Comparisons Between Hamiltonian Monte Carlo and Maximum A Posteriori For A Bayesian Model For Apixaban Induction Dose & Dose Personalization

A. Demetri Pananos APANANOS@UWO.CA
Department of Epidemiology and Biostatistics
Western University
London, Ontario, Canada

Daniel J. Lizotte DLIZOTTE@UWO.CA
Department of Epidemiology and Biostatistics
Department of Computer Science
Western University
London, Ontario, Canada

Editor: Editor's name

Abstract

Precision medicine's slogan is "right drug – right patient – right time", which is implicitly preceded by "right dose". However, determining the right dose for any one patient can be challenging because existing models for dose personalization require more data than can reasonably be obtained in order to truly personalize doses. Bayesian methods, with their ability to explicitly pass prior information to the model, have been proposed as a solution to this problem. Hamiltonian Monte Carlo (HMC) is largely considered the gold standard for fitting Bayesian models, and yet several dose personalization studies use Maximum A Posteriori (MAP) methods to fit their models. In this study, we put forth a new Bayesian pharmacokinetic model for apixaban pharmacokinetics written in an open source Bayesian language. We make the model code and posterior summaries of all parameters publicly available. We also perform a simulation study characterizing the differences between inferences made via MAP and HMC for personalized dosing strategies. We find that the differences between HMC and MAP are non-trivial and can greatly affect the choices surrounding dose selection for personalized medicine.

1. Introduction

Precision medicine's slogan is "right drug – right patient – right time", which is implicitly preceded by "right dose". However, determining the right dose for any one patient can be challenging. The anticoagulant Warfarin offers a good example of these challenges; physicians choose an initial dose based on guidelines and their own experience. They then closely monitor the patient's International Normalised Ratio (INR), which measures how long it takes blood to clot, and in response they adjust the dose over time.

Pharmacokinetic and statistical models of how drugs behave within an individual can alleviate some of these challenges. In some studies Sohrabi and Tajik (2017); Caldwell et al. (2007); Consortium (2009) a cohort of patients will have appropriate maintenance doses

determined empirically and these are then regressed onto patient covariates. In others Zhu et al. (2017); Xue et al. (2017) patient pharmacokinetics are directly modeled and can be simulated under different dosing regimens to find an appropriate dose. In both cases, uncertainty in the models can be assessed and can help guide clinical decisions as to what dose is best or what dose to try next.

Both types of models can provide guidance for individual patients, but only when there is enough data so that the models are accurate and reliable. Rarely is this amount of data available in practice. Obtaining sufficient data to learn a patient's pharmacokinetic parameters would require a lengthy observation period which few patients are willing or capable of committing to. Population pharmacokinetic models could be used in place of a patient's pharmacokinetics, but treating the patient as "average" is precisely what precision medicine seeks to improve upon.

When there is a paucity of data, Bayesian methods with strong priors have been proposed. Model priors allow analysts to specify their beliefs about model parameters prior to seeing data. This allows models to "hit the ground running" so to speak, and makes use of all available data to make inferences. All but the simplest of Bayesian models require computational techniques to obtain model estimates and predictions due to the presence of intractable integrals. Several approaches exist for generating approximate samples from the posterior distribution, with Hamiltonian Monte Carlo (HMC) being the gold standard Neal (1996); Matthew D. Hoffman (2014); Carpenter et al. (2017); Tripuraneni et al. (2017). Despite HMC being the preferred method by theorists and applied Bayesians alike, methods like Maximum A Posteriori (MAP), in which the posterior mode is computed via optimization and then a Laplace approximation is performed, continue to be used in population Bayesian pharmacokinetic studies as late as 2020 Brooks et al. (2016); Nguyen et al. (2016); Preijers et al. (2019); Stifft et al. (2020). HMC and MAP are two different approaches with different strengths and different theoretical motivations. Naturally, this raises questions regarding how decisions in personalized medicine may be affected by the use of different methods for performing inference. We seek to answer these questions by developing a new, high-fidelity Bayesian pharmacokinetic model and then investigating the impact of the choice of inference method on precision medicine decisions.

The main contributions of this paper are as follows:

1. A new Bayesian pharmacokinetic model for apixaban pharmacokinetics written in an open source Bayesian language. We make the model code and posterior summaries of all parameters publicly available.

