Development and evaluation of a gentamicin pharmacokinetic model that facilitates opportunistic gentamicin therapeutic drug monitoring in neonates and infants.

Authors: Eva Germovsek,[a,#] Alison Kent,[b] Tuuli Metsvaht,[c] Irja Lutsar,[c] Nigel Klein,[a] Mark A Turner,[d] Mike Sharland,[b] Elisabet I Nielsen,[e] Paul T Heath,[b] Joseph F Standing[a] and the neoGent collaboration*

Affiliations

Inflammation, Infection and Rheumatology section, Institute of Child Health, University College London, London WC1N 1EH, UK[a]; Paediatric Infectious Diseases Research Group, Institute of Infection and Immunity, St George's, University of London, Cranmer Terrace, London SW17 0RE, UK[b]; Department of Microbiology, University of Tartu, Tartu, Estonia[c]; Liverpool Women's Hospital, Crown Street, Liverpool, Merseyside L8 7SS, UK[d]; Department of Pharmaceutical Biosciences, Uppsala University, Uppsala, Sweden[e]

*Membership of the neoGent collaboration is provided in the Acknowledgments.

 # Contact details for the corresponding author

Telephone: 020 7905 2307

Fax: 020 7905 2882

E-mail: eva.germovsek.11@ucl.ac.uk

A short running title: Gentamicin PK model for TDM in neonates and infants

Key words: aminoglycoside antibiotics, PK modelling, NLME, infection, NONMEM

Abstract

Trough gentamicin therapeutic drug monitoring (TDM) is time-consuming, disruptive to neonatal clinical care and a patient safety issue. Bayesian models could allow TDM to be performed opportunistically at the time of routine blood tests. This study aimed to develop and prospectively evaluate a new gentamicin model and a novel Bayesian computer tool (neoGent) for TDM use in neonatal intensive care. We also evaluated model performance for predicting peak concentrations and AUC(0-t). A pharmacokinetic meta-analysis was performed on pooled data from three studies (1325 concentrations from 205 patients). A 3-compartment model was used with covariates being: allometric weight scaling, postmenstrual and postnatal age, and serum creatinine. Final parameter estimates (standard error) were: clearance: 6.2 (0.3) L/h/70kg; central volume (V) 26.5 (0.6) L/70kg; inter-compartmental disposition: Q=2.2 (0.3) L/h/70kg, V2=21.2 (1.5) L/70kg, Q2=0.3 (0.05) L/h/70kg, V3=148 (52.0) L/70kg. The model's ability to predict trough concentrations from an opportunistic sample was evaluated in a prospective observational cohort study that included data from 163 patients with 483 concentrations collected in five hospitals. Unbiased trough predictions were obtained: median (95% confidence interval (CI)) prediction error was 0.0004 (-1.07, 0.84) mg/L. Results also showed peaks and AUC(0-t) could be predicted (from one randomly selected sample) with little bias but relative imprecision with median (95% CI) prediction error being 0.16 (-4.76, 5.01) mg/L and 10.8 (-24.9, 62.2) mg h/L, respectively. NeoGent was implemented in R/NONMEM, and in the freely available TDMx software.

Introduction

The aminoglycoside antibiotic gentamicin is the most commonly used antimicrobial on neonatal units(1, 2) and is effective against Gram negative bacteria. Gentamicin use is limited by its narrow therapeutic index and risk of toxicity, specifically nephro- and ototoxicity(3). It is not metabolized in the liver(4) and is almost entirely eliminated by the kidneys; clearance therefore depends on renal function. During the first two weeks of life, renal and intra-renal blood flow increase rapidly, causing a steep rise in glomerular filtration rate (GFR)(5, 6).

Therapeutic drug monitoring (TDM) is required to ensure maximal efficacy and especially minimal toxicity, particularly in the neonatal population where variability in pharmacokinetic (PK) parameters is large. Dose individualization approaches focus on toxicity(7, 8) and include single-level methods and nomograms(9, 10), area under the curve (AUC) methods(11), and Bayesian methods(12). The use of nomograms is limited as they cannot readily incorporate covariates affecting PK parameters. AUC methods use a simplified 1-compartment PK model and require at least two gentamicin measurements, which is not appropriate in neonates with limited blood volumes. These drawbacks make Bayesian approaches the most attractive for newborn infants.

Deriving a Bayesian prior for TDM requires a non-linear mixed-effect PK model, and several such studies of neonatal gentamicin have been published(13-24). However, these studies are limited by their heterogeneity and use of sparse data (often identifying only a 1-compartment model when gentamicin follows multi-compartment kinetics(25, 26)) and fail to account for age-related differences in creatinine during the immediate newborn period. Although gentamicin is not a new drug, its dosing and monitoring is still a current issue as identified in the UK National Patient Safety alert (http://www.nrls.npsa.nhs.uk/alerts/?entryid45=66271) and a recent publication by Valitalo *et al*(27), who used simulations to define dosing guidelines.

We aimed to investigate whether opportunistic sampling can predict trough gentamicin concentrations so that standard TDM could be performed from a blood sample taken for other purposes (e.g. routine blood gases). As a secondary aim, we evaluated the model's ability to predict peak gentamicin concentrations and AUC(0-t) using one randomly selected sample.

Methods

Study population

This study used two datasets: a model-building dataset and a prospectively collected evaluation dataset.

To collect data for model development, the electronic bibliographic database PubMed was searched in January 2015 without time limitations. The search strategy included: (neonat* OR newborn*) AND (gentamicin) AND (pharmacokinetic* OR PK); gentamicin samples had to be prospectively collected and covariates (weight, gestational age (GA), postnatal age (PNA), serum creatinine measurements), also had to be reported. Additionally, we also searched the reference lists in identified papers. The authors of the publications that met the inclusion criteria (n=8) (11, 15, 21, 22, 28-31) were then invited to contribute their data.

