ORIGINAL ARTICLE

Morphine Pharmacodynamics in Mechanically Ventilated Preterm Neonates Undergoing Endotracheal Suctioning

PA Välitalo[1], EHJ Krekels[1], M van Dijk[2,3], SHP Simons[3], D Tibboel[2] and CAJ Knibbe[1,2,4*]

To date, morphine pharmacokinetics (PKs) are well quantified in neonates, but results about its efficacy are ambiguous. This work presents an analysis of a previously published study on pain measurements in mechanically ventilated preterm neonates who received either morphine or placebo to improve comfort during invasive ventilation. The research question was whether morphine reduces the pain associated with endotracheal or nasal suctioning before, during, and after suctioning. Because these neonates cannot verbalize their pain levels, pain was assessed on the basis of several validated pain measurement instruments (i.e., COMFORT-B, preterm infant pain profile [PIPP], Neonatal Infant Pain Scale (NIPS), and visual analogue scale (VAS)). The item response theory (IRT) was used to analyze the data in order for us to handle the data from multiple-item pain scores. The analysis showed an intra-individual relationship between morphine concentrations and pain reduction, as measured by COMFORT-B and VAS. However, the small magnitude of the morphine effect was not considered clinically relevant for this intervention in preterm neonates.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?
☑ Morphine PKs in neonates are well characterized, and PD studies so far have shown ambiguous efficacy for pain caused by invasive ventilation. However, the lack of efficacy could be an artefact due to difficulty in assessing pain in neonates.

WHAT QUESTION DID THIS STUDY ADDRESS?
☑ Is morphine effective in treating pain related to invasive ventilation and endotracheal suctioning in preterm neonates?

WHAT THIS STUDY ADDS TO OUR KNOWLEDGE
☑ This model-based analysis captures data from different pain scales into a single latent variable, pain, using IRT, thereby increasing the signal/noise ratio. The same approach can be used as a framework for future analyses of pain and distress studies. The analgesic effect of morphine in mechanically ventilated neonates is quantified.

HOW MIGHT THIS CHANGE DRUG DISCOVERY, DEVELOPMENT, AND/OR THERAPEUTICS?
☑ The analysis shows how the observational clinical pain scales in neonates can be better utilized so that they are more informative about the clinically relevant endpoint, thereby allowing for a more detailed characterization of the drug concentration-effect relationship.

Morphine is used in hospitalized neonates for the treatment of moderate to severe pain, however, results on the efficacy of morphine to improve comfort during invasive ventilation in this population are ambiguous.[1–3] A between-group statistical analysis of data from two large, double-blind, placebo-controlled studies in the early 2000s suggested that morphine has no advantage over placebo in one study,[1] whereas the other suggested an analgesic effect at some but not all time points.[2] However, a subsequent population analysis of the concentration-effect relationship using the same data[2] led the authors of the latter study to conclude that morphine has no analgesic effect.[3]

The gold standard for pain assessment in adults is self-report, but preverbal children cannot express their pain this way. Therefore, dedicated pain scales have been developed and validated for assessment of pain in preterm neonates.

In total, over 40 different pain assessment tools exist for quantifying pain in neonates.[4–6]

We hypothesized that the observed lack of morphine efficacy could be due to pain being difficult to quantify in neonates. We recently published a study about the informativeness of items from two neonatal pain scales, the COMFORT scale[7] and the preterm infant pain profile (PIPP) score,[8] using item response theory (IRT) modeling.[9] A high variability was found in the ability of the items of both pain scales to discriminate between pain levels. This may have obscured the results of the original studies, which used total COMFORT and PIPP scores to assess pain on morphine efficacy.[1,2] A more powerful approach would be to consider item-level data, possibly from multiple pain scores, to derive a single latent variable using IRT modeling, and then examine the drug effects on the derived latent

[1]Division of Pharmacology, Leiden University, Leiden, The Netherlands; [2]Intensive Care and Department of Pediatric Surgery Erasmus MC Sophia Children's Hospital, Rotterdam, The Netherlands; [3]Department of Pediatrics, Division of Neonatology, Erasmus MC Sophia Children's Hospital, Rotterdam, The Netherlands; [4]Department of Clinical Pharmacy, St Antonius Hospital, Nieuwegein, The Netherlands. *Correspondence: CAJ Knibbe (c.knibbe@antoniusziekenhuis.nl)

Table 1 Mean pain scores of individual items before, during, and after suctioning

Suction status	COMFORT (1–5; video)	NIPS (0–7; nurse)	PIPP (0–3; video)	VAS (0–10; nurse)	VAS (0–10; video)
Before	2 (0.76; 3,100)	0.77 (1.4; 622)		1.1 (1.2; 715)	0.91 (0.93; 619)
During	3 (0.83; 3,147)	4.4 (2.2; 633)	1.1 (0.89; 1,067)	3 (2.2; 725)	2.6 (1.5; 630)
After	2 (0.74; 3,064)	0.7 (1.3; 615)		1.1 (1.2; 710)	0.82 (0.9; 610)

COMFORT data consists of five behavioral COMFORT items; NIPS data consist of one compound score; PIPP data consist of three behavioral items, and VAS data consist of one variable assessed either by a nurse or by the investigator on the basis of video observations. Data are mean (SD; no.).
NIPS, Neonatal Infant Pain Scale; PIPP, Premature Infant Pain Profile; VAS, Visual Analogue Scale.

variable. This way, items are not assumed to be equally sensitive in detecting pain. IRT has been proposed for pharmacodynamic (PD) studies as an attractive statistical method to derive a latent variable from ordered categorical data.[10]

The aim of the current study was to assess the morphine concentration-effect relationship in mechanically ventilated neonates, using a previously developed population pharmacokinetic (PK) model[11] to predict morphine concentrations and IRT based on multiple neonatal pain scores to quantify pain.

METHODS
Study design
The details of the study are reported in the original publication[1] and are briefly repeated here.