2. A simulation study characterizing the differences between inferences made via MAP and HMC for personalized dosing strategies.

3. An induction dosing model for apixaban based on desired trough concentration level after a first dose.

2. Background

Pharmacokinetic Modelling

Broadly, pharmacokinetics is the study of the dynamics of a mass of drug in the body and is concerned with the absorption, distribution, metabolism, and excretion of that drug. Differential equations (equations which relate the derivative of an unknown function to itself) are often used to describe how these dynamics evolve over time. The differential equation models in pharmacokinetics are called "compartmental models" as they idealize different parts of the body as compartments from which drug can flow in and out at rates proportional to how much drug is presently in that compartment. If the differential equation is not too complex, the solution can be written in terms of analytic functions. In the case where the differential equation can not be solved in terms of analytic functions, a rich literature of numerical techniques exist to approximate the solution to within quantifiable precision. In either case, estimation of model parameters is of interest as they represent pharmacokinetic measures, such as the volume of distribution or rate constants for which the drug is absorbed into/excreted out of a compartment. If the parameters for such a model are known, we can use the models to make predictions about drug concentration as a function of time and dose. This in turn can be used to select a dose that meets given criteria about what the concentration function should look like.

Parameter estimation for these models can be done in both frequentist and Bayesian frameworks. In a Bayesian framework, parameter estimation begins by specifying a prior distribution which reflects the knowledge of parameters before seeing data. Once data is observed, Bayes' rule can be used to get the posterior distribution. This distribution provides information about what parameter values have most plausibly generated the observed data. By virtue of being a probability distribution, the posterior can be summarized by expectations to get point estimates of model parameters. Shown in fig. 1 is a visual summary of how Bayes' rule and Bayesian modelling of pharmacokinetics works using pseudodata. The leftmost panel is our prior distribution. Each concentration curve results from specific combinations of parameters for the model which are believed to be plausible before seeing data. Once data is observed (the middle panel), application of Bayes' rule yields the rightmost panel. Concentration curves in this panel correspond to combinations of parameters which have most plausibly generated the data, resulting in concentration curves which have most plausibly generated the data. Note that in this setting, because we have many measurements, the pharmacokinetic model is well-determined and the posterior uncertainty is small. Except in very simple cases, the integrals required to evaluate the posterior quickly become intractable, thus computational approximations are required to fit Bayesian models.

Dosing Decisions

Vitamin K antagonists, such as the popular oral anticoagulant Warfarin, are known to have narrow therapeutic windows as well as drug and food interaction. Determination of a maintenance dose is consequently a procedure with frequent monitoring and followup, with some sources recommending monitoring daily or every other day until the INR stabilizes for two days. The narrow therapeutic window forces investigators to also consider the pharmacody-

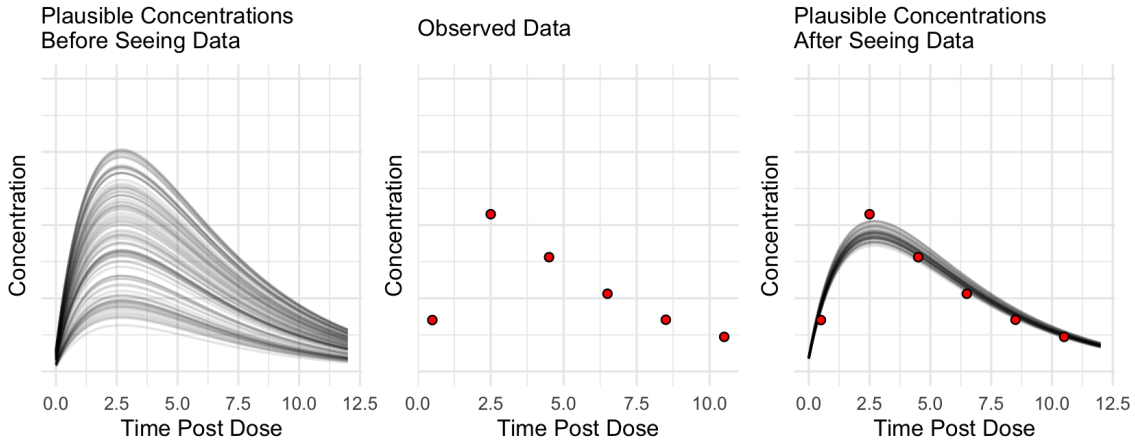

Figure 1: A demonstration of a Bayesian workflow for pharmacokientic models. The left-most panel represents the prior. Each curve corresponds to a unique set of model parameters which induce each concentration function. In the center panel is the data observed from a single patient. Conditioning on this data yields the right-most panel. Each curve corresponds to a unique set of model parameters drawn from the posterior distribution.

