS-only and mRS-only models by the joint estimation of the two mixture populations in combination with the dropout due to death survival component; one subpopulation of Non-survival patients with respect to mRS, NIHSS and dropout hazard, the other subpopulation taking on the values of the High-survival population for these endpoints. The correlation between the $Base_{mRS}$ and $Base_{NIHSS}$ was estimated, leading to a dOFV of -293.11 points. Fixing the IIV for $Slope_{mRS\ Non\text{-}survival}$ to zero, led to a non-significant increase of 1.12 ($p = 0.2897$), and this IIV was therefore dropped from the model. As a result, inter-individual differences in mRS profiles were only driven by the $Base_{mRS}$ parameter and mixture population assignment. IIVs for $ET50_{mRS}$ and $ET50_{NIHSS}$ could not be estimated and were fixed to zero.

Incorporation of DICHMAX on P1 led to a dOFV of -460.12 points. Additionally, incorporating log ICH0 on $Base_{mRS}$ and $Base_{NIHSS}$, led to a dOFV of -112.56, resulting in the final combined mRS/NIHSS model.

NONMEM-estimated parameters are provided in Table 1 and the model code is provided in the online

Table 1 NONMEM parameter estimates for the final mRS/NIHSS combined model

Parameter	Estimate (95% CI)[a]	SE (%CV)[b]	IIV[c]
$Base_{NIHSS}$	− 0.778 (− 0.829/− 0.727)	3.3	0.517[d]
$Base_{mRS}$	− 12.7 (− 14.2/− 11.3)	5.8	3.45[d]
$Emax_{NIHSS}$	− 0.458 (− 0.620/− 0.295)	18.1	0.694[e]
$D90_{NIHSS\ High\text{-}survival}$	− 1.29 (− 1.37/− 1.21)	3.1	0.708[e]
$ET50_{NIHSS}$ (days)	1.34 (0.462/2.22)	33.4	
$Emax_{mRS}$	7.63 (6.62/8.63)	6.7	
$ET50_{mRS}$ (days)	41.3 (17.6/64.9)	29.2	
$Slope_{NIHSS\ Non\text{-}survival}$	0.251 (0.232/0.270)	3.9	114%
$Slope_{mRS\ Non\text{-}survival}$	− 0.711 (− 1.06/− 0.365)	24.8	
Residual error $_{NIHSS}$	0.303 (0.300/0.307)	0.6	
B1	3.88 (3.12/4.64)	10.0	
B2	2.0 (1.65/2.34)	8.8	
B3	2.02 (1.71/2.33)	7.8	
B4	4.49 (4.02/4.95)	5.3	
Hazard rate $_{High\text{-}survival}$ (day^{-1})	0.000518 (0.000269/0.000767)	24.5	
Hazard rate $_{Non\text{-}survival}$ (day^{-1})	0.114 (0.0893/0.139)	11.1	
P1	0.871 (0.841/0.902)	1.8	
$EHVB_{NIHSS}$ (log IVH_0 on $Base_{NIHSS}$)	0.424 (0.370/0.478)	6.5	
$EHVB_{mRS}$ (log IVH_0 on $Base_{mRS}$)	− 1.65 (− 2.04/− 1.25)	12.2	
DICHEFF (DICHMAX on P1)	− 0.0629 (− 0.0836/− 0.0423)	16.7	

[a]95%CI is estimate ± 1.96 * the standard error for the estimate

[b]Standard errors of the estimate (SE) are reported as %CV: 100 * (standard error for the estimate)/estimate

[c]IIV is the CV of the IIV calculated as the square root of the diagonal element in the random-effect matrix

[d]Correlation between $\eta Base_{NIHSS}$ and $\eta Base_{mRS}$: − 0.654

[e]Correlation between $\eta Emax_{NIHSS}$ and $\eta D90_{NIHSS}$: 0.593

supplement. Supplementary Table S-1 compares estimates for the three longitudinal models, illustrating that most shared parameters are similar.

Standard non-parametric Kaplan–Meier survival curves can be constructed if the observed time to death is split by the NONMEM-estimated mixture model population. Results along with the predicted survival curves are provided by Fig. 1, illustrating the large difference in probability of survival for the two mixture populations, and the close correspondence between the model curves and the Kaplan–Meier curves. This figure suggests that the NONMEM estimated mixture accurately estimates the probability of survival in the two different groups, and a more complex survival model to describe time to death in these two populations is not required.

