NT-proBNP Qualifies as a Surrogate for Clinical End Points in Heart Failure

Walter Schmitt[1], Hauke Rühs[2], Rolf Burghaus[3], Christian Diedrich[3], Sulav Duwal[3], Thomas Eissing[1], Dirk Garmann[2], Michaela Meyer[1], Bart Ploeger[1] and Jörg Lippert[1,*]

N-terminal pro-B-type natriuretic peptide (NT-proBNP) is a well-established biomarker in heart failure (HF) but controversially discussed as a potential surrogate marker in HF trials. We analyzed the NT-proBNP/mortality relationship in real-world data (RWD) of 108,330 HF patients from the IBM Watson Health Explorys database and compared it with the NT-proBNP / clinical event end-point relationship in 20 clinical HF studies. With a hierarchical statistical model, we quantified the functional relationship and interstudy variability. To independently qualify the model, we predicted outcome hazard ratios in five phase III HF studies solely based on NT-proBNP measured early in the respective study. In RWD and clinical studies, the relationship between NT-proBNP and clinical outcome is well described by an E_{max} model. The NT-proBNP independent baseline risk (R_0, RWD/studies median (interstudy interquartile range): 5.5%/3.0% (1.7–4.9%)) is very low compared with the potential NT-proBNP–associated maximum risk (R_{max}: 55.2%/79.4% (61.5–89.0%)). The NT-proBNP concentration associated with the half-maximal risk is comparable in RWD and across clinical studies (EC_{50}: 3,880/2,414 pg/mL (1,460–4,355 pg/mL)). Model-based predictions of phase III outcomes, relying on short-term NT-proBNP data only, match final trial results with comparable confidence intervals. Our analysis qualifies NT-proBNP as a surrogate for clinical outcome in HF trials. NT-proBNP levels after short treatment durations of less than 10 weeks quantitatively predict hazard ratios with confidence levels comparable to final trial readout. Early NT-proBNP measurement can therefore enable shorter and smaller but still reliable HF trials.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?
☑ N-terminal pro-B-type natriuretic peptide (NT-proBNP) is known as a diagnostic and prognostic biomarker for heart failure.

WHAT QUESTION DID THIS STUDY ADDRESS?
☑ In this study, we investigate whether NT-proBNP concentrations qualify as a surrogate for clinical composite end points in heart failure trials. Based on data from 25 clinical studies and ~ 100,000 real-world medical records, we demonstrate that across all subtypes of heart failure the NT-proBNP event rate relationship follows an E_{max} relationship.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?
☑ Our quantitative model predicts mid-term to long-term event rates in clinical trials based on short-term response of NT-proBNP to treatment. The qualification of the model with data from five phase III heart failure studies demonstrates the quantitative accuracy of predictions.

HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?
☑ A model-based use of short-term NT-proBNP response can replace pivotal efficacy event end points in heart failure trials and allows reliable assessments of probability of technical success already in phase II.

N-terminal pro-B-type natriuretic peptide (NT-proBNP) is an established diagnostic biomarker for the presence of heart failure[1,2] (HF) reflected in diagnostic algorithms of current HF guidelines.[3,4] Over the past decades, a large number of publications investigated the prognostic properties of NT-proBNP for mortality and various cardiovascular events in patients with HF[5] and other cardiovascular (CV) diseases as well as in general elderly populations.[2,6] The association of NT-proBNP elevation with the future occurrence of CV events was shown in several studies. Van Veldhuizen[7] demonstrated that patients with comparable NT-proBNP levels bear the same risk for HF hospitalization and all-cause mortality independent of their type of HF, i.e., HF with reduced or preserved ejection fraction (HFrEF or HFpEF). This finding was recently confirmed by Lam et al.[8] where plasma levels of NT-proBNP were similarly predictive of death in the different HF phenotypes including HF patients with medium-range ejection fraction (HFmrEF).

While NT-proBNP–related measures were also used as primary end points in several phase II HF studies,[9–11] it is still a matter

of debate whether NT-proBNP is a reliable predictive response marker for treatment-associated changes of risk for clinical events and could thus serve as a surrogate outcome measure in pivotal clinical trials.[12,13]

Clinical data for current HF therapies indicate that changes in NT-proBNP can be determined within a few months[14] or even a few weeks[11,15] after the start of treatment. By contrast, clinical events occur over months and years with relatively low events rates. Establishing NT-proBNP as a qualified surrogate of clinical events would therefore have high relevance for clinical development and regulatory sciences. It could enable significantly shorter pivotal trials with smaller sample size resulting from the continuous nature of the NT-proBNP end point and the possibility of model-based trial analyses.[16,17]

A prerequisite for the establishment of NT-proBNP as a surrogate for clinical end points is a validated mathematical relationship between NT-proBNP concentrations and clinical end points enabling reliable, quantitative predictions. Previous investigations used various statistical approaches and implicit assumptions about the nature of the relationship between NT-proBNP concentration and CV risk (e.g., linear or logistic),[7–9] but none of the approaches could demonstrate a consistent relationship across different clinical studies. Recent investigations considered changes and temporal courses of NT-proBNP concentrations as predictors for clinical outcome.[13,18,19] However, a consistent quantitative framework is still missing.

In the present study we identified the quantitative relationship between NT-proBNP concentrations in plasma and the risk for CV events using real-world medical record data from a large cohort of 108,330 patients with HF of any type (for a formal definition see **supplementary material**). Using a model-based meta-analysis of data from 20 published interventional and observational studies, we confirmed the relationship derived from Real World Data (RWD) and quantified remaining uncertainty and empirical interstudy variability between these clinical trials. Finally, in an independent qualification, we confirmed the predictive power of the established statistical model. We predicted long-term clinical outcome in phase III studies with treatment durations of 10–49 months using only NT-proBNP measurements taken within the first 4 months after start of treatment.

Methods
Data sources
Our analysis of real-world electronic medical records is based on the IBM Watson Health (Armonk, NY) Explorys database (status as of March 18, 2020).

For the meta-analysis of published clinical study data, we performed a systematic literature search for clinical studies where NT-proBNP or BNP concentration values and corresponding information on CV event outcomes were reported per individual study strata. A search in Elsevier Embase on January 29, 2019 (for search terms see **supplementary material**) resulted in 3,818 potentially relevant publications. In three steps (see **Figure 1**) 21 clinical studies that either included patients with a chronic HF condition only or did not use specific diagnoses in the inclusion or exclusion criteria at all were identified. For our analysis, we added three phase III heart failure studies not captured by the EMBASE search[20–22] and one observational study[23] published after our EMBASE search.