Data for the evaluation of the PK model were collected as a prospective observational cohort study from five UK hospitals (St George's University Hospitals NHS Foundation Trust, Liverpool Women's NHS Foundation Trust, Oxford University Hospitals, Portsmouth Hospitals NHS Trust and Coventry & Warwickshire University Hospitals NHS Trust) from July 2012 to November 2013. Infants were eligible for inclusion if the following criteria were met: more than 36 hours gentamicin therapy anticipated, postnatal age of less than 90 days, not receiving extracorporeal membrane oxygenation, peritoneal dialysis or hemofiltration, and expected to survive the study period (as judged by the clinical team). Each patient provided a minimum of two gentamicin concentrations – a trough sample from routine TDM (i.e. a pre-dose sample taken before a non-initial dose) and an additional study sample (taken opportunistically during a course of gentamicin when the infant required blood sampling for clinical care). These samples will be referred to as routine (trough) and (opportunistic) study samples in this manuscript. Exact times of gentamicin dosing and sampling were recorded, along with the patient's weight, age and serum creatinine (Table 1). Written informed consent was obtained from parents and the study was approved by the London Central Ethics committee (reference 12/LO/0455).

Gentamicin dosing and sampling procedure in the prospective evaluation dataset

Gentamicin treatment was initiated at the discretion of the clinical team for possible infection and dosed and monitored using trough concentrations according to the standard practice at each hospital. Gentamicin was administered as a slow (<2 min) bolus via intravenous cannula, percutaneous long line, or umbilical venous catheter.

Bioanalytical techniques

An enzyme immunoassay (EMIT, Syva)(15), a fluorescence polarization immunoassay (TDx, Abbot)(15, 21), and high performance liquid chromatography coupled to tandem mass spectrometry (UHPLC-MS/MS) (32) were used to determine gentamicin concentration in the model-building dataset; and the Jaffe reaction (33) was used to determine serum creatinine concentrations. In the prospective evaluation dataset, gentamicin serum concentrations were analyzed using immunoassay techniques (Table S1); and creatinine concentrations were determined by either a Jaffe-based or an enzymatic method (137 neonates and 26 neonates, respectively).

Pharmacokinetic analysis

The observed concentration-time data from only the model-building studies were pooled and simultaneously analyzed with non-linear mixed-effects software NONMEM version 7.3(34). The first order conditional estimation method with interaction was used.

Basic model

One-, 2-, and 3-compartment structural models were considered when defining the basic structural population PK model. The inter-individual variability (IIV) was assumed to follow a log-normal distribution and tested on all parameters. An additive, a proportional, and a combination of both (Equation 1) residual error models were tested.

$$y_{ij} = f(t_{ij};\, \phi_i) + f(t_{ij};\, \phi_i) \cdot \varepsilon_{ij(proportional)} + \varepsilon_{ij(additive)}, \qquad \text{(Equation 1)}$$

where y_{ij} is an observed gentamicin concentration at time t_{ij}, f is the function that represents the gentamicin model, ϕ_i is a vector of parameters, ε_{ij} is a residual error term.

Inter-occasion variability (IOV) was also assumed to be log-normally distributed and it was tested for all parameters with an occasion defined as a single dosing interval.

Covariate model

Allometric scaling was used *a priori* to standardize all PK parameters to 70 kg (35), and a maturation function, describing the maturation of the GFR with postmenstrual age (PMA) (Equation 2) with fixed parameters from a previous study (5), was used to scale clearance. Allometric exponents were fixed to 0.632 for central clearance and 0.75 for inter-compartmental clearances. Different exponents were used because these values were shown best for describing the maturation of renal elimination(5) and tissue blood flows(36), respectively. Allometric exponents for volumes of distribution were fixed to 1. The combination of allometric weight scaling and sigmoidal maturation function was suggested as a standard method for scaling clearance in the pediatric population in a recent comparison of different approaches(37).

$$maturation\ function = \frac{PMA^{Hill}}{PMA_{50}^{Hill}+PMA^{Hill}}, \quad \text{(Equation 2)}$$

where *Hill* is the sigmoidicity coefficient and PMA_{50} is PMA when maturation of GFR reaches 50% of adult values.

As it is known that PNA and serum creatinine are important indicators of gentamicin clearance and also based on the posthoc estimates of etas versus covariates plots, they were tested on clearance. These time-varying covariates were considered to significantly improve the fit and therefore included in the model if the difference in objective function value (ΔOFV) after their inclusion was >3.84 (p<0.05). Additionally, linear extrapolations between observations were made. To account for endogenous creatinine, maternal creatinine and also the change in renal function with age, a typical value of serum creatinine (TSCr) for a specific PMA was determined using data from Cuzzolin *et al*(38) for preterm (GA<37 weeks) newborns and Rudd *et al*(39) for term newborns. A linear decline in TSCr with increasing PMA was found according to Equation 3:

$$TSCr = -2.849 \cdot PMA\ (weeks) + 166.48. \quad \text{(Equation 3)}$$

A possible influence of serum creatinine on clearance was tested according to the following Equation 4, where measured serum creatinine (MSCr) was standardized by TSCr for PMA and departures from it estimated as follows:

$$\left(\frac{MSCr}{TSCr}\right)^{\theta}. \quad \text{(Equation 4)}$$

The effect of PNA was investigated with a logistic function (Equation 5) to account for the rapid changes in gentamicin clearance in the first hours of life. The first day of life was defined as day 1.

$$postnatal\ age\ function = \frac{PNA}{PNA_{50}+PNA}, \quad \text{(Equation 5)}$$

where PNA_{50} is the PNA when clearance has reached 50% of typical adult's clearance.