namics (the study of the onset, intensity, and duration of the drug response and how these are related to the concentration of the drug at its site of action) of Warfarin in addition with the pharmacokinetics when determining dose size as concentration of the drug alone is not sufficient to infer the antithrombotic effect in patients. The introduction of factor Xa inhibitors like apixaban has alleviated some of the difficulties in prescribing anticoagulants. Factor Xa inhibitors have been shown to have lower risk for bleeding than Warfarin in patients with atrial fibrillation Vinogradova et al. (2018) and also allow for fixed dosing without frequent monitoring INR. Furthermore, unlike Warfarin, the pharmacodynamic effect of apixaban on clotting is closely correlated with the concentration in the plasma Byon et al. (2019), making pharmacokinetic modelling more informative on antithrombotic effect as compared to Warfarin. However, as of writing this paper there is little information on the therapeutic window, making selecting dose sizes large enough to avoid thromboembolism difficult. In this work, we develop personalized dosing whose goal is to find the minimum dose that avoids plasma concentrations that are too low.

3. Methods

Bayesian Model

We fit a hierarchical mixed effects model of apixaban pharmacokinetics using data from Beaton et al. (2018). Thirty-six participants were given 5 mg of apixaban and 100 ml of water in a fasted state. Blood plasma concentrations of apixaban were recorded over the course of 12 hours.

Table 1: Need a caption

	Female (N=184)	Male (N=104)	Overall (N=288)
Age			
Mean (SD)	48.8 (11.4)	51.7 (11.2)	49.8 (11.4)
Median [Min, Max]	49.0 [26.0, 67.0]	51.0 [31.0, 70.0]	50.0 [26.0, 70.0]
Weight			
Mean (SD)	84.6 (23.1)	94.0 (24.7)	88.0 (24.1)
Median [Min, Max]	82.8 [54.7, 136]	87.2 [62.0, 137]	83.5 [54.7, 137]
Creatinine			
Mean (SD)	68.5 (11.9)	67.2 (13.0)	68.0 (12.3)
Median [Min, Max]	66.0 [50.0, 95.0]	65.0 [50.0, 95.0]	65.0 [50.0, 95.0]
BMI			
Mean (SD)	29.9 (6.61)	31.5 (5.42)	30.5 (6.25)
Median [Min, Max]	29.5 [18.3, 42.3]	31.8 [23.3, 40.7]	31.3 [18.3, 42.3]

Since participants were given a single dose of apixaban in a fasted state, we use a single-compartment pharmacokinetic model with first order absorption and elimination. The profile of plasma concentration as a function of time is given as

$$y(t) = \frac{F \cdot D}{Cl} \frac{k_e \cdot k_a}{k_e - k_a} \left(e^{-k_a(t-\delta)} - e^{-k_e(t-\delta)} \right). \tag{1}$$

Here, D$D$is the size of the dose in mg, F s the bioavailability (fixed to 0.5 for apixaban Byon et al. (2019)), Cl is the clearance rate in units litres per hour, k_a is the rate constant of absorption into the volume of distribution in units 1/hours, and k_e is the elimination rate constant in units 1/hours. We include a time delay, δ, to relax the assumption that absorption begins immediately after ingestion. Parameters are considered as random effects, with some population mean and variance (which is estimated from the data).