Out of the resulting list of 25 studies, we selected 20 studies with a total of 47,590 patients for the quantification of the NT-proBNP / clinical end-point relationship (**Table 1**). Included in this data set are NT-proBNP baseline pretreatment measurements and outcome data from a pooled analysis of the two phase III trials, ATMOSPHERE and PARADIGM-HF.[24]

The remaining five publications of phase III studies with a total of 22,860 patients were kept separate to test the predictive performance of the established relationship in an independent qualification step (**Table 1**). These five studies qualified for this test, because for them NT-proBNP measurements were reported for timepoints long enough after treatment in order to already reflect the treatment effect (7–16 weeks), but on the other hand were way before the end of the follow-up periods (10.4–39 months). ATMOSPHERE and PARADIGM-HF were both also included individually in this test, but different NT-proBNP measurements after 16 weeks and 7–10 weeks of treatment reflecting drug-related changes were used for clinical outcome prediction. The data originate from the same studies, but they have been compiled following different criteria and are therefore partially independent from those used for model development. Model building relied on NT-proBNP baseline measurements, and strata were defined by NT-proBNP levels. Model validation used NT-proBNP measurements after start of treatment, and strata were defined by treatment (verum vs. control).

Preparation of analysis data sets

Real-word electronic medical records. First, we selected all patients in the Explorys database with at least one documented diagnosis of general heart failure (SNOMED concept ID 84114007) and at least one documented diagnosis of a specific heart failure subtype (see **Supplementary Material**). Out of these patients, we selected all patients with a documented NT-proBNP measurements within the interval between 1 month and 25 months after the first reported HF diagnosis (denoted follow-up interval hereafter). Data documented within the first month after HF diagnosis were excluded to prevent potential bias resulting from NT-proBNP elevations associated with acute decompensation crisis leading to incident HF diagnosis. We removed all patients for whom we could not determine survival status within the follow-up interval (either via a death code or documentation of consecutive encounters across the entire follow-up interval). For patients that had more than one NT-proBNP value reported within the follow-up interval the individual median was used. This resulted in 108,330 patients in our real-world analysis data set for which we had data pairs consisting of an NT-proBNP measurement and a binary 2-year survival outcome. A documented death code existed for 24.9% of these patients.

Literature data. For three clinical studies that reported BNP concentrations instead of NT-proBNP (see **Table 1**), we converted BNP into NT-proBNP estimates using a conversion factor of 6.67 (mean derived from values reported in four different published studies;[25–28] ±0.88 standard deviation; see **Table S3**). In the case of interventional studies, the first measurement after the start of treatment were used for model building if several NT-proBNP values were reported. When baseline NT-proBNP and change of NT-proBNP after treatment were reported, we calculated the resulting NT-proBNP concentrations after treatment (see **Table 1**).

The clinical end points differed between studies. If more than one clinical study end point was reported for a study, the one best matching those available for the other studies was selected. As a consequence, typical end points included are mortality and composites of mortality and hospitalization events (see **Table 1**).

Data analysis and modeling & simulation
The present study consists of three analyses which were conducted sequentially:
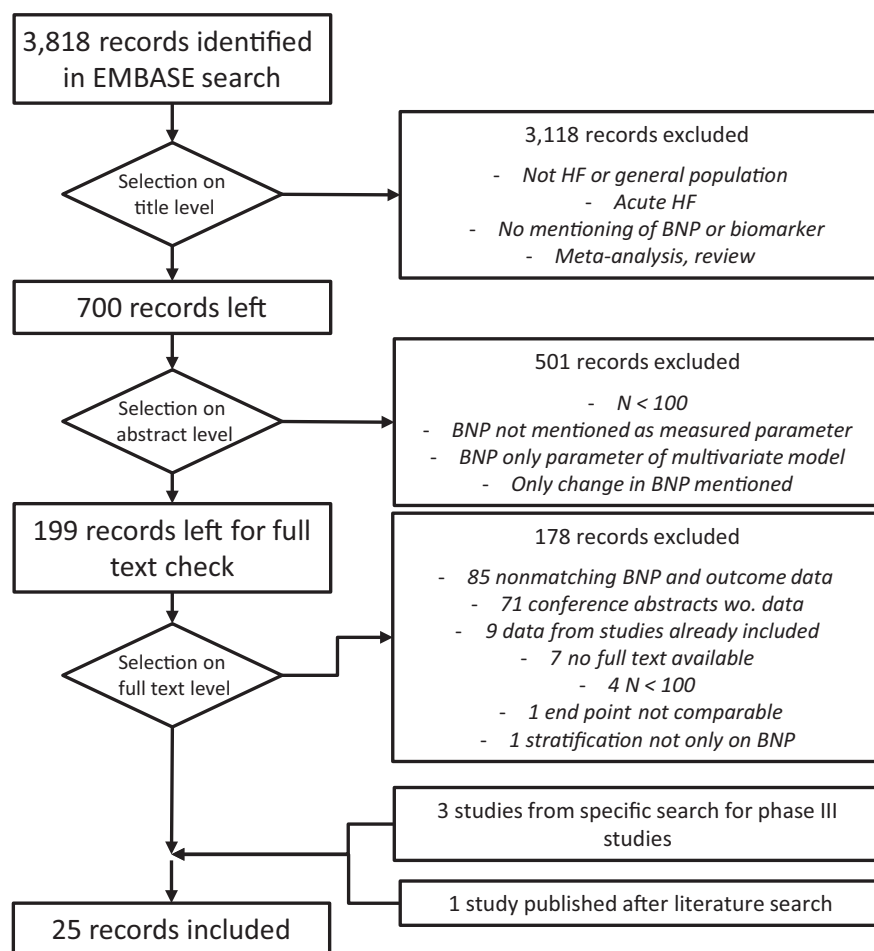
Figure 1 Flow chart describing the procedure for identification and selection of published clinical studies reporting per strata average (NT-pro) BNP concentrations and clinical event end-point rates or hazard ratios. HF, heart failure; BNP, brain natriuretic peptide; wo., without.

1. Analysis of individual patient-level real-world data from the Explorys database and developing a model describing the relation between NT-proBNP and risk, independent of any treatment.

2. Model-based meta-analysis of study cohort-level data from 20 published clinical studies employing the model structure developed in step 1, and confirming the performance of NT-proBNP as prognostic marker independent of treatment effects.

3. Testing the predictivity of the model with respect to treatment effects using the data from five interventional studies with clinical end points, for which NT-proBNP measurements after study intervention but long before end of study had been published. This data set allows for testing whether the effects of the study intervention on clinical outcome of the study could have been predicted at an early timepoint based on NT-proBNP measurements.

Visual inspection of the real-world data consisting of individual patient-level NT-proBNP measurements and 2-year survival data suggested that the relationship between logarithmic NT-proBNP concentrations and the risk for clinical end points is sigmoidal, with little dependency of risks at both very low and very high NT-proBNP concentrations. **Figure 2a** shows the real-world data.

Therefore, an E_{max}-type function was used to model the relation between the independent variable NT-proBNP and the dependent clinical risk R:

$$R(NT - proBNP) = R_0 + R_{max} \times (1 - R_0) \times \frac{NT - proBNP}{NT - proBNP + EC_{50}} \quad (1)$$

In the case of the real-world data, the clinical risk R represents the probability of an individual patient with his or her NT-proBNP measurement dying within a period of 2 years after the first HF diagnosis. In the case of the analysis of the clinical study data, the risk R represents the clinical event rate in a study arm or stratum with a given median NT-proBNP level of all the patients in the study arm/stratum.