After the forward selection (ΔOFV>3.84) of all covariates (full model), backward elimination was performed, with a *p*-value retention cut-off of 0.001 (ΔOFV<10.83).

Evaluation

Internal model evaluation

Basic goodness-of-fit plots for observations *versus* population and individual predictions, conditional weighted residuals *versus* population predictions and *versus* time after dose were produced using statistical software R version 3.1.0 (R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from: http://www.R-project.org/) and visually examined. The assumptions of normality and homogeneity of the residuals errors were investigated by inspecting a histogram and a qq-plot.

Standard errors from NONMEM covariance step and non-parametric bootstrap analysis with 1,000 replicates were used to determine the precision of the final PK parameter estimates.

Additionally, we simulated 1,000 datasets using parameter estimates from the final model, and plotted 95% confidence intervals (CI) around the 2.5th, 50th, and 97.5th prediction percentiles of the simulated data. Then, the observations were overlaid on the plot, also called the visual predictive check (VPC). Perl-speaks-NONMEM (PsN) software(40) was used for the bootstrap analysis and to produce the VPC, which was visualized using R-package Xpose4(41).

External model evaluation

The prospective evaluation dataset was used to evaluate the predictive performance of the model. No additional fitting was done, and the diagnostic plots and the VPC were generated as described above. Bayesian model-predicted trough concentrations were computed using the model as a prior and information from only the opportunistic study samples. These predictions were compared with the observed trough concentrations by calculating the prediction error (PE) (42), and also the mean PE (MPE) (i.e. a measure of bias), and root-mean-square error (RMSE), a measure of precision(43) (Equations 6).

$$PE = observed - predicted$$

$$MPE = \frac{1}{N} \cdot \sum_{i=1}^{N} \cdot PE_i \qquad \text{(Equations 6)}$$

$$RMSE = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^{N} \cdot PE_i^2}$$

Also, we counted the number of "correct" predictions that were below or above the currently recommended gentamicin trough concentration thresholds of 1 mg/L or 2 mg/L (the National Institute for Health and Care Excellence (NICE) (http://www.nice.org.uk/guidance/CG149/chapter/1-Guidance#therapeutic-drug-monitoring-for-gentamicin) and British National Formulary for Children (BNFc) (http://www.evidence.nhs.uk/formulary/bnfc/current/5-infections/51-antibacterial-drugs/514-aminoglycosides/gentamicin)).

Further analysis of paired samples (that is both study and routine samples taken in the same dosing interval) was undertaken for the following scenarios: study samples ≥1, ≥2, and ≥3 mg/L, compared with only unpaired samples.

Cross-validation

The subset with the study sample above 3 mg/L provided the most important comparison, since in this case the study sample was still above the pre-specified trough threshold. As there were only 18 pairs with opportunistic study concentration ≥3 mg/L in the evaluation dataset, these pairs were merged with paired samples of the same characteristics from the model-building dataset. The pooled dataset

was then randomly split into five subsets, and cross-validation was performed; meaning that in each subset 20% of the pairs were randomly removed and the model was re-estimated. The re-estimated model was then used as a prior to predict the troughs, and compared to the observed trough concentrations as previously described.

Whether the model is able to predict peak concentrations from one randomly selected non-peak sample was tested similarly as described above, using paired samples from both the model-building and the evaluation dataset, and performing cross-validations. Additionally, as a possible pharmacokinetic-pharmacodynamic target for aminoglycosides can also be AUC(0-24)/MIC (44), the model was also evaluated on how it predicts AUC(0-t). Only a subset of the data where five or more samples were collected after the same dose was used for defining AUC(0-t), and the model-predicted *versus* observed (non-compartmental) AUC(0-t) was compared.

Comparison with other models

To compare our mechanistic model which scales for size, age and expected renal function with previously published models using empirical covariate analysis, predictions for the measured trough from the routine opportunistic samples in our prospective dataset were generated.

neoGent software

The model was implemented using R and NONMEM (see Supplementary material). It works by reading an individual's data into R, then Bayesian estimates generated in NONMEM are used to predict outcomes of interest (e.g. the time when the concentration falls below 2 mg/L).

Results

Patients

Out of eight contacted authors identified in the literature search we obtained two large neonatal gentamicin datasets (15, 21). We received no response from four authors (11, 28-30); and although an initial response was received from two authors (22, 31) no data were actually shared. Additionally, we obtained some previously unpublished data taken during a PK study of ampicillin and penicillin (32). The data were pooled and comprised 1325 gentamicin concentrations from 205 neonates (Table 1). This dataset was used to derive the model.

For the model evaluation, gentamicin serum concentrations were prospectively collected from a total of 194 neonates. Of the enrolled patients, 163 were included in the PK analysis (Table 1). Reasons for exclusion (31 patients) included inexact sampling times, insufficient samples, or the gentamicin opportunistic study concentration being below the limit of quantification (n=12). The final evaluation dataset comprised 483 gentamicin serum measurements, with 229 study and 254 routinely taken trough concentrations. Median (range) time after dose was 13.3 (0.08-53.3) h and 31.1 (8.0-79.7) h for study and routine concentrations, respectively. Patients were on treatment for up to 20 days.

Pharmacokinetic analysis

Initially, a 2-compartment model provided a better fit to the data (ΔOFV=7.4 with a 3-compartment model) and was therefore chosen as the basic structural model. But, after the addition of the fixed allometric and renal function parameters, covariates and IOV, a 3-compartment model described the data better (47-unit drop in OFV). The IIV was described with an exponential error structure, and the best residual error model was a combination of a proportional and additive error.