The data were obtained from 140 preterm neonates undergoing mechanical ventilation in a randomized double-blind, double-dummy trial comparing morphine ($n = 71$) vs. placebo ($n = 69$) in treatment of pain related to invasive ventilation and endotracheal suctioning. The neonates received a loading dose of either morphine (100 mcg/kg) or saline followed by a continuous infusion of either morphine (10 mcg/kg) or saline. Both groups were allowed open-label rescue morphine of 50 mcg/kg bolus dose followed by 5–10 mcg/kg/h infusion, if needed. Some neonates in the placebo group had received an earlier loading dose of morphine at the start of mechanical ventilation, prior to the start of the current study. Sparse samples for concentrations of morphine and morphine glucuronides was available for each patient. In total, 25 patients in the placebo group received no morphine at all. Ten patients of the original trial[1] were not included in the current study because they were extubated early, or there were other reasons to terminate the study early. The local ethics committees of the study centers had approved the original study protocol.

The mean bodyweight of the included patients was 1400 g (SD = 725 g) and mean postmenstrual age was 211 days (SD = 24.4 days). Twenty-nine neonates were "small for gestational age" (i.e., birthweight in the lowest 2.5 percentile for that gestational age; this definition differs from the World Health Organization definition, which uses the lowest 10th percentile). One hundred thirty-three of the patients were recruited at a postnatal age of 1 day, and the remaining seven at postnatal age of 2 or 3 days.

Pain was assessed at baseline, before administration of study medication, 30 minutes after the loading dose, and twice daily at standardized time points around endotracheal suctioning. The time points of endotracheal suctioning were not standardized. For each assessment, pain was recorded before, during, and after suctioning. Suctioning was performed more frequently (range, 0–24 procedures per day) than the pain assessments. Pain was quantified using COMFORT,[7] PIPP scores,[8] and visual analogue scale (VAS) scores,[12] assessed by two investigators based on video recordings.

COMFORT is an eight-item pain assessment tool, with six behavioral items (alertness, calmness/agitation, respiratory response, body movement, muscle tension, and facial tension) and two physiological items (heart rate and blood pressure), each graded between one and five. PIPP is a seven-item pain assessment scale, with three behavioral items (brow bulge, eye squeeze, and nasolabial furrow) graded between zero and three and four remaining items (gestational age, behavior before painful stimulus, heart rate, and oxygen saturation). For COMFORT and PIPP scores, individual item-level data were available, although the COMFORT item "muscle tension" could not be assessed based on the video recordings. Only behavioral items of these scores were included in the analysis because physiological items have been shown previously to be uninformative of pain in mechanically ventilated preterm infants.[9] Furthermore, nurse-assessed VAS scores and composite scores for the Neonatal Infant Pain Scale (NIPS)[13] (range, 0–7) obtained at the time of the study, were used in the analysis. A total of 16,257 item-level COMFORT, PIPP, NIPS, and VAS recordings were included in the current analysis. Mean scores of the data, along with numbers of observations, are summarized in **Table 1**.

Modeling procedures
All modeling procedures were conducted with NONMEM 7.3,[14] using the Laplace method, with GFortran version 4.8.4 and Perl-speaks-NONMEM version 4.4.0 for run control.[15] Data management and graphical visualization was performed using the R statistical software, version 3.1.1.[16]

The modeling was performed sequentially, as summarized in **Figure 1**. First, an IRT model was fitted to the available pain data, which is detailed in the subsection "graded response model for pain." Subsequently, the parameters of the graded response model were fixed to the final estimates to reduce runtimes.

Second, as detailed in the subsection "pharmacodynamic modeling," morphine and morphine glucuronide concentrations were tested as predictors of pain. A previously established population PK model was used to predict the morphine and morphine glucuronide concentrations at the
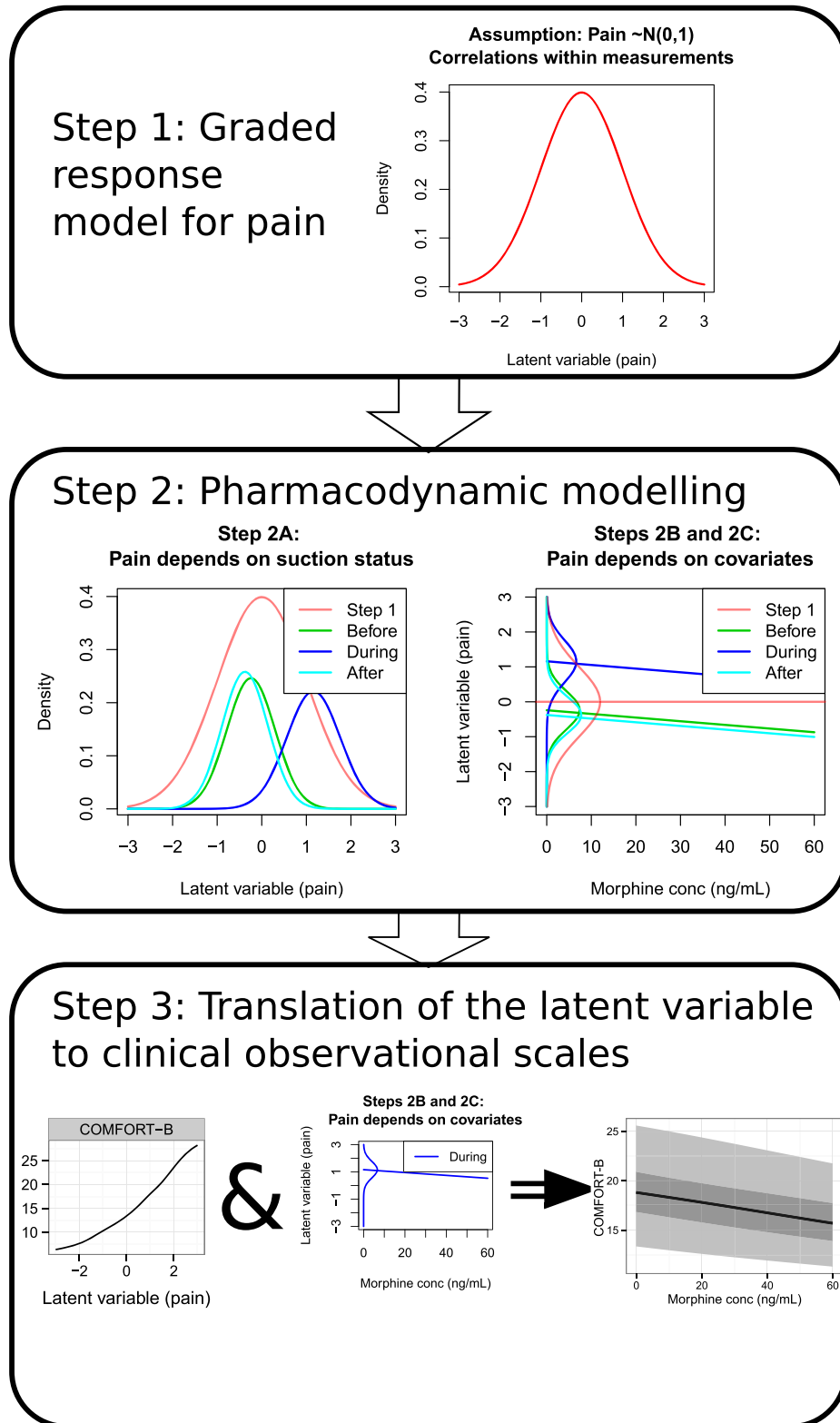
Figure 1 An overview of the modeling steps taken. The process starts by estimating the parameters for the graded response model to come up with a latent variable, pain (Step 1). Then, the latent variable is regressed on the basis of suctioning status, drug concentration, and other covariates (Step 2), whereas also showing the results of Step 1 as a comparison. Finally, the findings from Step 2 are transformed back to observational clinical scales (Step 3).