The parameters that need to be identified from the data sets are R_0, R_{max}, EC_{50}. R_0 describes the baseline risk independent of NT-proBNP observable at very low NT-proBNP concentrations. R_{max} represent the maximum additional risk associated with elevated NT-proBNP concentrations. EC_{50} is the NT-proBNP concentration at which 50% of the increase in risk from R_0 to $R_0 + R_{max}$ has been reached.

In the case of the Explorys data the parameters of the model (Eq. 1) were estimated by fitting the function to the individual median NT-proBNP concentration during the observation interval and the individual's survival status (i.e., all-cause death) at the end of the interval by nonlinear maximum likelihood regression.

The data derived from the 20 literature studies were modeled employing the same function (Eq. 1) in a hierarchical statistical model. The

Table 1 List of clinical studies used for model building and model validation respectively

Study First author (study name)	Therapy	Indication(s)	N[a]	Timepoint of NT-proBNP measurement	Follow-up period (month)	End point analyzed[b]	Ref.
Clinical studies selected for model building							
Alehagen	Observational	Elderly with HF symptoms	470	Baseline	156	CV mortality	32
Greene (ASTRONAUT)	Aliskiren	HFrEF	1,349	End of treatment	11.3	CV mort. + HF hosp.	33
Cleland (PEP-CHF)	Perindopril	CHF	375	Baseline	12	All cause mort. and HF hosp.	34
Van Veldhuisen (COACH)	Observational	HFrEF + HFpEF	615	Baseline[c]	18	All cause mort. and HF hosp.	7
De Fillipi (CV risk in elderly)	Observational	General elderly population	4,310	Baseline	143	HF and CV mort.	35
Franke	Observational	systolic CHF	504	6 month post inclusion	12	All-cause mortality	36
Frankenstein	Observational	CHF	618	Baseline	37	All-cause mortality	37
Gastelurrutia (MUSIC)	Observational	CHF	979	Baseline	44	All-cause mortality	38
Mishra (Heart & Soul)	Observational	Stable CAD, HF	635	End of follow up	615	CV mort. + HF hosp.	19
Anand (iPRESERVE)	Irbesartan	HFpEF	3,480	Baseline	49.5	CV mort. + HF hosp.	39
Kristensen (PARADIGM-HF, ATMOSPHERE)	Entresto, Aliskiren	HFrEF	14,737	Baseline	N.R.	CV mort. + HF hosp.	24
Kubanek	Observational	CHF	354	4–6 month post inclusion	38.8	All-cause mortality	40
Lam	Observational	HFpEF, HFmrEF, HFpEF[e]	2,039	Baseline	24	All-cause mortality	8
Olsson (COMET)	Carvedilol, Metoprolol	HFpEF	1,559	Baseline	58	All-cause mortality	41
Patton	Observational	General population	5,447	Baseline	150	Sudden cardiac death	42
Spinar	Observational	HFmrEF, HFrEF	1,088	Baseline	24	All cause death, Transplantation, LVAD	23
Toggweiler	Observational	Cardiac outpatients	2,313	Baseline	24	All cause mort., HF hosp. and MI	43
Tsuchida	Observational	General population	3,123	Occasional[c]	66	CV mort. + CV hosp.	44
Masson (ValHeft)	Valsartan	HFrEF	1,742	End of treatment[d]	24	All-cause mortality	45
Welsh (RED-HF)	Darbepoetin Alfa	HFrEF + anemia	1,853	Baseline	28	CV mort. + HF hosp.	46
Clinical phase III studies selected for validation of predictivity							
Athmosphere	Aliskiren, Aliskiren + Enalapril	HFrEF	7,016	16 w. after 1st dose	36.6	CV mort. + HF hosp.	21
Best	Bucindolol	HFrEF	2,708	12 w. after 1st dose[c]	12	CV mort. + CHF hosp.	20,47

(Continued)

Table 1 (Continued)

Study First author (study name)	Therapy	Indication(s)	N^a	Timepoint of NT-proBNP measurement	Follow-up period (month)	End point analyzed[b]	Ref.
Copernikus	Carvedilol	HFrEF	2,292	~12 w. after 1st dose	10.4	CV mort. + HF hosp.	22,48
PARADIGM	Entresto	HFrEF	8,399	7–10 w. after 1st dose	28.8	CV mortality + HF hospitalization	15,30,49
TOPCAT	Spironolactone	HFpEF	2,445	Baseline	39	CV mortality + HF hospitalization	50,51

BNP, brain natriuretic peptide; CAD, coronary artery disease; CHF, congestive heart failure; CV, cardiovascular; HF, heart failure; HFmrEF, HF with medium-range ejection fraction; HFpEF, HF with preserved ejection fraction; HFrEF, HF with reduced ejection fraction; hosp., hospitalization; LVAD, left ventricular assist device; MI, myocardial infarction; mort., mortality; N.R., not reported; NT-proBNP, N-terminal pro-B-type natriuretic peptide; w., week.
[a]Number of patients in total study or substudy dealing with NT-proBNP information if applicable. [b]End point selected for present evaluation if several end points were reported. [c]BNP values reported and converted to NT-proBNP. [d]NT-proBNP at follow-up calculated from reported baseline values and changes. [e]Results for HFrEF, HFmrEF, and HFpEF were reported separately and were thus considered as separate data in the modeling process.

general as well as the study-specific model parameters and their statistical distributions were estimated in a self-consistent way using a Bayesian Monte Carlo Markov chain (MCMC) approach. The binomial event rate model reflected sample size and resulting confidence in the study cohort event rates. The Bayesian framework ensured a systematic weighting of the data points in accordance with their respective information content. Further details of the modeling process and the statistical approaches can be found in the **Supplementary Material**.

In a consecutive step, the hierarchical statistical model was used to predict odds ratios between strata in five phase III studies based on the NT-proBNP median of strata determined within 4 months after the start of treatment. Predicted odds ratios were compared with reported odds ratios for study event end points at the end of the study to validate the model's predictive power.

All analyses were performed with the statistical programming languages R (The R Foundation, https://www.r-project.org/) and Stan (The Stan Development Team, https://mc-stan.org/).[29]

Results

The E_{max}-like model is an excellent representation of the data extracted from electronic medical records in Explorys (**Figure 2a**), motivating the choice of an E_{max}-like function for the consecutive model-based meta-analysis of clinical study data. The values identified with the model are 0.055 ± 0.002 (90% confidence interval) for the baseline 2-year-mortality (R_0), 0.552 ± 0.007 for the maximum NT-proBNP–associated mortality (R_{max}), and 3,880 ± 149 pg/mL for the NT-proBNP concentration where 50% of the NT-proBNP–associated mortality has been reached (EC_{50}).