Priors for k_e and k_a are not defined explicitly. Rather our model puts priors on the time to max concentration, which can be expressed as a function of the parameters in eq. (1)

$$t_{max} = \frac{\ln(k_a) - \ln(k_e)}{k_a - k_e} \tag{2}$$

and on the ratio between k_e and ka, which we call α

$$\alpha = \frac{k_e}{k_a}. \tag{3}$$

We choose to place a prior on the quantity alpha because it arises when non-dimensionalizing Lin and Segel (1988) the differential equation governing mass transit of the drug in and out of the volume of distribution. The plasma concentration function is a version of the "flip-flop" model Wakefield (1996); Salway and Wakefield (2008), since different parameterizations of this model can yield the same curve. This leads to model un-identifiability. To ensure the model is identifiable, we require $k_e < k_a$ as has been done in previous Bayesian analyses

of this model Wakefield (1996); Salway and Wakefield (2008). This requirement bounds alpha to the unit interval. In principle, information on the elimination rate constant could be obtained by examining the latter half of the concentration profiles where the drug is being eliminated from the body. To preserve as much data for model fitting, we forgo this approach. These two sets of equations are used to parameterize the absorption and elimination rate constants as follows

$$k_a = \frac{1}{t_{max}} \frac{\ln(\alpha)}{\alpha - 1} \tag{4}$$

$$k_e = k_a \alpha \tag{5}$$

Time to max concentration values for patient j are drawn from a log normal distribution

$$t_{max,j} | \mu_t, \sigma_t \sim \text{LogNormal}(\mu_t, \sigma_t) \tag{6}$$

and α is drawn from a weakly informative beta prior to prevent degenerate cases when α is 0 or 1

$$\alpha_j \sim \text{Beta}(2, 2) \tag{7}$$

Parameters $k_{e,j}$ and $k_{a,j}$ are determined from eqs. (4) and (5). The clearance rate is modelled hierarchically

$$Cl_j | \mu_{Cl}, \sigma_{Cl} \sim \text{LogNormal}(\mu_{Cl}, \sigma_{Cl}) \tag{8}$$

Each patient is observed to have a non-zero concentration at time 0.5, so the time delay for each patient is no larger than 0.5 hours. We place a beta prior on the delay

$$\delta_j | \phi, \kappa \sim \text{Beta}(\phi/\kappa, (1-\phi)/\kappa) \tag{9}$$

and multiply delta by 0.5 in our model to ensure the maximum delay is 0.5 hours. Here, ϕ is the mean of this beta distribution and κ determines the precision of the distribution. Shown below is a Bayes net to exposit model structure at a high level.

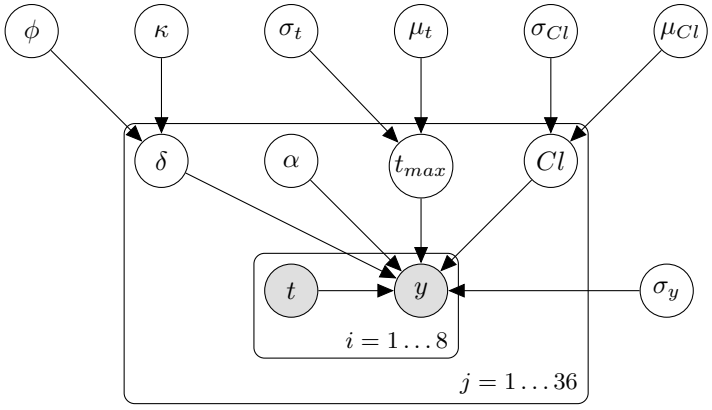

Figure 2: MODEL NEEDS CAPTION

Priors for Model Hyperparameters

Estimates of the time to max concentration for apixaban place the population median tmax near 3.3 hours after ingestion Byon et al. (2019) Assuming the median and the mean are similar, this provides information for μ_t and so we use specify

$$p(\mu_t) = \text{Normal}(\log(3.3), 0.25) \tag{10}$$

The standard deviation of the prior for $mu_t$was selected via prior predictive checks in which profiles are drawn and priors are assessed as realistic or not. We choose to err on the side of caution and inflate the uncertainty in this estimate to account for population differences between the measured patients in the data and the patients used in studies to determine the estimates of t_max. The population variability of t_max was modeled as

$$p(\sigma_t) = \text{Gamma}(10, 100) \tag{11}$$

Using these priors, we recover similar median, min, and max t_max values as reported in Byon et al. (2019). Similarly, we model the population mean and variability for the clearance rate as

$$p(\mu_{Cl}) = \text{Normal}(\log(3.3), 0.15) \tag{12}$$
$$p(\sigma_{Cl}) = \text{Gamma}(15, 100) \tag{13}$$

so that population estimates of the mean clearance rate are near 3.3 litres per hour with inflated uncertainty to account for possible population differences. We use weakly informative priors for ϕ and κ which have the effect of inducing an approximately uniform prior for δ.