Postnatal age and standardized serum creatinine had a significant effect on clearance (ΔOFV=134.1 and ΔOFV=17.2, respectively) and were thus included in the final model. Backward elimination (p=0.001) confirmed that these covariates remained significant with the 3-compartment model. The final gentamicin population PK model is summarized with Equations 7.

$$CL = \theta_{CL} \cdot \left(\frac{WT}{70}\right)^{0.632} \cdot \frac{PMA^{3.33}}{55.4^{3.33}+PMA^{3.33}} \cdot \left(\frac{MSCr}{TSCr}\right)^{\theta_{SCr}} \cdot \frac{PNA}{\theta_{P_{50}}+PNA} \cdot e^{(\eta_{CL}+\kappa_{CL})},$$

$$V = \theta_V \cdot \left(\frac{WT}{70}\right) \cdot e^{\eta_V}, \qquad \text{(Equations 7)}$$

$$Q = \theta_Q \cdot \left(\frac{WT}{70}\right)^{0.75} \cdot e^{\eta_Q},$$

where *CL* is gentamicin clearance, *V* is gentamicin volume of distribution, *Q* is inter-compartmental gentamicin clearance, *WT* is body weight in kilograms, η is IIV, κ is IOV.

There was only a small improvement in fit (ΔOFV=7.6) when the model was parameterized for time-varying covariates (linear extrapolation between observed covariate values), but as this model is more biologically plausible, it was chosen as the final model.

The OFV reduced from 2305.0 to 1217.5 between the basic and the final model. The inclusion of the covariates resulted in a reduction of the IIV on PK parameters: with the basic model the IIV on CL and V was 71.1% and 62.5%, respectively, and with the final model, 41.8% and 33.5%, respectively. The final PK parameter estimates with uncertainty are reported in Table 2.

Evaluation

Internal model evaluation

Figure 1 shows plots assessing goodness-of-fit by comparing observations and predictions. A VPC of the final model is shown in Figure 2.

External model evaluation

The basic diagnostic plots are presented in Figure 1, and the VPC performed using the evaluation dataset and the final parameters from the PK model without additional fitting in Figure 2.

Table 3 shows the number of correct predictions (for five different datasets from the evaluation data and pooled results from the cross-validation) for gentamicin trough thresholds of 1 and 2 mg/L together with prediction errors. In the total dataset, containing both paired and unpaired samples, the median (95% CI) PE was 0.0004 (-1.1, 0.8) mg/L. The MPEs when predicting trough and peak concentrations (using cross-validations) were 0.03 and 0.19 mg/L; and the RMSE 1.28 and 2.55 mg/L, respectively (Table 3). When AUC(0-t) prediction (from one random sample) was evaluated, MPE was 14.5 mg h/L, and RMSE 30.2 mg h/L.

Figure 3 shows the median and the range of PE for this model and previously published gentamicin population PK models.

NeoGent

Figure S1 shows an example of output from neoGent.

Discussion

A PK model for gentamicin in neonates was developed and evaluated with prospectively collected data. Through its use of mechanistic covariates the model gave unbiased predictions of trough concentration from an opportunistic sample. Using this model, concentrations from samples taken at any time can be used to generate informative TDM, potentially eliminating the need for specifically timed trough gentamicin samples and the safety concerns and inconvenience associated with them. An exploratory analysis to evaluate whether such an approach could be used for predicting individual peak concentration and AUC(0-t) showed that while predictions were unbiased, they were relatively imprecise (Table 3).

The small median PE (0.0004 mg/L) for trough concentrations suggests that the model implemented in neoGent performs well, although some outliers were not captured (range: -2.4 – 1.6 mg/L). The median prediction errors were in most cases negative (Table 3), indicating that the model slightly over-predicts the trough concentrations (i.e. predicts them to be higher than they are), which might be (from a safety perspective) preferable to under-predicting. Cross-validations confirmed that samples do not need to be taken at a specific time when using this model for TDM, as predictions of trough concentrations (using an opportunistic sample) were unbiased, with median PE of -0.04 mg/L (Table 3). Although we did not test the effect of the sampling time on model predictions; the samples were collected from a wide range of times (0.1-53.3 h after the dose), as they would be in routine hospital tests.

Comparison of the developed model with the existing published models showed that the predicted trough concentrations were the least biased (i.e. the median prediction error was the smallest) when our model was used (Figure 3). However, due to unavailability of some covariates in our dataset, three models were used without all of the covariates (APGAR score(15, 19), sepsis(19), co-medication with dopamine(23)) included, which could explain their worse predictive performance.