For a typical study, the model-based meta-analysis provides estimated median values for R_0 and R_{max} of 0.030 and 0.794, respectively. These values have little meaning since the absolute event rates reflected in R_0 and R_{max} depend on the definition of the clinical event end point. A narrow definition counting only severe but rare events such as mortality will lead to lower rates than a broad composite end point definition that also counts, e.g. hospitalization, New York Heart Association (NYHA) class and other reflections of progression.

The median EC_{50} of 2,414 pg/mL for all studies, however, has a clinical relevance independent of end-point choices since it reflects the relationships between relative clinical risks and NT-proBNP. The value obtained from clinical study data, importantly, is very similar to the value obtained from the RWD analysis. The empirical interstudy variability is relatively large in all three parameters, and extreme values can differ by a factor of 10 (for more details about individual study results see **Supplementary Material**, **Table S6**). As pointed out above, for baseline event probability R_0 and the maximum NT-proBNP–associated event rate R_{max}, a high interstudy variability is to be expected since all studies differed both in end-point definitions and inclusion and exclusion criteria. Both should modulate absolute levels of event probability. The heterogeneity is also evident in the raw data from clinical studies (**Figure 2b**). When corrected for these study-specific differences in R_0 and R_{max}, confidence intervals of empirical event probabilities of almost all clinical study strata overlap with the median of NT-proBNP–event probability model derived from clinical studies (54 of 97 overlapping with median model prediction; 96 of 97 overlapping with the 95% confidence interval of the model prediction) and also the model derived from RWD (58 of 97 overlapping with the model point estimate,
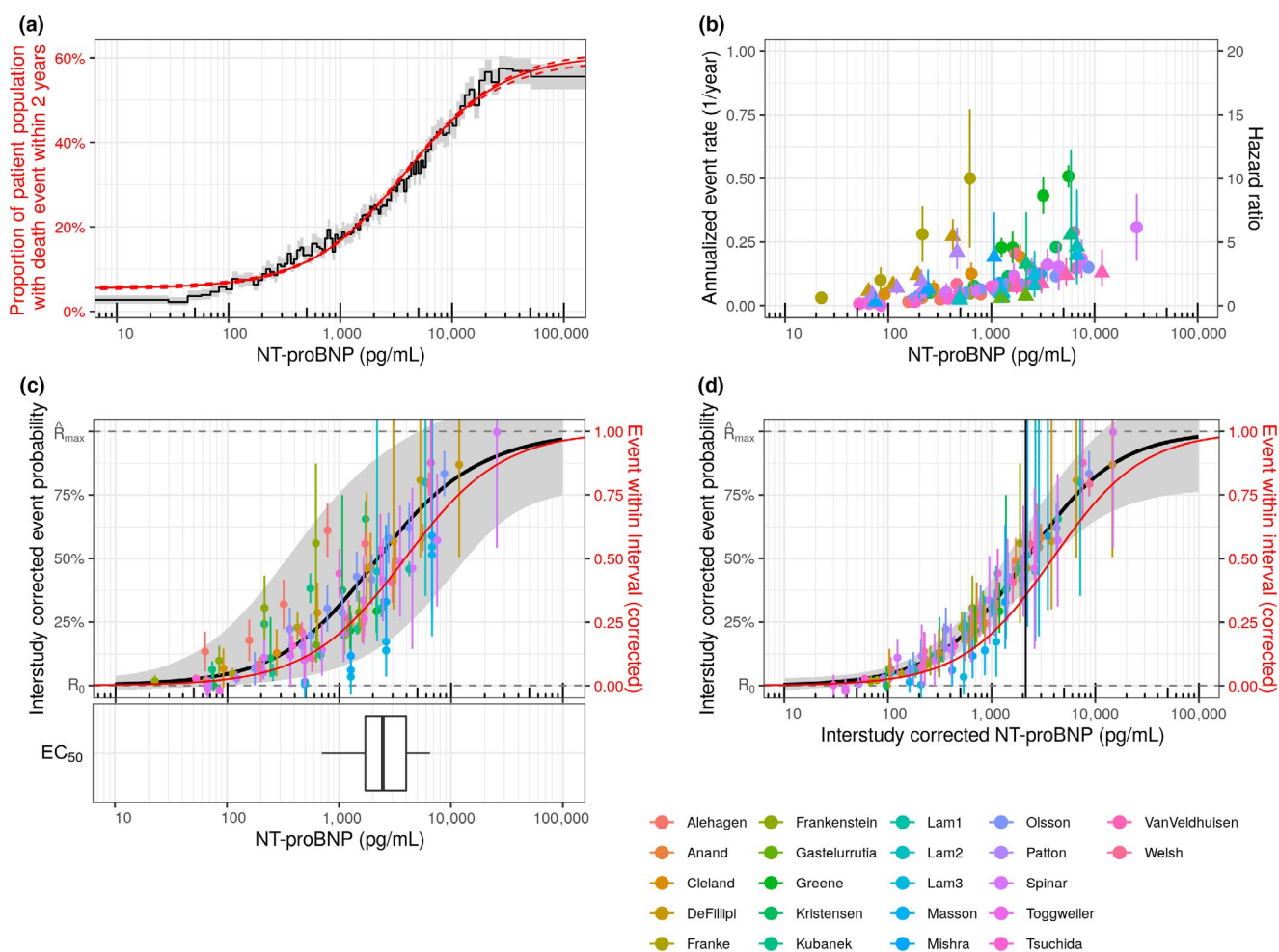
Figure 2. NT-proBNP dependent risk in RWD and clinical studies. (**a**) Proportion of patients dying within 2 years after initial HF diagnosis calculated for percentiles of the NT-proBNP distribution in the population. Data derived from Explorys medical record database (black line: point estimates; gray band: 95% CI). Data is compared with the model prediction obtained from the E_{max}-like function (red line: median; dashed line: 95% CI). By design, every NT-proBNP bin is supported by an even number of patients (one percentile) but spans an NT-proBNP range of varying width. (**b**) Hazard ratios (triangles) and annualized event rates (circles) vs. NT-proBNP concentrations as reported in original publications (see **Table 1**). (**c**) Black line: Median of model simulations. Gray area: 95% CI of simulations including uncertainties of all model parameters and interstudy variability of EC_{50}. Symbols: Event rates corrected for interstudy variabilities of R_0 and R_{max}. Red line: Model simulation from **a**. The box-whisker plot depicts the distribution of study-specific EC_{50} values and reflects interstudy variability. (**d**) Model simulations with 95% CI given by the uncertainty of the model parameters. To visualize the effect and extent of interstudy variability, the data points are corrected for interstudy variability of all three parameters. The vertical line depicts the median EC_{50} of the general model. $\widehat{R}_{max}$ represents $R_0 + R_{max}(1 - R_0)$. CI, confidence interval; EC_{50}, concentration associated with the half maximal risk; HF, heart failure; NT-proBNP, N-terminal pro-B-type natriuretic peptide; R_{max}, NT-proBNP-dependent maximum risk; R_0, NT-proBNP-independent baseline risk. [Colour figure can be viewed at wileyonlinelibrary.com]

Figure 2c). Despite all differences in inclusion and exclusion criteria, study duration, and end-point definition, after correction for R_0 and R_{max} but not EC_{50}, all 20 observational and interventional clinical studies are well described by an E_{max}-like relationship between NT-proBNP and event probabilities. **Figure 3** compares published clinical data with the model predictions derived from the model-based meta-analysis for every individual study.