A VPC applied to a stacked bar representation of the distribution of mRS scores for the different time points is provided in Supplementary Fig. S-4. The simulated range of trial outcomes closely matches the observed data. The stacked bar graph obscures the actual time profile (due to the use of equally spaced bars at unequally spaced time points) and shows the end result in terms of probabilities or fractions while the actual modeling takes place on the logit scale. An alternative VPC is presented in Fig. 2, where the x-axis uses the actual time scale and the y-axis uses the logit scale with corresponding probabilities as axis labels. This graph illustrates that the model correctly simulates observed cumulative mRS score fractions.

Overlaid individual NIHSS observed time profiles are provided as Supplementary Fig. S-5 along with a loess smooth through the data, compared to the median predicted profile and the 50% range of predicted outcomes, suggesting the data are well described by the model. The VPC for the NIHSS component, where the values associated with dropout due to death are replaced using a last observations carried forward (LOCF) procedure is provided by Fig. 3 illustrating that NIHSS profiles are accurately described by the model as well.

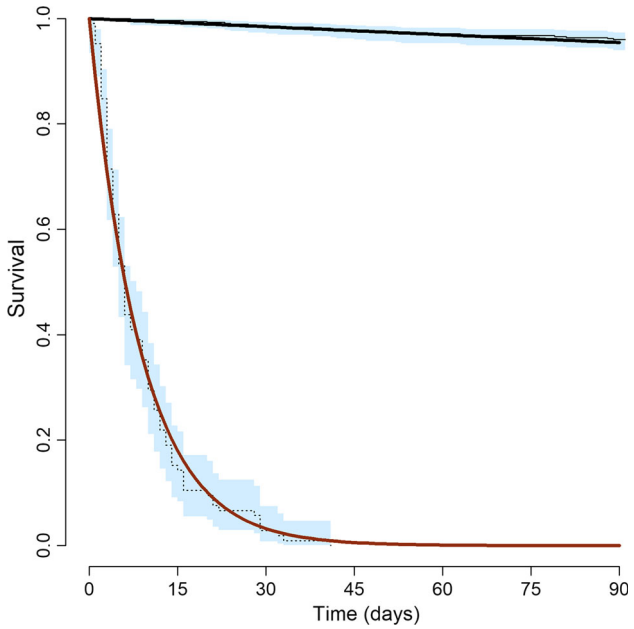

Fig. 1 Kaplan–Meier survival curve for the two mixture populations (solid black line: high survival population, black dashed line: Non-survival population, light-blue area: 95% confidence interval for the Kaplan–Meier curves) and predicted survival profiles using the final combined mRS/NIHSS model (black bold line: High-survival population, red solid line: Non-survival population). Numbers at the bottom of the graph provide subjects at risk (Color figure online)

Simulation of covariate effects

Simulation of mRS and NIHSS scores at day 90 was used to quantify the combined influence of ICH0 and DICHMAX using the median values of 8 equal sized quantiles for each covariate.

Figure 4 provides the probability of an mRS score < 4 where the different lines represent ICH0 categories: an increase in ICH0 was clearly associated with a decreased probability of a favorable outcome. Only the two highest categories for DICHMAX were associated with a steep drop in the probability of obtaining a favorable outcome. This suggests that an intervention is only effective if it could stop any increase in IHV of > 8 mL from happening. An alternative representation is provided as Supplementary Fig. S-6, which also provides the numbers of patients actually observed for the different covariate combination groups, indicating that not all covariate combinations are equally likely in practice. The graphs for an NIHSS drop > 4 (Supplementary Fig. S-7) and for the combined requirements of both an mRS < 4 and and NIHSS drop of > 4 (Supplementary Fig. S-8) suggest a similar conclusion.