$$p(\phi) = \text{Beta}(20, 20) \tag{14}$$
$$p(\kappa) = \text{Beta}(20, 20) \tag{15}$$

The tools used to measure the concentration of apixaban are believed to be within 10% of the real concentration. This implies that the observational model is heteroskedastic. We use a log-normal likelihood so that positivity of observed concentrations and heteroskedasticity are respected. We place a lognormal prior on the likelihood's variability with

$$p(\sigma_y) = \text{LogNormal}(\ln(0.1), 0.2) \tag{16}$$
$$C_j(t)|Cl_j, k_{a,j}, k_{a,j}, \delta_j \sim \text{LogNormal}(\ln(y(t)), \sigma_y) \tag{17}$$

Model Fitting and Diagnostics

For HMC, prior predictive checks and model fitting was performed in Stan Carpenter et al. (2017). Twelve chains were initialized and ran for 4000 iterations each (1000 for warmup allowing the Markov chain the opportunity to find the correct target distribution and 3000

to use as samples from the posterior). Stan monitors several diagnostics, which did not detect problematic HMC behavior [1].

We use Stan's optimization capabilities to compute the MAP estimate of the posterior distribution for our simulated patients. The L-BFGS optimizer was used to find the posterior mode. The optimizer terminated when either 10,000 iterations had been performed or when the value of the objective function stopped changing within a tolerance of 1e-10. Once the mode was located, 10,000 samples from the Laplace approximation to the posterior were obtained. Constrained parameters were transformed to the appropriate space before sampling.

Posterior Summarization and Generating New Data

Once our model was fit on the pharmacokinetic data, the marginal posteriors were summarized to create priors for the new model. Parameters for these priors were determined by using maximum likelihood on the posterior samples. The priors for the new model are as follows:

$$\mu_{Cl} \sim \text{Normal}(1.64, 0.09) \tag{18}$$
$$\sigma_{Cl} \sim \text{LogNormal}(-0.94, 0.11) \tag{19}$$
$$\mu_t \sim \text{Normal}(0.97, 0.05) \tag{20}$$
$$\sigma_t \sim \text{LogNormal}(-1.40, 0.12) \tag{21}$$
$$\alpha_j \sim \text{Beta}(2, 2) \tag{22}$$
$$\sigma_y \sim \text{LogNormal}(-1.76, 0.12) \tag{23}$$

Lognormal distributions were used to respect positivity of some parameters. The posterior predictive distribution was then used to simulate 100 new patients. The model with the summarized priors was then refit on the 100 simulated patients in order to examine differences between HMC and MAP in a "best case" scenario. Simulated patients were sampled between 0.5 and 12.0 hours after ingestion in increments of 0.5. Draws from the posterior were used to predict latent concentration for each patient at times 0.75 to 11.75 in increments of 0.5.

Determination of a Personalized Dose

We determine a personalized dose size by evaluating the pseudo patients' pharmacokinetics under different dose sizes and then computing posterior probabilities of failing to surpass concentration thresholds. In our first experiment, we determine the posterior probability of failing to exceed a concentration of 20 ng/ml 12 hours post dose for each pseudopatient. In our second experiment, we compute the expected amount of time spent below 20 ng/ml for each pseudopatient. We interpret these probabilities as the risk of being below the 20 ng/ml threshold. The chosen threshold is arbitrary, but our method generalizes to any threshold. These risks are computed across dose sizes of 0 mg to 60 mg, yielding risk as a function of

1. 0 divergences, all Gelman-Rubin diagnostics < 1.01, smallest effective sample size ratio 16%. See the supplement for a detailed diagnostic report for this model.

dose size. We interpolate these estimates using a monotone Hermite spline and then invert the risk curve for each pseudopatient. This allows us to determine a dose size which elicits a pre-specified risk level.

References

Melanie D Beaton, Rommel G Tirona, Ruth Strapp, Mala Ramu, Ute Schwarz, Richard Kim, Bandar Aljudaibi, and Zahra Kassam. Su1495 - apixaban and rosuvastatin pharmacokinetics in nonalcoholic fatty liver disease, 2018.