The general validness of an E_{max} relation can be further illustrated by finally also correcting for the study-specific EC_{50} differences by normalization to the EC_{50} of the typical study. The NT-proBNP measurements in the individual study are scaled such that the study-specific deviation of EC_{50} from the median value obtained for all studies is corrected for. As shown in **Figure 2d**, the

then still-remaining variability of the data collapses and, except for data from Lam *et al.*,[8] all data match the typical E_{max}-like behavior within the limits of the confidence interval (**Figure 2d**). Interstudy variability of EC_{50} therefore is the only relevant quantification of interstudy variability relative risk with the studies.

While the event probabilities for Lam *et al.*[8] might be systematically overpredicted by our model, the bias is small compared with uncertainty on event rates. We were not able to detect obvious differences in study characteristics that would explain the difference in behavior in the other 19 studies.

To validate the model and evaluate its predictive power, hazard ratios between study strata in phase III HF studies were predicted. These hazard ratios represent relative risks with the individual
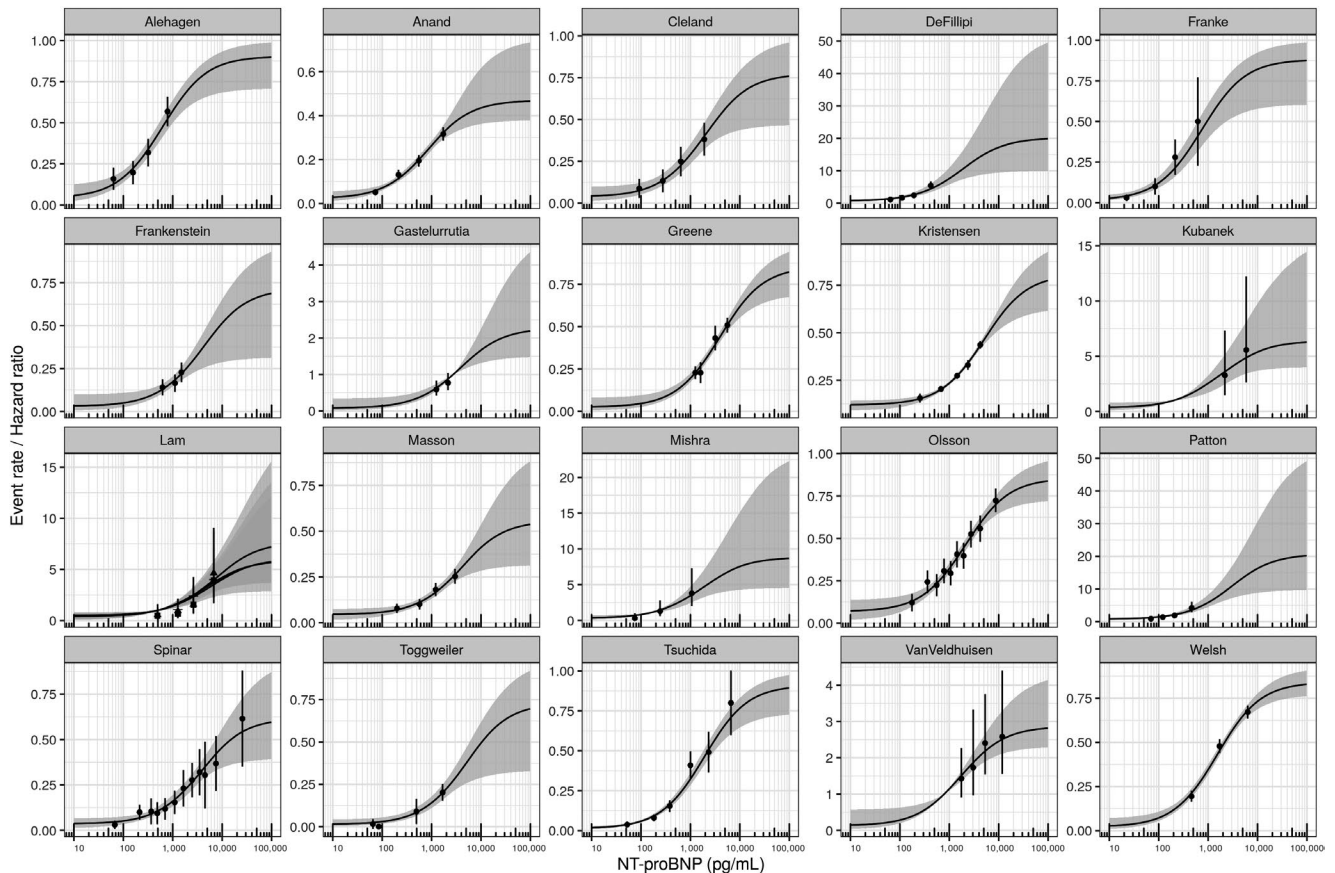
Figure 3 Comparison of individual study data and model predictions. Median (black line) and 95% percentile (gray area) show NT-proBNP response predictions per study based on individual study *post hoc* parameters. The symbols (point estimates) and vertical lines (confidence intervals) represent the original published data the model was built upon. In the case of Lam *et al.* rEF (cross), mrEF (square), and pEF (triangle) were handled separately. mrEF, medium-range ejection fraction; NT-proBNP, N-terminal pro-B-type natriuretic peptide; pEF, preserved ejection fraction; rEF, reduced ejection fraction.

study and can therefore be predicted with our model. Absolute event rates cannot be predicted since we have no possibility to predict the R_0 and R_{max} of an individual clinical study. An adjustment to study specifics is also not possible, so all predictions are based on the median EC_{50} obtained from our meta-analysis across all 20 studies.

For all five phase III studies, median NT-proBNP concentrations of the different study arms have been reported and were the only input into predictions. In addition, we also predicted the hazard ratios reported by Myhre *et al.*[30] based on median NT-proBNP for four strata in PARADIGM-HF that were defined by the quartiles of the BNP measurements in all 996 patients with a biomarker measurement at visit 7 (8–10 weeks of treatment with sacubitril/valsartan). None of our nine individual odds ratio predictions deviates significantly from the odds ratios determined at the end of the clinical studies (**Figure 4**, confidence intervals overlap with identity). The confidence intervals resulting from for our NT-proBNP–based prediction are smaller than those calculated based on event statistics at the end of the study. This indicates that NT-proBNP–based assessments of treatment effects are not only possible much earlier than those achieved with classical CV event end points, but they may also require smaller patient numbers (sample sizes) to achieve the same statistical power.

It is of note that we also included TOPCAT, where only baseline values for NT-proBNP were published. The model-predicted odds ratio deviates from 1, in line with the published point estimate based on final analysis of the event data. Since our prediction cannot reflect any action of the spironolactone therapy but is based on baseline differences only, we conclude that the nonsignificant positive trend in TOPCAT could retrospectively also be explained by a baseline NT-proBNP bias resulting from randomization.