The covariate model allows an investigation of the anticipated impact on outcome at day 90 of reducing the observed IHV increase. Table 2 provides the results under

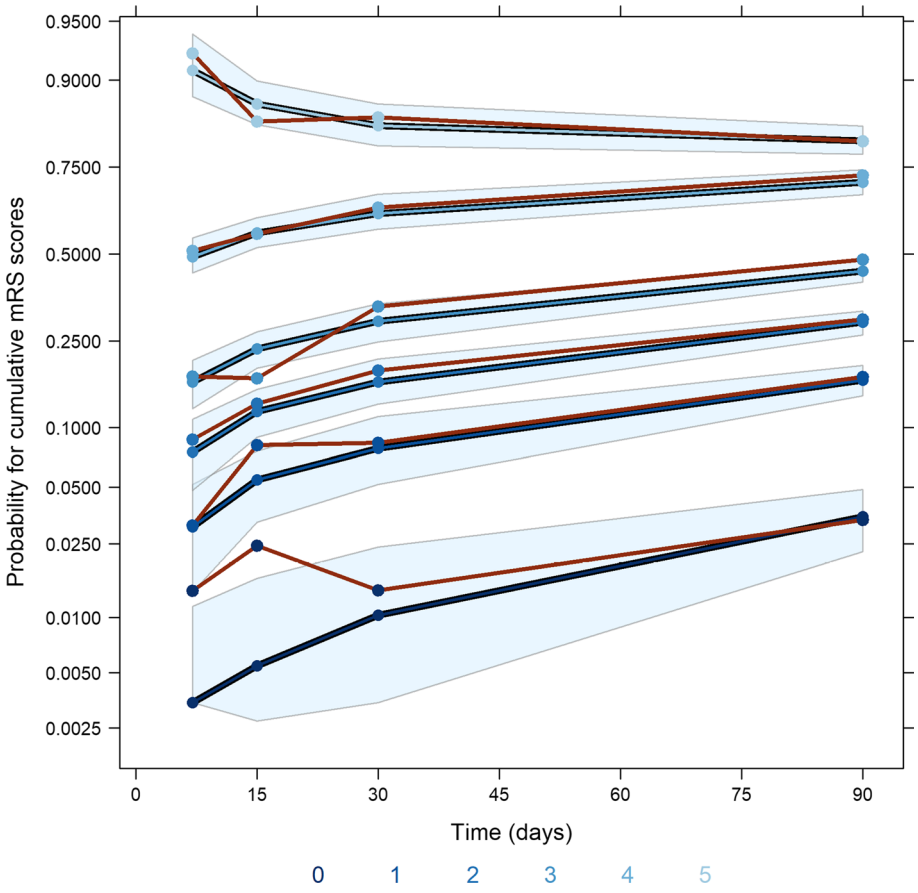

Fig. 2 Visual predictive check for cumulative mRS logits over time for the final combined mRS/NIHSS model. Red lines: observed logits of scores, blue colored lines: median simulated logits across 1000 trials, shaded areas show the 95% prediction interval using the simulated logits (Color figure online)

