Multistate models of developmental toxicity: Application to valproic acid-induced malformations in the zebrafish embryo

Ségolène Siméon, Rémy Beaudouin, Katharina Brotzmann, Thomas Braunbeck, Frédéric Bois

Multistate Models of Developmental Toxicity: Application to Valproic Acid-Induced Malformations in the Zebrafish Embryo

Ségolène Siméon[†], Rémy Beaudouin[˜], Katharina Brotzmann[‡], Thomas Braunbeck[‡], Frédéric Y. Bois[*,§].

[†] INERIS, METO unit, Parc ALATA BP2, Verneuil en Halatte, France.

[˜] UMR-I 02 SEBIO, INERIS, Parc ALATA BP2, Verneuil en Halatte, France.

[‡] University of Heidelberg, Aquatic Ecology and Toxicology, Centre for Organismal Studies (COS), Im Neuenheimer Feld 504, D-69120 Heidelberg, Germany.

[§] CERTARA UK Limited, Simcyp Division, Level 2-Acero, 1 Concourse Way, Sheffield, S1 2BJ, United Kingdom.

* Corresponding author: frederic.bois@certara.com

Abstract

For the determination of acute toxicity of chemicals in zebrafish (*Danio rerio*) embryos, the OECD test guideline 236, relative to the Fish Embryo Toxicity Test (FET), stipulates a dose-response analysis of four lethal core endpoints and a quantitative characterization of abnormalities including their time-dependency. Routinely, the data are analysed at the different observation times separately. However, observations at a given time strongly depend on the previous effects and should be analysed jointly with them. To solve this problem, we developed multistate models for occurrence of developmental malformations and live events in zebrafish embryos exposed to eight concentrations of valproic acid (VPA) the first five days of life. Observations were recorded daily per embryo. We statistically infer on model structure and parameters using a numerical Bayesian framework. Hatching probability rate changed with time and we compared five forms of its time-dependence; a constant rate, a piecewise constant rate with a fixed hatching time at 48 hours post fertilization, a piecewise constant rate with a variable hatching time, as well as a Hill and Gaussian form. A piecewise constant function of time adequately described the hatching data. The other transition rates were conditioned on the embryo body concentration of VPA, obtained using a physiologically-based pharmacokinetic model. VPA impacted mostly the malformation probability rate in hatched and non-hatched embryos. Malformation reversion probability rates were lowered by VPA. Direct mortality was low at the concentrations tested, but increased linearly with internal concentration. The model makes full use of data and gives a finer grain analysis of the teratogenic effects of VPA in zebrafish than the OECD-prescribed approach. We discuss the use of the model for obtaining toxicological reference values suitable for inter-species extrapolation. A general result is that complex multistate models can be efficiently evaluated numerically.

Keywords

Multistate model, Zebrafish embryo, Malformations, Survival analysis, Dose-response analysis.

1 Introduction

For human risk assessment, many *in vitro* and *in vivo* tools have been developed to improve the extrapolation of toxic effects from animals or non-animal systems to humans. Given its multiple advantages such as simple husbandry, transparency, small size, rapid development, documented genetic homologies to humans and animal welfare considerations, the zebrafish (*Danio rerio*) embryo has become an attractive species for toxicological and pharmacological testing (d'Amora and Giordani, 2018; Driessen et al., 2013; Hill et al., 2005; Hodgson et al., 2018; Kantae et al., 2016; Kanungo et al., 2014; Larisch et al., 2017; MacRae and Peterson, 2015; Sipes et al., 2011). For chemical toxicity testing in the zebrafish embryo, the OECD test guideline (TG) 236 on the Fish Embryo Acute Toxicity Test (Busquet et al., 2014; Embry et al., 2010; OECD, 2013) prescribes to record counts of four core endpoints (coagulation, lack of heart beat, somite formation and tail detachment) in addition to reporting malformations at various times following exposure of individual fish embryos to five test concentrations. The same guideline recommends performing a separate time-stratified dose-response analysis for the determination of the lowest observed effect (LOEC) as well as LC_{50} and LC_{100}. Only recently, full dose-response curves for all observations at different times have been recommended as an advanced approach for a more detailed data analysis (Brotzmann et al., 2020).