Discussion

NT-proBNP has been discussed as a potential surrogate marker in HF development for decades. With our study we demonstrate for the first time that NT-proBNP and clinical event outcome are linked by an E_{max} relationship. This relationship quantitatively describes data from a large real-world data cohort of 108,330 HF patients as well as median NT-proBNP concentrations and clinical event rates or hazard ratios of study strata in 25 clinical studies with 70,450 patients. The statistical model resulting from our analysis predicts long-term clinical outcome based on NT-proBNP concentration measured shortly after therapy onset in phase III trials. Our results are not affected by technical hurdles such as the need to convert BNP to estimated NT-proBNP level for data from three of the clinical studies analyzed or the extreme heterogeneity of the study designs covering
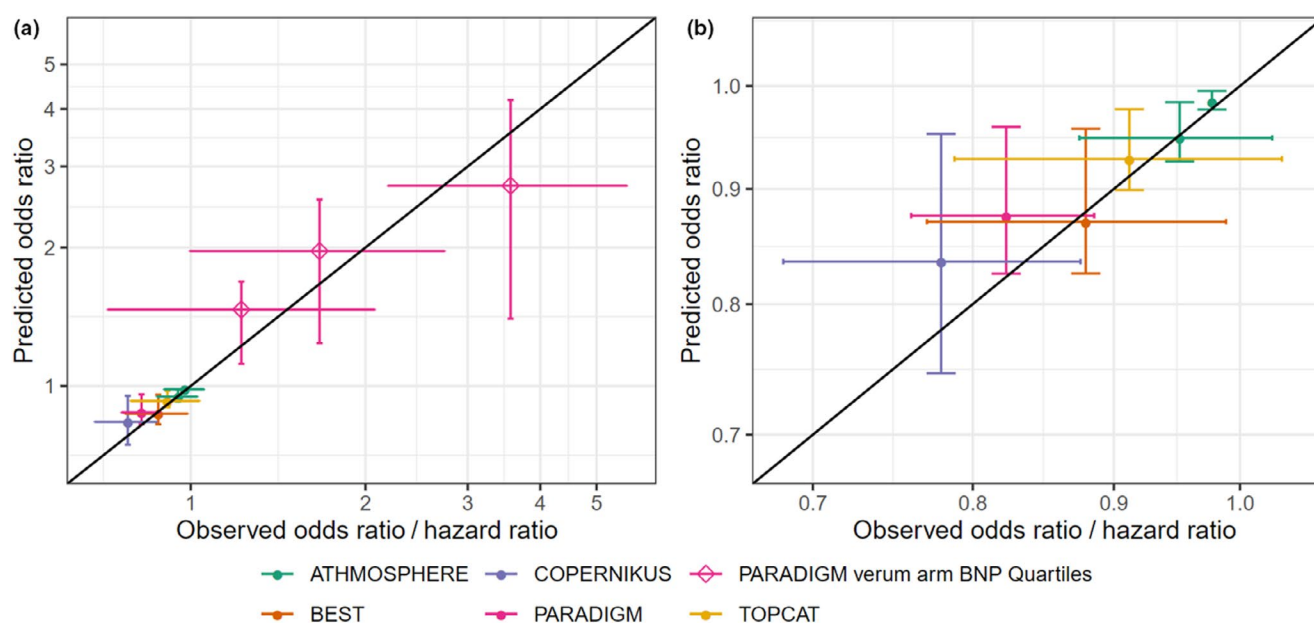
Figure 4 Comparison of predicted and observed risk ratios between control and treatment arms in five phase III heart failure studies (ATMOSPHERE, green symbols; COPERNIKUS, blue symbols; PARADIGM, red symbols; open diamonds represent data from a *post hoc* stratification of verum arm patients by quartiles of BNP measurements after 8–10 weeks of study drug treatment;[30] BEST, orange symbols; TOPCAT, yellow symbols). The black line is the line of identity. (**a**) Predictions for study treatments arms and strata in PARADIGM defined by BNP quartiles at visit 7 (open diamonds). (**b**) Zoom into figure: **a** lower left corner, on-treatment arms only. Hazard ratios are calculated relative to control arm and the first quartile of BNP, respectively. BNP, brain natriuretic peptide. [Colour figure can be viewed at wileyonlinelibrary.com]

observational and interventional clinical trials with varying inclusion and exclusion criteria, study durations and clinical event end-point definitions, and a broad range of pharmacological mechanisms of actions investigated. We conclude that the observed relationship is potentially universal across HF subtypes (as already assumed in previous studies[8,31]) and treatment types. Within the given framework of a specific study and its specific choice of a risk event end point, NT-proBNP concentrations predict clinical end-point performance, and our NT-proBNP model can explain up to a 10-fold to 25-fold difference in relative risk (i.e., R_{max}/R_0) of patient strata with very small unexplained variability of event rates (**Figure 3**). It is also worth mentioning that our results could be obtained without explicitly controlling and correcting for clinical covariates such as ejection fraction, age, body mass index, and comorbidities known to massively impact mortality and hospitalization risk. A detailed analysis of the remaining variability in the data explained by these covariates will be a valuable follow-up study. It would have been highly interesting to investigate absolute risk rates and their dependence on NT-proBNP also. Unfortunately, the covariate information required for such an analysis was not provided in the study publications, and the heterogeneity of the studies and their limited number would have prevented an adequate powering of the analysis.

In a recent meta-analysis Green *et al.*[12] have investigated the mismatch between successful surrogate end point–based phase II trials and consecutive phase III failure. The authors describe the specific issues identified with several surrogate candidates including, among several other plasma, imaging and functional markers, NT-proBNP, and postulate four requirements to be fulfilled by any surrogate. With our analysis of real-world data and 25

clinical studies, we can demonstrate that NT-proBNP fulfills all four requirements:

1. NT-proBNP correlates with outcome in patients independent of a given medication as shown in 108,330 patients from the real-world database Explorys and in all 15 clinical studies that report NT-proBNP baseline measurements. In all cases NT-proBNP and clinical outcome are linked via the same E_{max} model.
2. NT-proBNP under different specific treatments also correlates with long-term clinical outcome. In all eight studies that reported NT-proBNP measurements after the start of treatment the relationship to clinical outcome is again in line with the same E_{max} model.
3. The therapeutic modulation predicts the net effect of treatment as demonstrated by predicting event end-point odds ratios for four phase III trials (**Figure 4**).
4. The relationship between NT-proBNP and clinical HF outcomes is consistent across a very diverse set of interventions (beta-blockers, mineralocorticoid receptor antagonists, and ACE, renin, and angiotensin-receptor-neprilysin inhibitors) and patient population definitions as demonstrated by the diversity of observational and interventional clinical trials and the real-world data set included in our analysis.