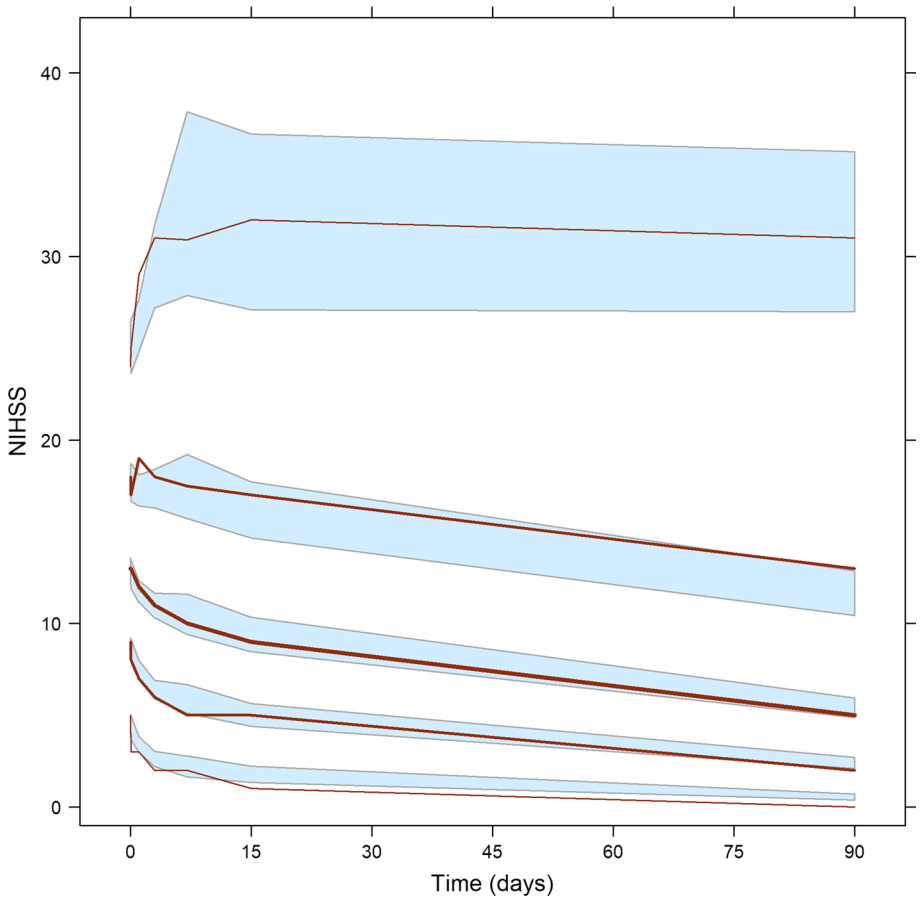

Fig. 3 Visual predictive check for NIHSS time profiles using LOCF for dropout values for the final combined mRS/NIHSS model. Red lines are the 5th, 25th, 50th (median), 75th and 95th percentiles of the observed data and the light blue areas show the 95% of the simulated quantiles (Color figure online)

the observed condition (0% reduction) and subsequent reductions in volume increase in 20% steps up to a complete reduction of volume increase (100% reduction). Complete reduction predicts 88.0% of patients to survive compared to 80.7% without an intervention.

If it were possible to identify patients beforehand that were likely to have a large increase in IHV if left untreated, then treatment of these patients could result in a more favorable outcome. 25% of the patients in the VISTA-ICH database have an untreated IHV increase of > 8 mL. If the IHV increase in these patients could be blocked completely, then the percentage patients alive would increase from 60.4% when left untreated, to 87.9% (Table 2). The remaining 75% of patients with an untreated IHV increase of ≤ 8 mL show no further benefit, even with a total reduction in IHV increase (Table 2).

Trial simulations and power calculations

Trial simulations were performed by changing model parameters with impact on predicted outcome at day 90, to allow qualitative assessment of required sample sizes for the different analysis models and sampling schedules. The arbitrary effect size resulted in an increase in the simulated percentage of patients alive at day 90 from 81 to 88%, a change in mRS of -0.50, and a change in NIHSS of -2.21.

The three longitudinal models for both the standard and the enriched sampling schedule were compared using 500 simulated trials with 150 patients per arm. The standard sampling scheme led to a low fraction of successful parameter estimation results for both the mRS-alone (28.8%) and the combined mRS/NIHSS models (47.6%);