318 The rich data in our model-building dataset (6.5 samples per patient) supported a 3-compartment
319 model, where the final estimates for the third compartment were: inter-compartmental clearance 0.3
320 L/h/70kg and peripheral volume of distribution of 148 L/70kg. Additionally, the terminal half-life for
321 a typical subject from the prospective evaluation dataset (weight 2.0 kg, PMA 34.9 weeks, PNA 6
322 days, MSCr 47.0 µmol/L, TSCr 66.4 µmol/L) was 189.7 hours. This could indicate uptake of
323 gentamicin into the renal cortex, and slow excretion from it (45); and is in agreement with previously
324 found evidence of deep tissue accumulation of gentamicin (26, 46).
325
326 Unfortunately many authors were unwilling or unable to share their data and we only managed to
327 obtain data from two (15, 21) out of eight identified studies for our model building dataset. We did
328 obtain one further subsequent dataset where assays from another pharmacokinetic study in neonates
329 also receiving gentamicin were used (32). Due to differences in model structure and parameterization,
330 it was not possible to extract relevant information for model building from the published reports.
331 However, in part because data from Nielsen *et al*(21) was of such high quality with multiple samples
332 per patient, our final model described both model building and the evaluation datasets well, as shown
333 in Figures 1 and 2. The histogram and the qq-plot of the conditional weighted residuals (data not
334 shown) confirmed that they follow a normal distribution. The final estimates for clearance (CL) and
335 volume of distribution (V) were (mean (standard error)) 6.21 (0.30) L/h/70 kg and 26.5 (1.11) L/70kg,
336 respectively (Table 2). The values of the PK parameters for a typical infant from the model-building
337 dataset (weight 2.12 kg, PMA 33.0 weeks, PNA 5.4 days, MSCr 78 µmol/L, TSCr 71.4 µmol/L) were
338 0.077 L/h and 0.80 L (and 0.10 L/h and 0.78 L for a neonate from the evaluation dataset) for CL and
339 V, respectively. These values are in agreement with estimates for clearance from previous neonatal
340 studies of gentamicin pharmacokinetics(13, 14, 18, 22-24). The reported value for CL from Nielsen *et*
341 *al*(21) may appear to be lower (0.026 L/h), but when our median demographic values were used in
342 their model, the CL became similar to our estimates (0.095 L/h). The final estimate for volume of
343 distribution is consistent with the estimate from Fuchs *et al*(23) and Botha *et al*(24), but it is not in
344 accordance with what was found by Garcia *et al* (20) (0.252 L). The probable reason for this is a

different studied population, as when the median weight from our dataset was used in their model, the resulting V was 0.968 L, in agreement with our estimate.

We did not attempt to estimate the allometric power exponents and constants of the maturation function as the PMA in the studied neonates (23.3-43.8 weeks) was insufficient to capture the age when maturation is complete (PMA_{50}=55.4 weeks(5)); instead, these constants were fixed to the values from another study in which the main focus was renal maturation(5). This type of scaling was used to improve the model usefulness by allowing it to be extrapolated to different subpopulations (for example, neonates with a different weight, or PMA). In addition to changes in clearance due to long-term maturation that extends throughout gestation and into the first two years of life, we attempted to capture the short-term changes in clearance that occur after birth regardless of gestational age. A benefit of fixing the long-term maturation based on known relationships between PMA and renal function was that this short-term maturation was apparent with our estimate of PNA_{50} of 40.8 hours, indicating that clearance rapidly increases over the first few days of life. In the first day of life the clearance was at 37% of the value for a typical adult, and it reached 95% by the end of the first month of age.

The typical serum creatinine (used in the model) was determined using SCr concentrations, determined by the Jaffe assay, because the same method was used to determine SCr in the model-building dataset. But to determine SCr in the evaluation dataset, assays, based on both the Jaffe and the enzymatic methods, were used. However, the goodness-of-fit to the evaluation dataset and the predictive performance of the model were good, therefore no correction factor was included. Also, the enzymatic assay was only used in 16% of patients. Due to the range of the data that was used to determine typical-for-PMA SCr the model can be used for a neonate with PMA <44 weeks or a term neonate of <4weeks of age. The power exponent on the creatinine function was estimated to be -0.13, meaning that if observed SCr and typical SCr were70 µmol/L and 60 µmol/L, respectively, clearance would be 2% lower.

Large η-shrinkage indicates that the data do not contain enough information to make a reliable individual estimation. And whilst the shrinkage was large on the peripheral volumes of distribution (V2 and V3), it was relatively small on clearance (6.9%) (Table 2), which is important for making predictions of trough gentamicin concentrations and AUC(0-t). The η-shrinkage was also relatively small (15%) on the central volume of distribution (Table 2).

Although the main aim was to evaluate whether the model can predict trough concentrations, the ability of the model to predict peak gentamicin concentration (from a randomly-selected non-peak sample) was also examined. Cross-validations showed that the median prediction error (95% CI) when predicting peaks was 0.16 (-4.76, 5.01) mg/L, indicating unbiased, but not very precise predictions. This is perhaps not surprising, given that concentrations collected at a median time after dose of 19.3 hours were used to predict concentrations at median 1h post dose. The prediction of AUC(0-t) (also from one sample) was similarly unbiased (median prediction error 10.8 mg h/L), but imprecise (95% CI: -24.9, 62.2 mg h/L) (Table 3). However, normalized RMSEs (by the range of observed data) for peak and AUC(0-t) prediction were 7.0% and 17.6%, respectively; indicating that considering the range of possible values, the precision is perhaps more acceptable. Target AUC(0-24) or peak values have not been defined in neonates, and slow clearance and a narrow therapeutic index mean that adjusting doses to target efficacy in this population may not be realistic. However, our model does now give unbiased predictions of both metrics from an opportunistically collected single sample, which should prove useful in future clinical research to define efficacy targets in this age group. At present, due to their imprecision, these predictions (for peak concentration and AUC(0-t)) should currently only be used for research purposes, and not for dose adjustment.

Conclusion

A new gentamicin model has been developed and evaluated with prospectively collected data. We used mechanistic covariate parameterization informed by principles of allometric size scaling, known scaling of glomerular filtration maturation, and standardization for age-expected creatinine. This "biological prior" information gave a model with better predictive performance on prospectively

collected external data than any previously published gentamicin model. Using this we developed a software tool neoGent (see Supplementary material for provisional stand-alone version, and implemented in the web TDM application TDMx (http://www.tdmx.eu/) (47)), which can be used to predict when the trough concentration will fall below 2 mg/L and so guide the dosing interval. Furthermore, peak concentration or AUC(0-24) from any post-dose sample can also be predicted with little bias.

Acknowledgements

The authors wish to thank Alison Thomson for making the data freely available online (http://www.rfpk.washington.edu, NIH/NIBIB grant P41-EB01975; when the website was last accessed (31/10/2014) it was no longer operational) and all patients, families and staff from the hospitals participating in the neoGent study. We would also like to thank the members of the neoGent collaboration: Dr Mark Anthony, Dr Tim Scorrer, Dr Prakash Satodia, Dr Nasreen Aziz and their research teams. We acknowledge the contribution of the UK National Institute for Health Research Clinical Research Network.