time of pain observations, on the basis of individual concentration-time data and postnatal age and body-weight.[11] Other tested covariates for pain include measures for bodyweight, age, and study time.

Third, model-predicted changes in the latent variable were statistically transformed back to the observational clinical scales, detailed in subsection "Translation of the Latent Variable to Observational Clinical Scales."

Graded response model for pain

A graded response model was used to model ordered categorical data of the pain scores in an IRT framework, predicting the probabilities of observing a specific grade for each of the items, as a function of pain. The pain measurement data consisted of five investigator-assessed COMFORT items, three investigator-assessed PIPP items, the nurse-assessed compound scores of NIPS, and VAS scores assessed both by the investigators and nurses.

Similar to ordered categorical data, for measurement i at suctioning state τ (before, during, and after suctioning), for item j, the probability of observation y being score k can be calculated on the basis of the latent variable $D_{i\tau}$ (i.e., pain), a discrimination parameter a_j and difficulty parameters b_{jk}[17] as outlined in Eq. 1. The item j refers to the subscales of COMFORT or PIPP, or the total score of NIPS.

$$p\left(y_{i\tau j} > k\right) = \frac{e^{a_j\left(D_{i\tau} - b_{jk}\right)}}{e^{a_j\left(D_{i\tau} - b_{jk}\right)} + 1} \tag{1}$$
$$p\left(y_{i\tau j} = k\right) = p\left(y_{i\tau j} > k-1\right) - p\left(y_{i\tau j} > k\right)$$

Where difficulty parameters b are the latent variable thresholds for 50% probability that the grade is higher than the subscripted value, and discrimination parameters a describe how well the item can separate different levels of pain.

The scale of the latent variable $D_{i\tau}$ is a relative one, therefore, its distribution needed to be defined. Within the nonlinear mixed-effects modeling context this was done by regarding this variable as a random effect[18] with a fixed standard normal distribution $N(0,1)$. For the purposes of fitting the graded response model, each measurement i was interpreted as an independent observation, so that all the observed variability in pain measurements is captured.

After fitting the graded response model, the model parameters were fixed and continuous data of the VAS scores was added. The relationship between the latent variable and the continuous VAS data was described with Eq. 2.

$$Pred_{vas,i\tau} = 10 \cdot \frac{e^{a_{vas}\left(D_{i\tau} - b_{vas}\right)}}{1 + e^{a_{vas}\left(D_{i\tau} - b_{vas}\right)}} \tag{2}$$

where the parameter a_{vas} is a steepness factor and the parameter b_{vas} is the threshold for the VAS score being 5. Separate parameters were estimated for nurse-assessed and investigator-assessed VAS scores. Additive and combined additive and logit residual error models were tested for the residual error of the VAS scores. A logit error model alone was unfeasible because boundary values of VAS

scores were present. Once the parameters for VAS data had been estimated, they were also fixed to their final values.

To compare model predictions to observed data, summaries of the model-predicted and empirical probabilities of any given grade k occurring as a function of the latent variable D_i were compared. The empirical probabilities were calculated by fitting an ordered categorical smoothing model for the grades as a function of the latent variable. The mgcv package[19] version 1.8-7 in R was used for this purpose.

Pharmacodynamic modeling

After defining the scale of the latent variable and mapping the parameters of the graded response model to this scale, the parameter values of the graded response model were fixed to allow the assessment of the influence of covariates on changes in the latent variable within individuals. For this PD modeling, measurements were treated as correlated within individuals.

The pain ($D_{hi\tau}$) for individual h at measurement i at state τ was described with Eq. 3.

$$D_{hi\tau} = f\left(\tau, \theta_\tau, \eta_{\tau,h}\right) - g\left(C_{hi}, \theta_{drug}\right) + h(cov_{hi}, \theta_{cov}) \tag{3}$$

where θ are vectors of fixed effects and η are vectors of random effects specific for individual h. The random effects were assumed normally distributed with means of zero and estimated SDs of ω_τ. Covariances between the random effects were estimated to account for intra-individual correlations between the measurements at different states. The modeling proceeded sequentially, and is described in detail in the following paragraphs.

The first part of the above equation, $D_{hi\tau} = f\left(\tau, \theta_\tau, \eta_{\tau,h}\right)$, is the model for baseline pain without drug effects or other covariate effects. Different population estimates (θ_τ) and individual estimates (η_τ) of baseline pain were estimated before, during, and after suctioning.

The second part of the above equation, $g\left(C_{hi}, \theta_{drug}\right)$, defines the effect of morphine, M3G, and M6G concentrations on pain. Using a previously developed morphine PK model for this population,[11] individual PK parameter values were obtained based on the individual morphine dosing schedule (including both scheduled doses and rescue doses), available concentration-time data for both morphine and the morphine glucuronides, and postnatal age and bodyweight of the patients. The individual PK parameters were subsequently used to predict the drug concentrations at the times of pain observations in the current study, which is known as the individual PK parameter method of sequential PK-PD data analysis.[20] Morphine and metabolite effects on pain were tested with a linear relationship, power function, and maximum effect (E_{max}) models with and without the Hill coefficient. The latter ones were tested only if the linear relationship was statistically significant. Finally, differential morphine effects before, during, and after suctioning were examined.