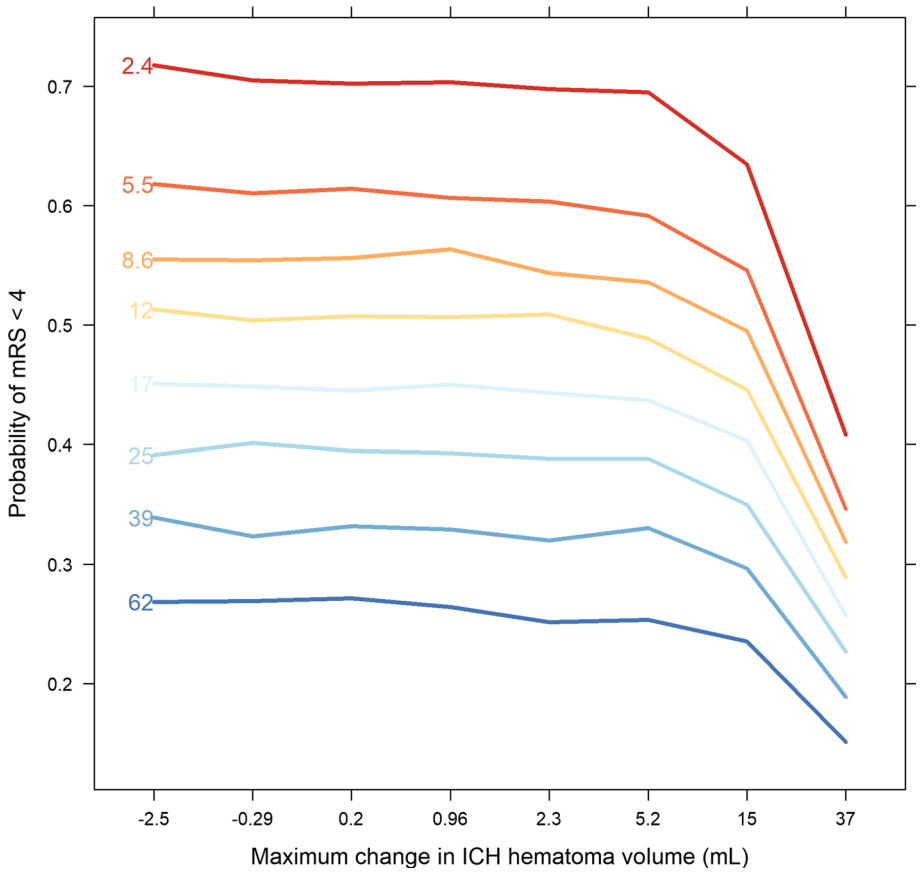

Fig. 4 Simulated probability of obtaining an mRS score < 4 at 90 days as a function of baseline hematoma volume (colored lines with associated baseline hematoma volume category value) and maximum change in volume categories (Color figure online)

switching to an enriched sampling scheme increased success rates to 85.0% and ≥ 95.0%, respectively.

Power curves are provided in Fig. 5; the combined mRS/NIHSS model provided the most powerful analysis, followed by the NIHSS-only model and then the mRS-only model. Switching from a standard to an enriched sampling schedule led to an increase in power. All longitudinal models provided a substantial reduction in sample size compared to the binomial logistic regression approach on dichotomized mRS scores.

Discussion

In order to provide a modeling and simulation framework to aid in the design and analysis of ICH trials, longitudinal models for mRS and NIHSS scores for the placebo data in the VISTA-ICH database were developed. The combined mRS/NIHSS model provided an accurate description of the evolution of mRS and NIHSS scores over time, and dropout due to death was well described by a constant hazard survival model with separate and highly different survival rates for the two mixture populations.

The current analysis indicates that reduction of IHV increase will have to be substantial to impact treatment outcome in cerebral hemorrhage patients, and is in line with the failed Recombinant Activated Factor VII trials in ICH [3]. Additionally, the results suggest that only patients with IHV increases of more than 8 mL would benefit from treatment that reduces the volume increase. These assertions depend on the assumption that the difference in clinical outcome between a placebo patient with a small ICH volume increase and a placebo patient with a large ICH volume increase can be used to deduce the effect of ICH volume reduction within a patient. This may or may not be the case; an intervention that reduces the ICH

Table 2 Summary measures for percentage of patients alive, mRS scores, and NIHSS scores at day 90 assuming different % reduction in intra-cranial hemorrhage hematoma volume increase, for all patients, for patients split by the subset that would have a volume increase of > 8 mL if left untreated, and patients with ≤ 8 mL increase

Reduction in volume increase	Percentage alive	Mean mRS	Delta mRS	Percentage mRS < 4	Mean NIHSS	Delta NIHSS
All patients						
0% reduction	80.7	3.53	0.00	44.7	8.85	0.00
20% reduction	82.3	3.49	− 0.04	45.3	8.55	− 0.30
40% reduction	83.8	3.45	− 0.08	46.0	8.24	− 0.61
60% reduction	85.4	3.41	− 0.12	46.7	7.93	− 0.92
80% reduction	86.9	3.37	− 0.16	47.3	7.64	− 1.21
100% reduction	88.0	3.34	− 0.19	47.8	7.43	− 1.42
Patients with > 8 mL volume increase (25% of patients)						
0% reduction	60.4	4.40	0.00	26.1	14.38	0.00
20% reduction	66.3	4.25	− 0.15	28.5	13.24	− 1.14
40% reduction	72.2	4.09	− 0.31	31.0	12.08	− 2.30
60% reduction	78.2	3.93	− 0.47	33.6	10.91	− 3.47
80% reduction	84.0	3.78	− 0.62	36.0	9.77	− 4.61
100% reduction	87.9	3.68	− 0.72	37.6	8.99	− 5.39
Patients with ≤ 8 mL volume increase (75% of patients)						
0% reduction	87.6	3.24	0.00	50.9	6.98	0.00
20% reduction	87.7	3.24	0.00	51.0	6.96	− 0.02
40% reduction	87.8	3.24	0.00	51.0	6.95	− 0.03
60% reduction	87.9	3.23	− 0.01	51.1	6.93	− 0.05
80% reduction	87.9	3.23	− 0.01	51.1	6.92	− 0.06
100% reduction	88.0	3.23	− 0.01	51.2	6.91	− 0.07