Funding

This project was funded by Action Medical Research (grant code SP4650, GN1834). EG received funding from the NeoMero study, part of the European Union Seventh Framework Programme for research, technological development and demonstration (Grant Agreement number 242146). JFS received funding from a United Kingdom Medical Research Council Fellowship (grant number G1002305).

Transparency declarations

None to declare.

427 **References**

428

429 1. **Turner MA, Lewis S, Hawcutt DB, Field D.** 2009. Prioritising neonatal medicines research:
430 UK Medicines for Children Research Network scoping survey. BMC Pediatr **9:**50.

431 2. **Cantey JB, Wozniak PS, Sanchez PJ.** 2014. Prospective Surveillance of Antibiotic Use in
432 the Neonatal Intensive Care Unit: Results from the SCOUT Study. Pediatr Infect Dis J
433 doi:10.1097/INF.0000000000000542.

434 3. **Pacifici GM.** 2009. Clinical pharmacokinetics of aminoglycosides in the neonate: a review.
435 Eur J Clin Pharmacol **65:**419-427.

436 4. **Ramirez MS, Tolmasky ME.** 2010. Aminoglycoside modifying enzymes. Drug Resist Updat
437 **13:**151-171.

438 5. **Rhodin MM, Anderson BJ, Peters AM, Coulthard MG, Wilkins B, Cole M, Chatelut E,**
439 **Grubb A, Veal GJ, Keir MJ, Holford NH.** 2009. Human renal function maturation: a
440 quantitative description using weight and postmenstrual age. Pediatr Nephrol **24:**67-76.

441 6. **Kearns GL, Abdel-Rahman SM, Alander SW, Blowey DL, Leeder JS, Kauffman RE.**
442 2003. Developmental pharmacology--drug disposition, action, and therapy in infants and
443 children. N Engl J Med **349:**1157-1167.

444 7. **Turnidge J.** 2003. Pharmacodynamics and dosing of aminoglycosides. Infect Dis Clin North
445 Am **17:**503-528.

446 8. **Begg EJ, Barclay ML, Kirkpatrick CM.** 2001. The therapeutic monitoring of antimicrobial
447 agents. Br J Clin Pharmacol **52 Suppl 1:**35S-43S.

448 9. **Dersch-Mills D, Akierman A, Alshaikh B, Yusuf K.** 2012. Validation of a dosage
449 individualization table for extended-interval gentamicin in neonates. Ann Pharmacother
450 **46:**935-942.

451 10. **Boyle EM, Brookes I, Nye K, Watkinson M, Riordan FA.** 2006. "Random" gentamicin
452 concentrations do not predict trough levels in neonates receiving once daily fixed dose
453 regimens. BMC Pediatr **6:**8.

11. **Stickland MD, Kirkpatrick CM, Begg EJ, Duffull SB, Oddie SJ, Darlow BA.** 2001. An extended interval dosing method for gentamicin in neonates. J Antimicrob Chemother **48:**887-893.

12. **Burton ME, Brater DC, Chen PS, Day RB, Huber PJ, Vasko MR.** 1985. A Bayesian feedback method of aminoglycoside dosing. Clin Pharmacol Ther **37:**349-357.

13. **Kelman AW, Thomson AH, Whiting B, Bryson SM, Steedman DA, Mawer GE, Samba-Donga LA.** 1984. Estimation of gentamicin clearance and volume of distribution in neonates and young children. Br J Clin Pharmacol **18:**685-692.

14. **Grasela TH, Ott R, Faix RG.** 1985. Population pharmacokinetics of gentamicin in neonates using routine clinical data, abstr American Society for Clinical Pharmacology and Therapeutics, 86th Annual Meeting, San Antonio, Texas, USA,

15. **Thomson AH, Way S, Bryson SM, McGovern EM, Kelman AW, Whiting B.** 1988. Population pharmacokinetics of gentamicin in neonates. Dev Pharmacol Ther **11:**173-179.

16. **Jensen PD, Edgren BE, Brundage RC.** 1992. Population pharmacokinetics of gentamicin in neonates using a nonlinear, mixed-effects model. Pharmacotherapy **12:**178-182.

17. **Weber W, Kewitz G, Rost KL, Looby M, Nitz M, Harnisch L.** 1993. Population kinetics of gentamicin in neonates. Eur J Clin Pharmacol **44 Suppl 1:**S23-25.

18. **Lanao JM, Calvo MV, Mesa JA, Martin-Suarez A, Carbajosa MT, Miguelez F, Dominguez-Gil A.** 2004. Pharmacokinetic basis for the use of extended interval dosage regimens of gentamicin in neonates. J Antimicrob Chemother **54:**193-198.

19. **Lingvall M, Reith D, Broadbent R.** 2005. The effect of sepsis upon gentamicin pharmacokinetics in neonates. Br J Clin Pharmacol **59:**54-61.

20. **García B, Barcia E, Pérez F, Molina IT.** 2006. Population pharmacokinetics of gentamicin in premature newborns. Journal of Antimicrobial Chemotherapy **58:**372-379.

21. **Nielsen EI, Sandstrom M, Honore PH, Ewald U, Friberg LE.** 2009. Developmental pharmacokinetics of gentamicin in preterm and term neonates: population modelling of a prospective study. Clin Pharmacokinet **48:**253-263.

22. **Arenas-Lopez S, Mulla H, Durward A, Tibby SM.** 2010. Extended-interval gentamicin: population pharmacokinetics in pediatric critical illness. Pediatr Crit Care Med **11:**267-274.