The standard stipulated by OECD TG 236 might also be improved by the following quantitative and qualitative principles: First, translation of toxicity assessment from zebrafish to humans should account for differences in pharmacokinetics between the two species (Quignot et al., 2014); this has been addressed in a previous article on physiologically-based pharmacokinetic (PBPK) modeling for the zebrafish embryo (Siméon et al., 2020a, 2020b). Second, inter-species extrapolation might be improved by a more in-depth mechanistic analysis of the animal model, which would also contribute to a refinement of the test method in line with the 3Rs principle (Russell and Burch, 1959). Although the design of mechanistic models for developmental toxicity is difficult (Battistoni et al., 2019), at least the general features of the malformation process should be accounted for. In quantitative terms, there are two major problems with time-stratified statistical analyses: (1) Events recorded at a given time are at least partly dependent on previously manifested effects; as a consequence, malformations occurrences are not independent. For example, embryos may die before completion of heart development; since dead

embryos cannot subsequently develop heart malformation, the prevalence of cardiac malformation at lethal test concentrations seems to decrease − a typical intricacy of competing risks. (2) Furthermore, stratifying the data by times of observation reduces the number of data points per analysis, potentially making their analysis impossible or at least reducing their power considerably.

The standard analysis of such time-of-occurrence data uses survival models (Bradburn et al., 2003a, 2003b; Clark et al., 2003a, 2003b; Jiang and Fine, 2007; Lee and Go, 1997), which are commonly used in clinical studies or failure-time analyses (Gould et al., 2015; Karmen et al., 2019). Multistate Markov models (Andersen et al., 2012; Beyersmann et al., 2012; Jackson, 2011) are extensions of survival models able to deal with competing risks.

The present communication illustrates the application of multistate models to the analysis of zebrafish embryo malformation data following exposure to a model compound: valproic acid (VPA). VPA is a notorious teratogenic antiepileptic and thymoregulator widely used in humans, even though its mechanism of action is not fully understood (Chuang et al., 2012; Fathe et al., 2014; Phiel et al., 2001). The final model includes five physiological states the zebrafish embryos may be in: "normal" (*i.e.*, pre-hatched), "hatched", "non-hatched with effects" (*i.e.*, malformed), "hatched with effects" and "coagulated" (*i.e.*, dead). The probabilities of transitions between states may depend on time (*e.g.*, for hatching) or on internal chemical exposure concentration (*e.g.*, for the transitions to malformed or dead states). At all times and all exposures levels, data collected were jointly analyzed, and an innovative route for solving the associated equations is proposed. Additionally, we computed internal VPA concentrations as a function of time and water-concentration using a zebrafish embryo physiologically-based pharmacokinetic (PBPK) model previously developed (Siméon et al., 2020a, 2020b). Combining PBPK and multistate effect models allows to establish concentration-time-response relationships, which can be used for the establishment of toxicity reference values.

2 Background on multistate models

Survival analysis models the statistical distribution of the time of occurrence of events of interest (*e.g.*, death), after a fixed starting time, using data from several subjects or observation units (patients, zebrafish embryos, *etc.*) (Bradburn et al., 2003a). The simplest survival model (Figure 1) considers only

two states, *e.g.*, "alive" and "dead", with no possible reversion from the second to the first. The deaths observed correspond to transitions between states 1 and 2, characterized by a hazard rate (also called transition intensity) $q_{1,2}$. This transition intensity is of fundamental interest, because it can link instantaneous exposure to drugs or toxicants and observed death times in exposed individuals.

Survival analysis can be generalized to multistate Markov models, which describe transitions, eventually reversible, between more than two states and can be used to obtain the distribution of residence time in the various states (Andersen et al., 2012; Beyersmann et al., 2012; Welton and Ades, 2005). If several states are accessible from another one (Figure 2), we have a competing risks process.

Therefore, multistate Markov models form a general class of competing risk models, and can consider many intermediate events (Farewell and Tom, 2014; Putter et al., 2007) (see for example Figure 3).