The third part of the overall model, $h(cov_{hi}, \theta_{cov})$, defines the remaining covariates for pain. Based on perceived potential clinical relevance, the following covariates were
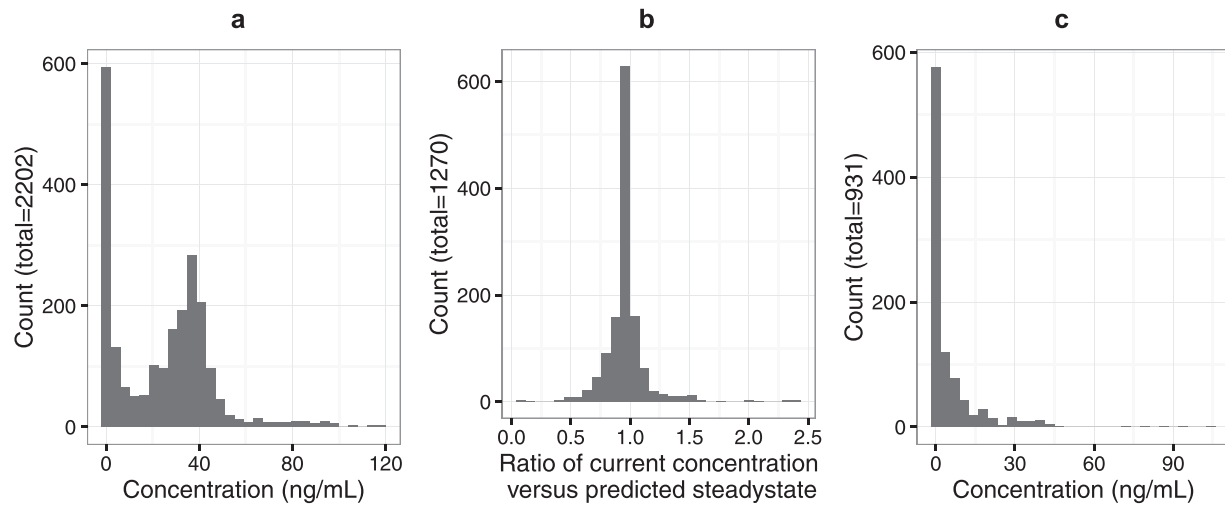
Figure 2 (**a**) A histogram of all morphine concentrations. (**b**) A histogram of the ratio between the predicted morphine concentration and the predicted morphine steady state at the time of pain assessment. (**c**) A histogram of morphine concentrations when pain was assessed but no infusion was ongoing (placebo group, for whom morphine rescue analgesia was allowed).

tested with a linear relationship: average number of suctioning events per day, birth bodyweight, gestational age, postnatal age, and study time, which was defined as the time after the first dose or as the time after the first pain recording in patients of the placebo group who did not receive rescue morphine before the first pain recording. "Small for gestational age" was tested as a categorical covariate.

Statistical testing was done based on Likelihood Ratio Test with a $P < 0.001$ criterion for significance. Additionally, we conducted an empirical assessment of significance levels for our study according to Wåhlby *et al.*[21] The details of this procedure are described in **Supplementary Electronic Material S1** "Assessment of actual significance levels." The procedure was used to verify that the objective function is truly χ^2 distributed.

Models were evaluated by assessing the precision of the parameters in terms of SEs. Nonparametric bootstrap with 1,000 resampled datasets was used to obtain an alternative set of confidence intervals for the model.[22] Given the similar number of patients per treatment group, the bootstrap was not stratified for this. Condition number, defined as the largest eigenvalue of the covariance matrix divided by the smallest eigenvalue of the covariance matrix, was used as a general diagnostic of model overparameterization.

Translation of the latent variable to observational clinical scores

To assess the clinical relevance of the covariate effects, the expected values of the scores for item j, with scores k and maximum score of K, were calculated as the sum of each score k times the probability of that score occurring.

$$E(y_j) = \sum_{k=1}^{K} p_{jk} \cdot k$$

where p_{jk} is the probability of kth grade occurring for item j, calculated with the graded response model. With this in

mind, the expected values of each item were calculated as a function of pain.

To calculate the expected value of total COMFORT-B scores in the absence of the COMFORT item "muscle tension," the sum of the expected values of the included COMFORT-B items was multiplied by 6/5, thus assuming that the grade for muscle tension is equal to the average score of the five other items. The same multiplication by 6/5 was done for the observed COMFORT-B items. Because, for the PIPP scale, only 3 of 7 items were included in the model, no expected value of the total score for this scale was calculated.

Changes in pain by drug concentrations and other covariates were visualized for total COMFORT-B score and VAS, because these are currently the most commonly used scales in clinical practice.