23. **Fuchs A, Guidi M, Giannoni E, Werner D, Buclin T, Widmer N, Csajka C.** 2014. Population pharmacokinetic study of gentamicin in a large cohort of premature and term neonates. Br J Clin Pharmacol doi:10.1111/bcp.12444.

24. **Botha JH, du Preez MJ, Adhikari M.** 2003. Population pharmacokinetics of gentamicin in South African newborns. Eur J Clin Pharmacol **59:**755-759.

25. **Heimann G.** 1983. Renal toxicity of aminoglycosides in the neonatal period. Pediatr Pharmacol (New York) **3:**251-257.

26. **Laskin OL, Longstreth JA, Smith CR, Lietman PS.** 1983. Netilmicin and gentamicin multidose kinetics in normal subjects. Clin Pharmacol Ther **34:**644-650.

27. **Valitalo PA, van den Anker JN, Allegaert K, de Cock RF, de Hoog M, Simons SH, Mouton JW, Knibbe CA.** 2015. Novel model-based dosing guidelines for gentamicin and tobramycin in preterm and term neonates. J Antimicrob Chemother doi:10.1093/jac/dkv052.

28. **Nakae S, Yamada M, Ito T, Chiba Y, Sasaki E, Sakamoto M, Tada K, Yamada T, Mori S.** 1988. Gentamicin dosing and pharmacokinetics in low birth weight infants. Tohoku J Exp Med **155:**213-223.

29. **Ali AS, Farouq MF, Al-Faify KA.** 2012. Pharmacokinetic approach for optimizing gentamicin use in neonates during the first week of life. Indian J Pharmacol **44:**36-40.

30. **Lannigan R, Thomson AH.** 2001. Evaluation of 22 Neonatal Gentamicin Dosage Protocols Using a Bayesian Approach. Paediatric and Perinatal Drug Therapy **4:**92-100.

31. **Rastogi A, Agarwal G, Pyati S, Pildes RS.** 2002. Comparison of two gentamicin dosing schedules in very low birth weight infants. Pediatr Infect Dis J **21:**234-240.

32. **Metsvaht T, Tasa T, Kipper K, Padari H, Ilmoja M, Lutsar I.** 2015. Population pharmacokinetics (PK) of gentamicin in term and near-term neonates, abstr The world society for pediatric infectious diseases (WSPID), 9th world congress, Rio de Janeiro, Brazil,

33. **Jaffe M.** 1886. Ueber den Niederschlag, welchen Pikrinsäure in normalem Harn erzeugt und über eine neue Reaction des Kreatinins. Zeitschrift für physiologische Chemie **10:**391-400.

509 34. **Boeckmann AJ, Beal SL, Sheiner LB.** 1999. NONMEM users guide, University of
510 California at San Francisco, San Francisco.

511 35. **Anderson BJ, Holford NH.** 2009. Mechanistic basis of using body size and maturation to
512 predict clearance in humans. Drug Metab Pharmacokinet **24:**25-36.

513 36. **West GB, Brown JH, Enquist BJ.** 1997. A general model for the origin of allometric scaling
514 laws in biology. Science **276:**122-126.

515 37. **Germovsek E, Barker CI, Standing JF.** 2015. An Argument for Standardised Scaling:
516 Comparison of Methods for Scaling Clearance in Children. PAGE 24 (2015) Abstr 3635
517 [www.page-meeting.org/?abstract=3635], Crete, Greece.

518 38. **Cuzzolin L, Fanos V, Pinna B, di Marzio M, Perin M, Tramontozzi P, Tonetto P,**
519 **Cataldi L.** 2006. Postnatal renal function in preterm newborns: a role of diseases, drugs and
520 therapeutic interventions. Pediatr Nephrol **21:**931-938.

521 39. **Rudd PT, Hughes EA, Placzek MM, Hodes DT.** 1983. Reference ranges for plasma
522 creatinine during the first month of life. Arch Dis Child **58:**212-215.

523 40. **Lindbom L, Pihlgren P, Jonsson EN.** 2005. PsN-Toolkit--a collection of computer intensive
524 statistical methods for non-linear mixed effect modeling using NONMEM. Comput Methods
525 Programs Biomed **79:**241-257.

526 41. **Jonsson EN, Karlsson MO.** 1999. Xpose--an S-PLUS based population
527 pharmacokinetic/pharmacodynamic model building aid for NONMEM. Comput Methods
528 Programs Biomed **58:**51-64.

529 42. **Brendel K, Dartois C, Comets E, Lemenuel-Diot A, Laveille C, Tranchand B, Girard P,**
530 **Laffont CM, Mentre F.** 2007. Are population pharmacokinetic and/or pharmacodynamic
531 models adequately evaluated? A survey of the literature from 2002 to 2004. Clin
532 Pharmacokinet **46:**221-234.

533 43. **Sheiner LB, Beal SL.** 1981. Some suggestions for measuring predictive performance. J
534 Pharmacokinet Biopharm **9:**503-512.

44. **Ambrose PG, Bhavnani SM, Rubino CM, Louie A, Gumbo T, Forrest A, Drusano GL.** 2007. Pharmacokinetics-pharmacodynamics of antimicrobial therapy: it's not just for mice anymore. Clin Infect Dis **44:**79-86.

45. **Prayle A, Watson A, Fortnum H, Smyth A.** 2010. Side effects of aminoglycosides on the kidney, ear and balance in cystic fibrosis. Thorax **65:**654-658.

46. **Schentag JJ, Jusko WJ.** 1977. Renal clearance and tissue accumulation of gentamicin. Clin Pharmacol Ther **22:**364-370.