Using estimates from the final mRS/NIHSS combined model. Delta mRS = change in mRS from the 0% reduction value. Percentage mRS < 4 = % of patients at day 90 with an mRS score of 0, 1, 2 or 3. Delta NIHSS = change in NIHSS from the 0% reduction value. Mean and Delta NIHSS values were calculated with last observation carried forward replacement for patients that died during treatment

volume increase may have a much more profound effect in ICH patients, but the current data with only placebo patients cannot provide supporting evidence.

Selectively targeting patients that would have large volume increases if left untreated and assuming a methodology would be capable of detecting such patients, would result in much larger treatment effects if the intervention results in a substantial reduction of a hematoma volume increase. Trials in ICH can be further optimized by choosing sensitive analysis strategies and appropriate sampling schedules. Our results suggest that the combined mRS/NIHSS model is the most sensitive analysis strategy amongst the three analysis strategies considered, followed by the NIHSS-only model and finally the mRS-only model. Analysis with binomial logistic regression using only dichotomized day 90 mRS data required much larger sample sizes. The enriched sampling schedule with mRS measurements at 7, 15, 30, 45 and 90 days, and NIHSS measurements at 0 and 1 h, and at 1, 3, 7, 15, 30, 45 and 90 days, consistently required a smaller sample size than the standard sampling schedule with mRS measurements at 15 and 90 days, and NIHSS measurements at 0 and 1 h, and at 1, 3, 15, and 90 days, and a sampling schedule with only two mRS assessments at 15 and 90 days may lead to difficulties in parameter estimation, both with and without use of NIHSS information. If the burden of increasing the number of mRS assessments is considered too high in clinical practice, further increase in the number of mRS assessments might be feasible using an alternative, less labor-intensive assessment of mRS scores, e.g., assessment of mRS over phone which has been shown to be reliably comparable with a face-to-face assessment and can significantly reduce the burden on subjects [15, 16].