RESULTS

Individual predicted morphine concentrations at the time of pain assessment were mostly within the range of 0–60 ng/mL (**Figure 2a**) and, thus, our modeling results are restricted to that concentration range. We assessed how close morphine concentrations during infusion were to steady-state values at the times of the observational clinical measurement. This was done by comparing the predicted morphine concentration to the expected steady-state concentration given the current infusion rate. A histogram of concentrations in relation to the steady state of the current individual is presented in **Figure 2b**. This figure suggests the data may not be supportive of estimating a distribution delay in PD effects, because most of the concentrations are at or near steady state. Most of the nonsteady-state observations were from individuals who had morphine concentrations of ∼0 (**Figure 2c**).
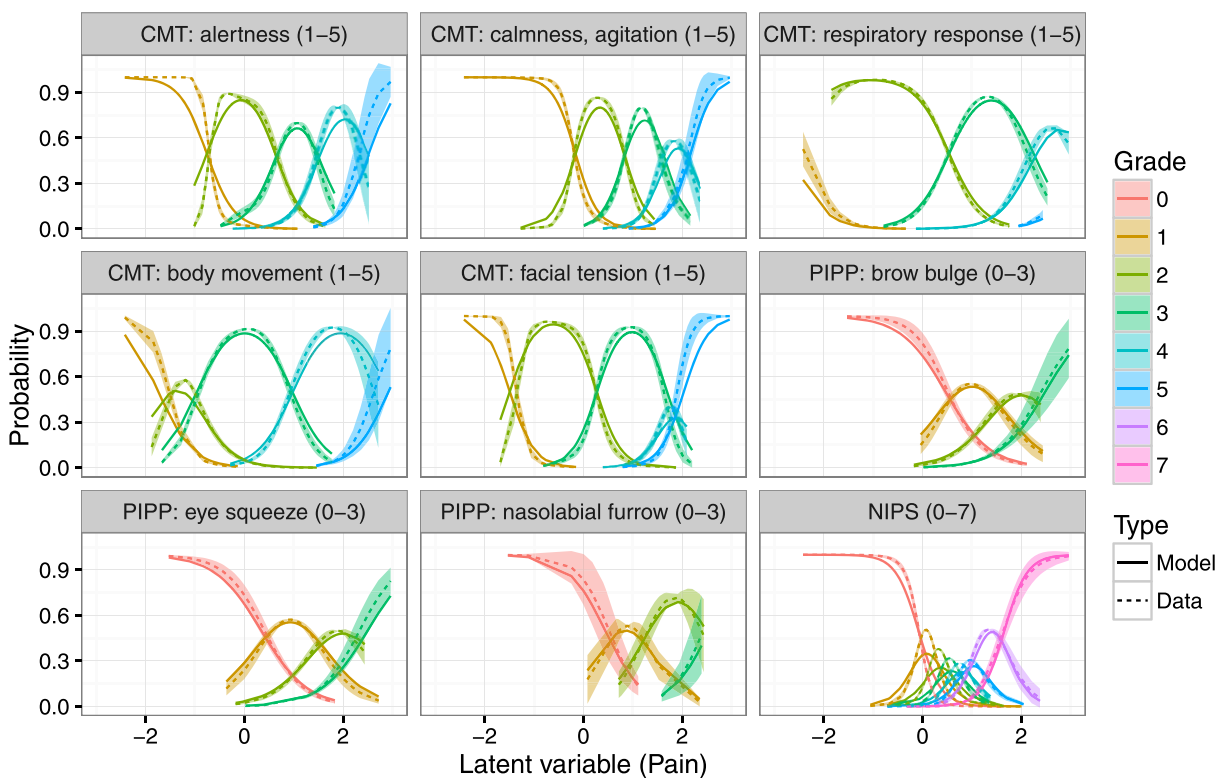
Figure 3 Item-grade probability curves. X-axis is the latent variable and Y-axis represents the probability of each grade occurring. Each color represents a specific grade, between 0 and 7. Solid lines are predictions from the graded response model and dashed lines with shaded areas are a smoothed mean and 95% confidence interval for the raw data. For COMFORT (abbreviated as CMT), colors representing grades from 1–5 are present, for Neonatal Infant Pain Scale (NIPS) color representing grades from 0–7 are present, and for preterm infant pain profile (PIPP) colors representing grades from 0–3 are present, according to the discrimination and difficulty values of the grades of these items.

Graded response model for pain

In the model for the latent variable pain, VAS scores were described with an additive residual error model, as the presence of VAS scores of 0 and 10 prevented the use of a logit-error model.

Figure 3 shows the model predictions of the probability for a given grade to occur (solid, colored lines), and the empirical probability densities of the grades (colored areas). Based on the agreement between the model predictions and the empirical probabilities, it can be concluded that the graded response model is able to describe pain as a latent variable.

The parameters relating to the graded response model and the model for VAS scores as a function of the latent variable are provided as **Supplementary Electronic Material S2**. A moderate variability was found in the discrimination parameters *a*, suggesting that the informativeness of the items is not equal.

Pharmacodynamic model

Morphine was found to decrease pain in a concentration-dependent manner, and increasing study time was associated with increased pain. The final model is described with the equation below and the final model parameter estimates are presented in **Table 2**. The SEs of the

parameters were low, and the condition number of 47.89 was well below the critical value of 1,000, demonstrating that the model is supported well by the data.[23]

$$D_{\tau hi} = f\left(\tau, \theta_\tau, \eta_{\tau,h}\right) - C_{hi,morph} \cdot \theta_{morph} + t_{hi} \cdot \theta_{time}$$

where the first part of the equation means that both the median and individual baseline pain state are different depending on the suctioning status τ.

The typical latent variable during suctioning was about one unit higher than the typical latent variable values before or after suctioning (**Table 2**). The correlation between random effects before and during suctioning was 54%, correlation during and after suctioning was 44%, and correlation before and after suctioning was 69%.

A linear model best described the relationship between morphine concentration and pain ($P < 0.001$, dOFV$=-72.91$). More complex concentration-effect relationships did not improve the model statistically significantly, nor were the morphine glucuronide concentrations found to have an effect on pain after the inclusion of morphine as a predictor. No delay, by means of an effect compartment model, could be estimated between morphine plasma concentration and effect. A model differentiating between morphine effects before, during, and after suctioning confirmed

Table 2 Parameter estimates from final model and bootstrap ($n = 1,000$)

Parameter	Symbol	Value (SE)	Bootstrap CI (5–95%)
Baseline pain, before suction	$\theta_{bl,1}$	−0.27 (0.0615)	−0.387 to −0.177
	$\omega_{bl,1}$	0.558 (0.0271)	0.483 to 0.636
Baseline pain, during suction	$\theta_{bl,2}$	1.15 (0.0581)	1.04 to 1.24
	$\omega_{bl,2}$	0.583 (0.0237)	0.514 to 0.65
Baseline pain, after suction	$\theta_{bl,3}$	−0.393 (0.0602)	−0.502 to −0.297
	$\omega_{bl,3}$	0.563 (0.0231)	0.494 to 0.634
Morphine effect slope $(ng/mL)^{-1}$	θ_{morph}	0.0091 (0.0016)	0.00567 to 0.0116
Study time slope (1/day)	θ_{time}	0.0476 (0.0185)	0.018 to 0.0782

The symbol θ denotes a fixed effect for a typical population estimate and the symbol ω denotes the magnitude of between-subject variability, expressed as a SD.
CI, confidence interval.

that the morphine effect is consistent across the three states, as the efficacy estimates θ_{morph} of morphine were largely similar regardless of suctioning status. A model with differential morphine effects did not result in significant improvement. Adding a random effect on morphine efficacy resulted in unacceptably high relative SE (71%) on the population estimate of morphine efficacy. As a sensitivity analysis, the pain data subsequent to rescue analgesia were excluded and the final model was rerun. In this case, the morphine effect became highly uncertain with a relative SE of 132%.