We therefore conclude that NT-proBNP is qualified as a surrogate end point in interventional HF trials. It is not only suited as a reliable measure of efficacy in proof-of-concept and dose-finding settings, but it may enable much shorter pivotal HF trials without

losing statistical power. We see three necessary prerequisites if primary objectives in HF trials shall be based on NT-proBNP: (i) A disbalance in NT-proBNP baseline concentrations between study arms needs to be prevented by adequate randomization measures; (ii) NT-proBNP response needs to be assessed only after treatment effects have reached a steady-state level; and (iii) NT-proBNP–based predictions of clinical outcome need to adequately reflect the uncontrolled interstudy variability and resulting uncertainty with a model as described in this study. Simplified analysis concepts, e.g., based on thresholds for a relative reduction from baseline NT-proBNP, should not be used. If these requirements are met, NT-proBNP surrogate end point–based assessments of the efficacy of HF drugs is expected to be as reliable as conventional clinical end point–based approaches. NT-proBNP–based development strategies have the potential to significantly shorten development times without increasing attrition risk.

Supporting Information

Supplementary information accompanies this paper on the *Clinical Pharmacology & Therapeutics* website (www.cpt-journal.com).

CONFLICT OF INTEREST

The authors are employed by Bayer AG, a pharmaceutical company active in the development of heart failure treatment.

AUTHOR CONTRIBUTIONS

W.S. and J.L. wrote the manuscript. W.S., R.B., C.D., T.E., M.M., and J.L. designed the research. W.S., H.R., C.D., S.D., D.G., and B.P. performed the research and analyzed the data.

1. Oremus, M. *et al*. A systematic review of BNP and NT-proBNP in the management of heart failure: overview and methods. *Heart Fail. Rev.* **19**, 413–419 (2014).
2. Tang, W.H.W. B-type natriuretic peptide: a critical review. *Congest. Heart Fail.* **13**, 48–52 (2007).
3. Ponikowski, P. *et al*. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) developed with the special contribution of the Heart Failure Association (HFA) of the ESC. *Eur. Heart J.* **37**, 2129–2200 (2016).
4. Yancy, C.W. *et al*. 2013 ACCF/AHA guideline for the management of heart failure: executive summary: a report of the American College of Cardiology Foundation/American Heart Association Task Force on practice guidelines. *Circulation* **128**, 1810–1852 (2013).
5. Balion, C. *et al*. AHRQ comparative effectiveness reviews. In *Use of Natriuretic Peptide Measurement in the Management of Heart Failure* (Agency for Healthcare Research and Quality, Rockville, MD, 2013).
6. Geng, Z., Huang, L., Song, M. & Song, Y. N-terminal pro-brain natriuretic peptide and cardiovascular or all-cause mortality in the general population: A meta-analysis. *Sci. Rep.* **7**, 41504 (2017).
7. van Veldhuisen, D.J. *et al*. B-type natriuretic peptide and prognosis in heart failure patients with preserved and reduced ejection fraction. *J. Am. Coll. Cardiol.* **61**, 1498–1506 (2013).
8. Lam, C.S.P. *et al*. Mortality associated with heart failure with preserved vs. reduced ejection fraction in a prospective international multi-ethnic cohort study. *Eur. Heart. J.* **39**, 1770–1780 (2018).
9. Gheorghiade, M. *et al*. Effect of vericiguat, a soluble guanylate cyclase stimulator, on natriuretic peptide levels in patients with worsening chronic heart failure and reduced ejection fraction: the SOCRATES-REDUCED randomized trial. *JAMA* **314**, 2251–2262 (2015).
10. Pieske, B. *et al*. Vericiguat in patients with worsening chronic heart failure and preserved ejection fraction: results of the SOluble guanylate Cyclase stimulatoR in heArT failurE patientS with PRESERVED EF (SOCRATES-PRESERVED) study. *Eur. Heart J.* **38**, 1119–1127 (2017).
11. Solomon, S.D. *et al*. The angiotensin receptor neprilysin inhibitor LCZ696 in heart failure with preserved ejection fraction: a phase 2 double-blind randomised controlled trial. *Lancet* **380**, 1387–1395 (2012).
12. Greene, S.J. *et al*. Reassessing the Role of Surrogate End Points in Drug Development for Heart Failure. *Circulation* **138**, 1039–1053 (2018).
13. Vaduganathan, M. *et al*. Natriuretic peptides as biomarkers of treatment response in clinical trials of heart failure. *JACC Heart Fail.* **6**, 564–569 (2018).
14. Edelmann, F. *et al*. Effect of spironolactone on diastolic function and exercise capacity in patients with heart failure with preserved ejection fraction: the Aldo-DHF randomized controlled trial. *JAMA* **309**, 781–791 (2013).
15. Zile, M.R. *et al*. Prognostic implications of changes in N-terminal Pro-B-type natriuretic peptide in patients with heart failure. *J. Am. Coll. Cardiol.* **68**, 2425–2436 (2016).
16. Wason, J., McMenamin, M. & Dodd, S. Analysis of responder-based endpoints: improving power through utilising continuous components. *Trials* **21**, 427 (2020).
17. Karlsson, K.E., Vong, C., Bergstrand, M., Jonsson, E.N. & Karlsson, M.O. Comparisons of analysis methods for proof-of-concept trials. *CPT Pharmacometrics Syst. Pharmacol.* **2**, e23 (2013).
18. den Boer, S.L. *et al*. Usefulness of Serial N-terminal Pro-B-type natriuretic peptide measurements to predict cardiac death in acute and chronic dilated cardiomyopathy in children. *Am. J. Cardiol.* **118**, 1723–1729 (2016).
19. Mishra, R.K., Judson, G., Christenson, R.H., DeFilippi, C., Wu, A.H.B. & Whooley, M.A. The association of five-year changes in the levels of N-terminal fragment of the prohormone brain-type natriuretic peptide (NT-proBNP) with subsequent heart failure and death in patients with stable coronary artery disease: the heart and soul study. *Cardiology* **137**, 201–206 (2017).
20. Beta-Blocker Evaluation of Survival Trial Investigators; Eichhorn, E.J., Domanski, M.J., Krause-Steinrauf, H., Bristow, M.R. & Lavori, P.W. A trial of the beta-blocker bucindolol in patients with advanced chronic heart failure. *N. Engl. J. Med.* **344**, 1659–1667 (2001).
21. McMurray, J.J.V. *et al*. Aliskiren, enalapril, or aliskiren and enalapril in heart failure. *N. Engl. J. Med.* **374**, 1521–1532 (2016).
22. Packer, M. *et al*. Effect of carvedilol on the morbidity of patients with severe chronic heart failure: results of the carvedilol prospective randomized cumulative survival (COPERNICUS) study. *Circulation* **106**, 2194–2199 (2002).
23. Spinar, J. *et al*. Prognostic value of NT-proBNP added to clinical parameters to predict two-year prognosis of chronic heart failure patients with mid-range and reduced ejection fraction - A report from FAR NHL prospective registry. *PLoS One* **14**, e0214363 (2019).
24. Kristensen, S.L. *et al*. Prognostic value of N-terminal pro-B-type natriuretic peptide levels in heart failure patients with and without atrial fibrillation. *Circ. Heart Fail.* **10**, e004409 (2017).
25. Vanderheyden, M., Bartunek, Claeys, J., Claeys, G., Manoharan, G., Beckers, J.F., Ide L. Head to head comparison of N-terminal pro-B-type natriuretic peptide and B-type natriuretic peptide in

patients with/without left ventricular systolic dysfunction. *Clin. Biochem.* **39**, 640–645 (2006).