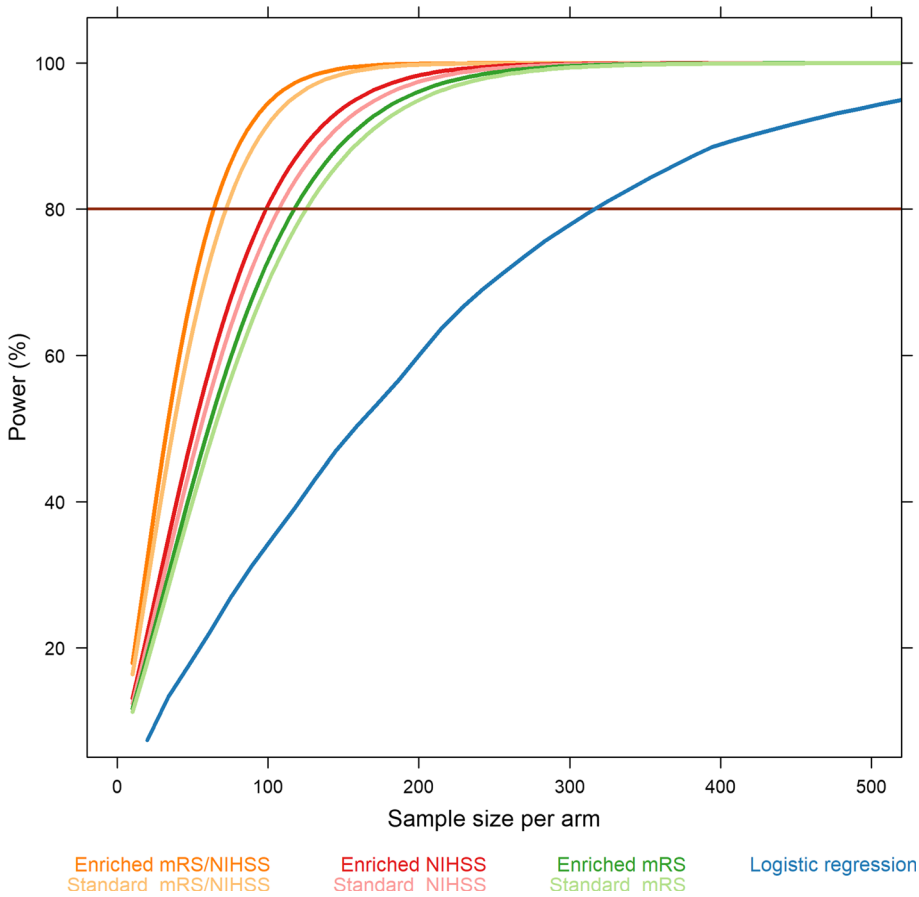

Fig. 5 Power versus sample size per arm for the three longitudinal models with either standard or enhanced sampling, for the survival and improvement effect size, compared to a binomial logistic regression on mRS ≥ 5 at day 90 data

Conclusion

Longitudinal analysis models using multiple endpoints provide an increase in power compared to the standard approach of using only mRS measurements at day 90. By using a combined mRS/NIHSS model as illustrated with the current work, the power of the longitudinal models can be further increased, allowing smaller studies to detect treatment effects. The proposed approach allows designing trials more efficiently, especially in a condition in very vulnerable patients, for which effective treatments are highly needed.

Acknowledgements The authors would like to thank Bill Denney for his valuable input in early stages of model development. VISTA-ICH Steering Committee: Daniel F. Hanley (Chair), Kenneth S. Butcher, Stephen Davis, Barbara Gregson, Kennedy R. Lees, Patrick Lyden, Stephan Mayer, Keith Muir and Thorsten Steiner.

Author contributions RS, SN, LH, and MOK wrote the manuscript; RS modelled the data.

Compliance with ethical standards

Conflict of interest At the time this manuscript was submitted for publication, L.H. and S.N were full-time employees of Pfizer Ltd, and R.S and M.O.K. were paid consultants for Pfizer Ltd.

References

1. Broderick J et al (2007) Guidelines for the management of spontaneous intracerebral hemorrhage in adults 2007 update. Stroke 38:2001–2023
2. Mozaffarian D et al (2015) Heart disease and stroke statistics—2015 update: a report from the American Heart Association. Circulation 131:e29–e322
3. Mayer SA et al (2008) Efficacy and safety of recombinant activated factor VII for acute intracerebral hemorrhage. N Engl J Med 358:2127–2137
4. http://www.virtualtrialsarchive.org/vista-ich/. Accessed 17 July 2018
5. Lynden P (2017) Using the National Institutes of Health Stroke Scale, A Cautionary tale. Stroke 48:513–519
6. https://stroke.nih.gov/documents/NIH_Stroke_Scale.pdf. Accessed 17 July 2018
7. Banks JL, Marotta CA (2007) Outcomes validity and reliability of the modified rankin scale: implications for stroke clinical trials: a literature review and synthesis. Stroke 38:1091–1096
8. mRS. http://rankinscale.org/. Accessed 25 May 2016
9. Sheiner LB (1994) A new approach to the analysis of analgesic drug trials, illustrated with bromfenac data. Clin Pharmacol Ther 56:309–322
10. Beal SL, Sheiner LB, Boeckmann AJ, Bauer RJ (Eds) (1989–2011) NONMEM users guides. Icon Development Solutions, Ellicott City
11. Lindbom L, Pihlgren P, Jonsson EN (2005) PsN-toolkit—a collection of computer intensive statistical methods for non-linear

mixed effect modeling using NONMEM. Comput Methods Progr Biomed 79:241–257
12. R Development Core Team (2015) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
13. Karlsson MO, Holford N (2008) A tutorial on visual predictive checks. PAGE 17:1434
14. Ueckert S, Karlsson MO, Hooker AC (2016) Accelerating Monte-Carlo power studies through parametric power estimation. J Pharmacokinet Pharmacodyn 43:223–234
15. Bruno A et al (2011) Simplified modified rankin scale questionnaire. Reproducibility over the telephone and validation with quality of life. Stroke 42:2276–2279
16. Savio K et al (2013) Reliability of the modified Rankin Scale applied by telephone. Neurol Int 5:e2