Of the other covariates, study time and postnatal age were associated with increased pain ($P < 0.001$, dOFV$= -105.54$). However, the two covariates are collinear because an overwhelming majority ($n = 133$; 95%) of the neonates entered the study within a postnatal age of one day. Of the two, study time was selected as the clinically most likely causal covariate and included in the model. The number of suctionings per day in each patient was not correlated to the baseline pain (θ_τ) of patients.

The effect of fixing IRT parameters values for the PD model is discussed in **Supplementary Electronic Material S3**, and the identifiability of the model without fixing these parameters is discussed in **Supplementary Electronic Material S4**. The PD model would be unidentifiable if the IRT model parameters were estimated together with the mean and variance of the latent variable.

Translation of the latent variable to observational clinical scales

The observed vs. predicted VAS and COMFORT-B scores before, during, and after suctioning are shown in **Figure 4**. A moderate level of agreement can be seen in individual predictions vs. observations, and a mild agreement between population predictions and observations.

Supplementary Figure S1 presents the expected values of individual pain grades as a function of the latent variable. **Figure 5** summarizes the expected values of COMFORT-B and VAS during suctioning as a function of morphine concentration and study time. As morphine concentrations increase from 0 to 20 ng/mL, the COMFORT-B score is expected to drop by one unit, and VAS score by <0.5 units, on average. As study day increases from 0 to 7, the COMFORT-B score is expected to increase by about two units, and VAS score by approximately one unit, on average. Although **Figure 5** shows the effects of morphine and

study time only during suctioning, the drug effects are similar before and after suctioning.

DISCUSSION

In this study, we have quantified the concentration-effect relationship of morphine in mechanically ventilated preterm neonates undergoing endotracheal suctioning. To our knowledge, this is the first analysis using item-level pain measurement data in a latent variable model to estimate analgesic morphine effects. In contrast to the statistical analysis of the original study,[1] we have detected an analgesic effect, although clinically insignificant, of morphine using advanced modeling techniques. Morphine PDs are notoriously hard to quantify, with large interindividual variability in morphine efficacy[24,25] and large variability in pain readings between individuals and between assessments, as observed in the current study.

We argue that the ability to detect a morphine effect in the current study results from the increased statistical power obtained by using advanced data analysis methods; nonlinear mixed-effects modeling together with IRT. The original analysis featured a statistical between-group comparison of NIPS, PIPP, and VAS scores in a morphine arm vs. placebo arm, whereas both treatment arms were allowed open-label morphine as rescue analgesia. The study was intended as a pragmatic trial to inform about choices between treatments,[1] and the original analysis was certainly best suited for that purpose. However, for the purposes of quantifying drug effects, concentration-effect modeling is more powerful as it relates drug concentrations to effects instead of relating treatment arm to group effects. Moreover, individuals who experience more pain are also more likely to receive rescue analgesia; therefore, a population-level comparison of drug concentrations vs. pain levels could appear to indicate that morphine causes pain. Using nonlinear mixed-effects modeling to separate interindividual variability in baseline pain from the intra-individual drug effects, we observe a small intra-individual drug effect (**Figure 5**). Furthermore, this drug effect cannot be quantified if pain data subsequent to rescue analgesia are omitted, which results in patients either having drug concentrations of zero (placebo arm) or steady-state morphine concentrations (treatment arm).
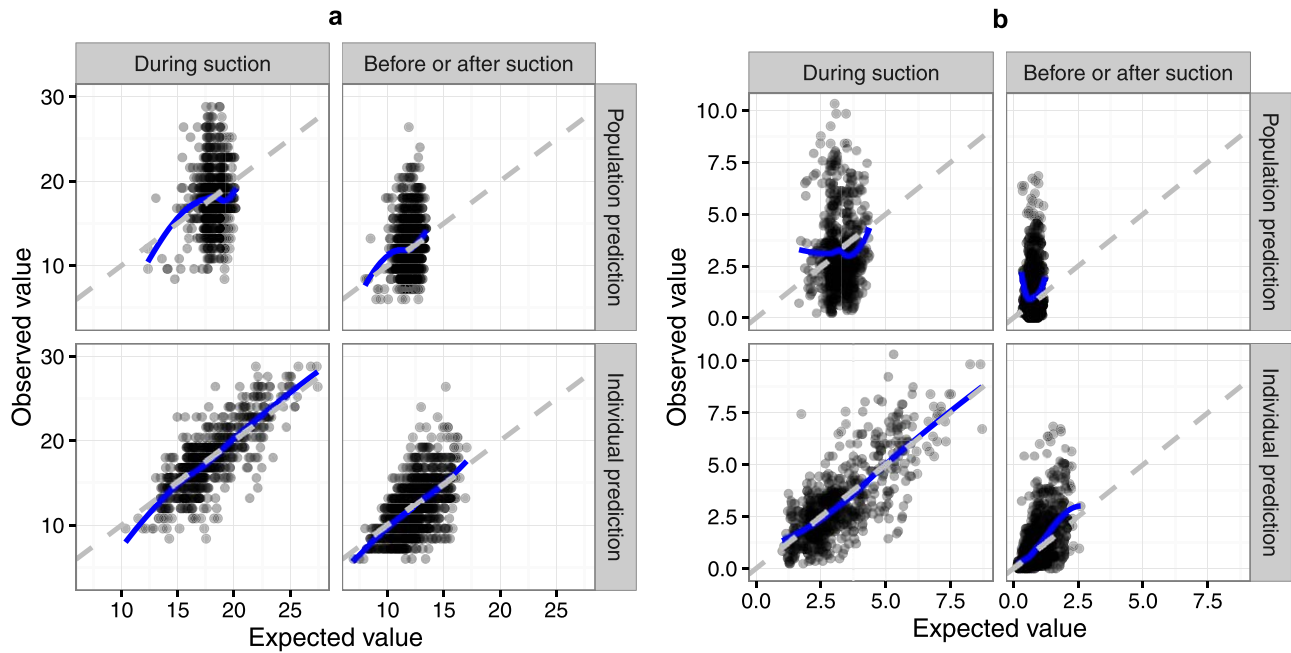
Figure 4 Observed vs. predicted plot for COMFORT-B (**a**) and visual analogue scale (VAS) (**b**) with population predictions in the upper rows and individual predictions in the lower rows. The gray dashed line is the line of unity and solid blue line is local regression smoothing. Because the pain levels were almost identical before and after suctioning, these two states were combined within this plot. To correct for the missing item "muscle tension," the observed and predicted sums of the existing items were multiplied by 6/5 to arrive at the total COMFORT-B score.
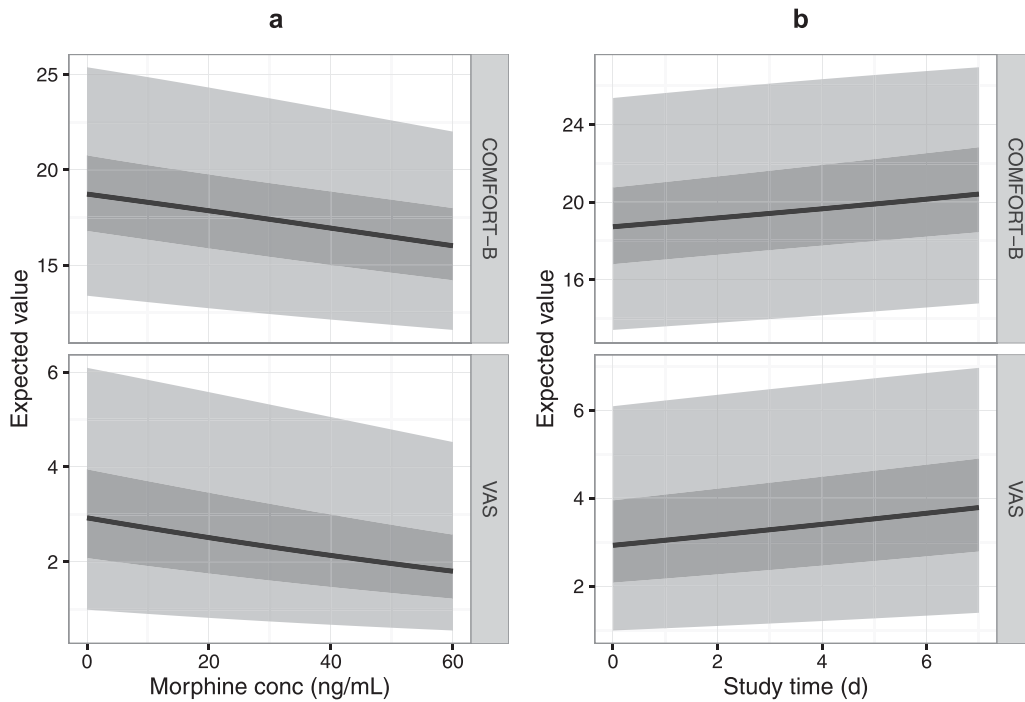