Emily Brooks, Susan E Tett, Nicole M Isbel, and Christine E Staatz. Population pharmacokinetic modelling and bayesian estimation of tacrolimus exposure: Is this clinically useful for dosage prediction yet? *Clin. Pharmacokinet.*, 55(11):1295–1335, November 2016.

Wonkyung Byon, Samira Garonzik, Rebecca A Boyd, and Charles E Frost. Apixaban: A clinical pharmacokinetic and pharmacodynamic review. *Clin. Pharmacokinet.*, 58(10): 1265–1279, October 2019.

Michael D Caldwell, Richard L Berg, Kai Qi Zhang, Ingrid Glurich, John R Schmelzer, Steven H Yale, Humberto J Vidaillet, and James K Burmester. Evaluation of genetic factors for warfarin dose prediction. *Clin. Med. Res.*, 5(1):8–16, March 2007.

Bob Carpenter, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A probabilistic programming language, 2017.

International Warfarin Pharmacogenetics Consortium. Estimation of the warfarin dose with clinical and pharmacogenetic data. *New England Journal of Medicine*, 360(8):753–764, 2009.

C C Lin and L A Segel. Mathematics applied to deterministic problems in the natural sciences, 1988.

Andrew Gelman Matthew D. Hoffman. The No-U-Turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, 2014.

Radford M Neal. Bayesian learning for neural networks, 1996.

Thu Thuy Nguyen, Henri Bénech, Alain Pruvost, and Natacha Lenuzza. A limited sampling strategy based on maximum a posteriori bayesian estimation for a five-probe phenotyping cocktail. *Eur. J. Clin. Pharmacol.*, 72(1):39–51, January 2016.

Tim Preijers, Britta A P Laros-vanGorkom, Ron A A Mathôt, and Marjon H Cnossen. Pharmacokinetic-guided dosing of factor VIII concentrate in a morbidly obese severe haemophilia a patient undergoing orthopaedic surgery, 2019.

Ruth Salway and Jon Wakefield. Gamma generalized linear models for pharmacokinetic data. *Biometrics*, 64(2):620–626, June 2008.

Mohammad Karim Sohrabi and Alireza Tajik. Multi-objective feature selection for warfarin dose prediction. *Comput. Biol. Chem.*, 69:126–133, August 2017.

Frank Stifft, Franciscus Vandermeer, Cees Neef, Sander van Kuijk, and Maarten H L Christiaans. A limited sampling strategy to estimate exposure of once-daily modified release tacrolimus in renal transplant recipients using linear regression analysis and comparison with bayesian population pharmacokinetics in different cohorts. *Eur. J. Clin. Pharmacol.*, February 2020.

Nilesh Tripuraneni, Mark Rowland, Zoubin Ghahramani, and Richard Turner. Magnetic hamiltonian monte carlo. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML'17, pages 3453–3461. JMLR.org, August 2017.

Yana Vinogradova, Carol Coupland, Trevor Hill, and Julia Hippisley-Cox. Risks and benefits of direct oral anticoagulants versus warfarin in a real world setting: cohort study in primary care. *bmj*, 362:k2505, 2018.

Jon Wakefield. The bayesian analysis of population pharmacokinetic models, 1996.

Ling Xue, Nick Holford, Xiao-liang Ding, Zhen-ya Shen, Chen-rong Huang, Hua Zhang, Jing-jing Zhang, Zhe-ning Guo, Cheng Xie, Ling Zhou, Zhi-yao Chen, Lin-sheng Liu, and Li-yan Miao. Theory-based pharmacokinetics and pharmacodynamics of s- and r-warfarin and effects on international normalized ratio: influence of body size, composition and genotype in cardiac surgery patients, 2017.

Yu-Bin Zhu, Xian-Hua Hong, Meng Wei, Jing Hu, Xin Chen, Shu-Kui Wang, Jun-Rong Zhu, Feng Yu, and Jian-Guo Sun. Development of a novel individualized warfarin dose algorithm based on a population pharmacokinetic model with improved prediction accuracy for chinese patients after heart valve replacement. *Acta Pharmacol. Sin.*, 38(3): 434–442, March 2017.