47. **Wicha SG, Kees MG, Solms A, Minichmayr IK, Kratzer A, Kloft C.** 2015. TDMx: a novel web-based open-access support tool for optimising antimicrobial dosing regimens in clinical routine. Int J Antimicrob Agents **45:**442-444.

547 **Tables and figures**

548

549 Table 1: A summary of demographics and dosing

	Model-building dataset	Evaluation dataset
n	205	163
weight (g) [a]	2.12 (0.53-5.05)	2.03 (0.48-5.05)
gestational age (weeks) [a]	34.0 (23.3-42.1)	34.3 (23.9-42.3)
postnatal age (days) [a]	5.4 (1-66)	6 (1-78)
postmenstrual age (weeks) [a]	33.0 (23.3-43.8)	34.9 (24-43.3)
females (%)	89 (43%)	68 (41.7%)
gentamicin samples per patient [b]	6.5	3.0
gentamicin concentration (mg/L) [a]	3.4 (0.3-37.6)	1.0 (0.1-13.2)
time after the dose (h) [a]	8.0 (0.02-54.1)	23.5 (0.08-79.7)
occasion [a]	2 (1-22)	2 (1-7)

550 Weight and gestational age are values at treatment initiation, the rest are values at time of gentamicin

551 sampling/dosing; an occasion was defined as a dose with subsequent gentamicin samples taken; day

552 of birth was defined as day 1; [a]median (range); [b]mean

553

554 Table 2: Final parameter estimates from NONMEM output file and from the bootstrap analysis

	Parameters from the final model				Bootstrap analysis		
	mean	SE	%CV	η-shrinkage	median	2.5%ile	97.5%ile
CL (L/h/70kg)	6.21	0.30	-	-	6.14	5.47	6.75
θ_SCr	-0.13	0.055	-	-	-0.13	-0.25	-0.03
PNA_{50} (days)	1.70	0.30	-	-	1.68	1.15	2.30
V (L/70kg)	26.5	1.11	-	-	26.3	23.6	28.4
Q (L/h/70kg)	2.15	0.32	-	-	2.19	1.68	3.25
V2 (L/70kg)	21.2	1.50	-	-	20.9	17.9	24.2
Q2 (L/h/70kg)	0.27	0.047	-	-	0.28	0.19	0.38
V3 (L/70kg)	148	52.0	-	-	152	65.2	534
IIV on CL	0.175	0.038	41.8	6.9	0.170	0.104	0.254
IIV on V	0.112	0.032	33.5	15.2	0.113	0.057	0.190
covariance CL-V	0.116	0.030	-	-	0.115	0.060	0.184
IIV on V2	0.132	0.060	36.3	57.8	0.117	0.023	0.281
IIV on V3	0.177	0.216	42.1	85.0	0.114	0.00002	4.18
inter-occasion variability	0.014	0.007	11.8	-	0.013	0.001	0.029
residual error (proportional)	0.036	0.006	19.0	-	0.036	0.025	0.049
residual error (additive)	0.016	0.007	-	-	0.015	0.000002	0.032

555 CL is clearance, V is volume of distribution, Q is inter-compartmental CL, IIV is inter-individual
556 variability, SE is standard error obtained with NONMEM 7.3 covariance step, CV is coefficient of
557 variation.

558
559

Table 3: Summary of external evaluation with the evaluation dataset

dataset	Limit = 1 mg/L			Limit = 2 mg/L			PE (mg/L)	MPE (mg/L)	RMSE (mg/L)
	n correct (%)	OP	UP	n correct (%)	OP	UP			
paired + unpaired	214/254 (84.3)	20	20	242/254 (95.3)	10	2	0.0004 (-1.07, 0.84)	0.007	0.45
paired: study≥1mg/L	53/57 (93.0)	3	1	56/57 (98.2)	1	0	-0.04 (-0.57, 0.70)	-0.03	0.32
paired: study≥2mg/L	31/33 (93.9)	2	0	33/33 (100)	0	0	-0.08 (-0.50, 0.74)	-0.05	0.35
paired: study≥3mg/L	19/20 (95.0)	0	1	20/20 (100)	0	0	-0.06 (-0.56, 0.82)	-0.02	0.42
unpaired	136/161 (84.5)	14	11	155/161 (96.3)	5	1	0.02 (-1.11, 0.70)	-0.001	0.43
XV: paired: study≥3mg/L	478/502 (95.2)	12	12	460/502 (91.6)	21	21	-0.04 (-1.77, 3.03)	0.03	1.28
XV: peaks [a]	-	-	-	-	-	-	0.16 (-4.76, 5.01)	0.19	2.55
AUC(0-t) [a]	-	-	-	-	-	-	10.8 (-24.9, 62.2) [b]	14.5 [b]	30.2 [b]

Correct indicates that the predicted trough concentration agrees with the measured concentration (is above/below the limit); OP is overprediction, UP is underprediction; PE is prediction error (median (95% confidence interval)), MPE is mean prediction error, RMSE is root mean square error, XV is cross-validation. Except [a] all results refer to trough prediction evaluation. [b] in mg h/L.

Figure legends

Figure 1: Observed versus population predicted gentamicin serum concentrations (top left for the model-building dataset and bottom left for the evaluation dataset) and conditional weighted residuals versus time after dose (top right for the model-building dataset and bottom right for the evaluation dataset).

Figure 2: Visual predictive check of 1000 simulated concentration-time datasets from the final model, using the model-building dataset (left) and the evaluation dataset (right). Points are the observations, black lines are the 2.5th, 50th, and 97.5th percentiles, and the shaded areas are the 95% confidence intervals of the corresponding predicted gentamicin concentrations.

Figure 3: Comparison of predictive performance of the developed model (shaded box plot) and previously published neonatal gentamicin PK models.