Figure 5 Expected values of COMFORT-B and visual analogue scale (VAS), vs. (**a**) morphine concentration and (**b**) time after study start. Solid black line is the population median value of pain, the dark shaded area is the 25–75% most commonly occurring pain range and light shaded area is the 2.5–97.5% most commonly occurring pain range, calculated based on the between-subject variability of pain during endotracheal suctioning.

A further increase in statistical power is gained by the use of IRT modeling, or latent variable modeling in general. We recently showed that the informativeness of the PIPP and COMFORT individual items is highly variable.[9] Therefore, if the items are summed together, some of the items may add more noise than information to the analysis. On the other hand, if the data are analyzed with a latent variable model, each item affects the objective function differently, which can be interpreted as the item being "weighted" by its informativeness. Differences in the discrimination parameters *a* between the included items (**Supplementary Electronic Material S2**), suggest that the items within the dataset have differing sensitivity to pain and their contribution to the model fit are scaled accordingly. Latent variable modeling also enables the use of observation sets with one or more items missing, thus enabling more comprehensive use of the available data.

Morphine target concentrations of around 20 ng/mL have been reported[26,27] for postoperative pain in neonates, concentration well above this value have been observed in the current analysis. On average, morphine concentration increasing from 0–20 ng/mL decreases total COMFORT-B scores by approximately one grade, and VAS scores by approximately one-half a grade (**Figure 5a**). Considering that the average COMFORT-B grade is 18 during suctioning and 12 before/after suctioning (**Table 1**), the applied morphine-dosing algorithm does not yield concentrations that provide clinically relevant analgesia that would justify its use for suctioning-related acute pain. Given the linear relationship between morphine concentration and effect, extrapolations of our findings to higher concentrations are not possible, moreover, higher concentrations would likely yield an increase in undesirable side effects. It should also be noted that the current results of morphine efficacy cannot be generalized to other, potentially more severe types of pain experienced by children, such as postoperative pain or pain caused by necrotizing enterocolitis.

The pain scales used in our analysis are validated, widely accepted, and show good inter-rater reliability.[7,8,12,28–30] However, it should be acknowledged that the items included in this analysis are measuring the signals of pain and distress from the neonate, not the absolute pain itself.[31] The observed signals can be affected (i) by the pain experienced by the neonate, (ii) by the distress experienced by the neonate, and (iii) by the ability of the neonate to express pain and distress. As such, it is possible that the modest association between study time and pain (**Figure 5b**) results from antenatal maturation of the neonates, and them becoming more able to express pain. However, it is also possible that the suctioning procedure gradually becomes more painful as it is repeated many times per day over the duration of the study.

In conclusion, we have built a model to describe and predict pain in mechanically ventilated preterm neonates undergoing endotracheal suctioning, based on several validated pain measures for this patient population. Using these, we have constructed a latent variable to represent pain, and performed a PD analysis on the latent variable. IRT enabled us to quantify a mild analgesic effect of morphine in mechanically ventilated preterm neonates, however, this analgesic effect

is not enough to advocate routine morphine use in these patients.

Additional Supporting Information may be found in the online version of this article.

Author Contributions: P.A.V., E.H.J.K., M.v.D., S.H.P.S., D.T., and C.A.J.K. wrote the manuscript. P.A.V., E.H.J.K., S.H.P.S., D.T., and C.A.J.K. designed the research. P.A.V., M.v.D., S.H.P.S., and D.T. performed the research. P.A.V. analyzed the data.