26. Sykes, E. *et al*. Analytical relationships among Biosite, Bayer, and Roche methods for BNP and NT-proBNP. *Am. J. Clin. Pathol.* **123**, 584–590 (2005).

27. Yeo, K.-T.J. *et al*. Multicenter evaluation of the Roche NT-proBNP assay and comparison to the Biosite Triage BNP assay. *Clin. Chim. Acta* **338**, 107–115 (2003).

28. Pfister, R., Scholz, M., Wielckens, K., Erdmann, E. & Schneider, C.A. Use of NT-proBNP in routine testing and comparison to BNP. *Eur. J. Heart Fail.* **6**, 289–293 (2004).

29. Carpenter, B. *et al*. Stan: a probabilistic programming language. *J. Stat. Softw.* **76**, 1–32 (2017).

30. Myhre, P.L. *et al*. B-type natriuretic peptide during treatment with sacubitril/valsartan: the PARADIGM-HF trial. *J. Am. Coll. Cardiol.* **73**, 1264–1272 (2019).

31. Savarese, G. *et al*. Associations with and prognostic and discriminatory role of N-terminal pro-B-type natriuretic peptide in heart failure with preserved versus mid-range versus reduced ejection fraction. *J. Card. Fail.* **24**, 365–374 (2018).

32. Alehagen, U., Dahlström, U., Rehfeld, J.F. & Goetze, J.P. Association of copeptin and N-terminal proBNP concentrations with risk of cardiovascular death in older patients with symptoms of heart failure. *JAMA* **305**, 2088–2095 (2011).

33. Greene, S.J. *et al*. Clinical profile and prognostic significance of natriuretic peptide trajectory following hospitalization for worsening chronic heart failure: findings from the ASTRONAUT trial. *Eur. J. Heart. Fail.* **17**, 98–108 (2015).

34. Cleland, J.G.F., Taylor, J., Freemantle, N., Goode, K.M., Rigby, A.S. & Tendera, M. Relationship between plasma concentrations of N-terminal pro brain natriuretic peptide and the characteristics and outcome of patients with a clinical diagnosis of diastolic heart failure: a report from the PEP-CHF study. *Eur. J. Heart Fail.* **14**, 487–494 (2012).

35. deFilippi, C.R., Christenson, R.H., Gottdiener, J.S., Kop, W.J. & Seliger, S.L. Dynamic cardiovascular risk assessment in elderly people. The role of repeated N-terminal pro-B-type natriuretic peptide testing. *J. Am. Coll. Cardiol.* **55**, 441–450 (2010).

36. Franke, J. *et al*. Is there an additional benefit of serial NT-proBNP measurements in patients with stable chronic heart failure receiving individually optimized therapy? *Clin. Res. Cardiol.* **100**, 1059–1067 (2011).

37. Frankenstein, L. *et al*. Relation of N-terminal pro-brain natriuretic peptide levels and their prognostic power in chronic stable heart failure to obesity status. *Eur. Heart J.* **29**, 2634–2640 (2008).

38. Gastelurrutia, P. *et al*. Obesity paradox and risk of sudden death in heart failure results from the MUerte Subita en Insuficiencia cardiaca (MUSIC) study. *Am. Heart J.* **161**, 158–164 (2011).

39. Anand, I.S. *et al*. Prognostic value of baseline plasma amino-terminal pro-brain natriuretic peptide and its interactions with irbesartan treatment effects in patients with heart failure and preserved ejection fraction: findings from the I-PRESERVE trial. *Circ. Heart Fail.* **4**, 569–577 (2011).

40. Kubánek, M., Goode, K.M., Lánská, V., Clark, A.L. & Cleland, J.G.F. The prognostic value of repeated measurement of N-terminal pro-B-type natriuretic peptide in patients with chronic heart failure due to left ventricular systolic dysfunction. *Eur. J. Heart Fail.* **11**, 367–377 (2009).

41. Olsson, L.G. *et al*. Prognostic importance of plasma NT-pro BNP in chronic heart failure in patients treated with a beta-blocker: results from the Carvedilol Or Metoprolol European Trial (COMET) trial. *Eur. J. Heart Fail.* **9**, 795–801 (2007).

42. Patton, K.K., Sotoodehnia, N., DeFilippi, C., Siscovick, D.S., Gottdiener, J.S. & Kronmal, R.A. N-terminal pro-B-type natriuretic peptide is associated with sudden cardiac death risk: the Cardiovascular Health Study. *Heart Rhythm* **8**, 228–233 (2011).

43. Toggweiler, S. *et al*. NT-proBNP provides incremental prognostic information in cardiac outpatients with and without echocardiographic findings. *Clin. Cardiol.* **34**, 183–188 (2011).

44. Tsuchida, K. & Tanabe, K. Plasma brain natriuretic peptide concentrations and the risk of cardiovascular events and death in general practice. *J. Cardiol.* **52**, 212–223 (2008).

45. Masson, S. *et al*. Prognostic value of changes in N-terminal pro-brain natriuretic peptide in Val-HeFT (Valsartan Heart Failure Trial). *J. Am. Coll. Cardiol.* **52**, 997–1003 (2008).

46. Welsh, P. *et al*. Prognostic importance of emerging cardiac, inflammatory, and renal biomarkers in chronic heart failure patients with reduced ejection fraction and anaemia: RED-HF study. *Eur. J. Heart Fail.* **20**, 268–277 (2018).

47. Frantz, R.P. *et al*. Baseline and serial neurohormones in patients with congestive heart failure treated with and without bucindolol: results of the neurohumoral substudy of the Beta-Blocker Evaluation of Survival Study (BEST). *J. Cardiac. Fail.* **13**, 437–444 (2007).

48. Hartmann, F. *et al*. NT-proBNP in severe chronic heart failure: rationale, design and preliminary results of the COPERNICUS NT-proBNP substudy. *Eur. J. Heart Fail.* **6**, 343–350 (2004).

49. McMurray, J.J.V. *et al*. Angiotensin-neprilysin inhibition versus enalapril in heart failure. *N. Engl. J. Med.* **371**, 993–1004 (2014).

50. Anand, I.S. *et al*. Interaction between spironolactone and natriuretic peptides in patients with heart failure and preserved ejection fraction: from the TOPCAT trial. *JACC Heart Fail.* **5**, 241–252 (2017).

51. Pitt, B. *et al*. Spironolactone for heart failure with preserved ejection fraction. *N. Engl. J. Med.* **370**, 1383–1392 (2014).