1. Simons, S.H. *et al.* Routine morphine infusion in preterm newborns who received ventilatory support: a randomized controlled trial. *JAMA* **290**, 2419–2427 (2003).
2. Anand, K.J. *et al.* Effects of morphine analgesia in ventilated preterm neonates: primary outcomes from the NEOPAIN randomised trial. *Lancet* **363**, 1673–1682 (2004).
3. Anand, K.J. *et al.* Morphine pharmacokinetics and pharmacodynamics in preterm and term neonates: secondary results from the NEOPAIN trial. *Br. J. Anaesth.* **101**, 680–689 (2008).
4. Hummel, P. & van Dijk, M. Pain assessment: current status and challenges. *Semin. Fetal. Neonatal Med.* **11**, 237–245 (2006).
5. Ranger, M., Johnston, C.C. & Anand, K.J. Current controversies regarding pain assessment in neonates. *Semin. Perinatol.* **31**, 283–288 (2007).
6. Maxwell, L.G., Malavolta, C.P. & Fraga, M.V. Assessment of pain in the neonate. *Clin. Perinatol.* **40**, 457–469 (2013).
7. Ambuel, B., Hamlett, K.W., Marx, C.M. & Blumer, J.L. Assessing distress in pediatric intensive care environments: the COMFORT scale. *J. Pediatr. Psychol.* **17**, 95–109 (1992).
8. Stevens, B., Johnston, C., Petryshen, P. & Taddio, A. Premature Infant Pain Profile: development and initial validation. *Clin. J. Pain* **12**, 13–22 (1996).
9. Välitalo, P.A. *et al.* Pain and distress caused by endotracheal suctioning in neonates is better quantified by behavioural than physiological items: a comparison based on item response theory modelling. *Pain* **157**, 1611–1617 (2016).
10. Ueckert, S. *et al.* Improved utilization of ADAS-cog assessment data through item response theory based pharmacometric modeling. *Pharm. Res.* **31**, 2152–2165 (2014).
11. Knibbe, C.A. *et al.* Morphine glucuronidation in preterm neonates, infants and children younger than 3 years. *Clin. Pharmacokinet.* **48**, 371–385 (2009).
12. van Dijk, M., Koot, H.M., Saad, H.H., Tibboel, D. & Passchier, J. Observational visual analog scale in pediatric pain assessment: useful tool or good riddance? *Clin. J. Pain* **18**, 310–316 (2002).
13. Lawrence, J., Alcock, D., McGrath, P., Kay, J., MacMurray, S.B. & Dulberg, C. The development of a tool to assess neonatal pain. *Neonatal Netw.* **12**, 59–66 (1993).
14. Beal, S.L., Sheiner, L.B., Boeckmann, A.J. & Bauer, R.J. NONMEM Users Guide – part III. <nonmem.iconplc.com> (2013).
15. Lindbom, L., Pihlgren, P. & Jonsson, E.N. PsN-Toolkit–a collection of computer intensive statistical methods for non-linear mixed effect modeling using NONMEM. *Comput. Methods Programs Biomed.* **79**, 241–257 (2005).
16. R Development Core Team R. A language and environment for statistical computing. (Vienna, Austria). <http://www.R-project.org> (2014).
17. Samejima, F. Estimation of latent ability using a response pattern of graded scores. *Psychometrika* **1968**, 1–169 (1968).
18. Rijmen, F., Tuerlinckx, F., De Boeck, P. & Kuppens, P. A nonlinear mixed model framework for item response theory. *Psychol. Methods* **8**, 185–205 (2003).
19. Wood, S.N. Generalized Additive Models: An Introduction with R (Chapman and Hall/CRC Press, 2006).
20. Lacroix, B.D., Friberg, L.E. & Karlsson, M.O. Evaluation of IPPSE, an alternative method for sequential population PKPD analysis. *J. Pharmacokinet. Pharmacodyn.* **39**, 177–193 (2012).
21. Wählby, U., Jonsson, E.N. & Karlsson, M.O. Assessment of actual significance levels for covariate effects in NONMEM. *J. Pharmacokinet. Pharmacodyn.* **28**, 231–252 (2001).
22. Yafune, A. & Ishiguro, M. Bootstrap approach for constructing confidence intervals for population pharmacokinetic parameters. I: a use of bootstrap standard error. *Stat. Med.* **18**, 581–599 (1999).
23. Montgomery, D.C., Peck, E.A. & Vining, G. Introduction to linear regression analysis, 5th ed. (Wiley, Hoboken, NJ, 1982).
24. Sverrisdóttir, E., Lund, T.M., Olesen, A.E., Drewes, A.M., Christrup, L.L. & Kreilgaard, M. A review of morphine and morphine-6-glucuronide's pharmacokinetic-pharmacodynamic relationships in experimental and clinical pain. *Eur. J. Pharm. Sci.* **74**, 45–62 (2015).

248

25. Juul, R.V. *et al.* A pharmacokinetic-pharmacodynamic model of morphine exposure and subsequent morphine consumption in postoperative pain. *Pharm. Res.* **33**, 1093–1103 (2016).

26. Olkkola, K.T., Maunuksela, E.L., Korpela, R. & Rosenberg, P.H. Kinetics and dynamics of postoperative intravenous morphine in children. *Clin. Pharmacol. Ther.* **44**, 128–136 (1988).

27. Lynn, A.M., Nespeca, M.K., Bratton, S.L. & Shen, D.D. Intravenous morphine in postoperative infants: intermittent bolus dosing versus targeted continuous infusions. *Pain* **88**, 89–95 (2000).

28. van Dijk, M., de Boer, J.B., Koot, H.M., Tibboel, D., Passchier, J. & Duivenvoorden, H.J. The reliability and validity of the COMFORT scale as a postoperative pain instrument in 0 to 3-year-old infants. *Pain* **84**, 367–377 (2000).

29. Stevens, B., Johnston, C., Taddio, A., Gibbins, S. & Yamada, J. The premature infant pain profile: evaluation 13 years after development. *Clin. J. Pain* **26**, 813–830 (2010).

30. Taddio, A., O'Brien, L., Ipp, M., Stephens, D., Goldbach, M. & Koren, G. Reliability and validity of observer ratings of pain using the visual analog scale (VAS) in infants undergoing immunization injections. *Pain* **147**, 141–146 (2009).

31. Pillai Riddell, R., Fitzgerald, M., Slater, R., Stevens, B., Johnston, C. & Campbell-Yeo, M. Using only behaviours to assess infant pain: a painful compromise? *Pain* **157**, 1579–1580 (2016).

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