In these models, a set of transition intensities $q_{i,j}(t,z(t))$ determines the risk of moving from state i to state j, as a function of time and individual-specific variables $z(t)$, which can themselves be time-dependent (Jackson, 2011). An important class of variables $z(t)$ are exposures to hazards, measured for example by the body or organ concentration of a toxicant.

Assume we start from state $X(0) = i$ among a set of states X, and call $p_{i,j}(t)$ the transition probability from state $X(0) = i$ to state $X(t) = j$ during time length t. Said otherwise, by definition:

$$p_{i,j}(t) = P\{X(t) = j|\mathrm{X}(0) = i\} \quad (1)$$

Determining these transition probabilities is important, because they condition the likelihood of observing transitions between states as time passes (*i.e.*, the data). To calibrate a multistate model from data, we need to link variables $z(t)$, transition intensities $q_{i,j}(t,z(t))$, transition probabilities $p_{i,j}(t)$ and observations. A system of ordinary differential equations (ODE), called forward Kolmogorov differential equations, can be written to link transition intensities and probabilities (Brinks and Hoyer, 2018; Fisz, 1976). This system of equations can be expressed quite simply with a matrix notation:

$$\frac{\partial \boldsymbol{P}(t)}{\partial t} = \boldsymbol{P}(t)\boldsymbol{Q}(t) \quad (2)$$

The transition probabilities $p_{i,j}(t)$ are the elements of the probability matrix $\boldsymbol{P}(t)$, and the corresponding transition intensities $q_{i,j}(t,z(t))$ are the elements of the transition matrix $\boldsymbol{Q}(t)$. The ODE system is obtained by element-wise multiplication of the rows of matrix $\boldsymbol{P}(t)$ by the columns of matrix $\boldsymbol{Q}(t)$. The probability matrix $\boldsymbol{P}(t)$ we need is the solution of the above system of differential equations (Jackson, 2011) with suitable initial conditions (*i.e.*, $\boldsymbol{P}(0) = I$, the identity matrix).

The final link between transition probabilities and data likelihood depends on the observational process, *i.e.* on study design and type of data collected (Jackson, 2011). In an intermittently-observed process, like the OECD 236 test, the data is a set of n observation times $\{t_{1,m}, \ldots, t_{n,m}\}$ and states $\{s(t_{1,m}), \ldots, s(t_{n,m})\}$ in which individual m was at these times. This is also called time-censored "panel" data, because the exact time of occurrence of the malformation is not known with precision. The likelihood of observing a pair of successive states $(s(t_{k,m}), s(t_{k+1,m}))$ at times $(t_{k,m}, t_{k+1,m})$ is simply, by definition, the probability of going from state $s(t_{k,m})$, to state $s(t_{k+1,m})$ during the length of time $t_{k+1,m} - t_{k,m}$:

$$L_{k,m} = p_{s(t_k,m)s(t_k+1,m)}(t_{k+1,m} - t_{k,m}) \tag{3}$$

The full likelihood of the data, L, is simply the product of each $L_{k,m}$ over all observed transitions and individuals. This essentially amounts to assuming a multinomial distribution of transition counts. Observations times do *not* need be equally spaced for Eq. 3 to apply.

Two simplifications of the above general model can be made:

Time-homogeneous multistate Markov models consider that transition intensities only depend on the starting state and their parameters do not depend on time (Putter et al., 2007; Zare et al., 2014):

$$q_{i,j}\big(t, z(t)\big) = q_{i,j} \tag{4}$$

In this case, the ODE system is linear and can be solved analytically using the so-called matrix exponentials (Jackson, 2011; Jones et al., 2017), which can be quite computationally expensive (Tsiros et al., 2019).

Time-homogeneous semi-Markov models consider that transition intensities depend only on the starting state i and on the entry time t_j into the next state j (Meira-Machado et al., 2009):

$$q_{i,j}\big(t, z(t)\big) = q_{i,j}\big(t - t_j\big) \tag{5}$$

The model can also be extended to hidden Markov processes if, for example, the exact states may be identified wrongly or indirectly by the observations. We will not consider such hidden Markov models here, because the definition and observation of malformations are quite crisp in the zebrafish embryo. However, we are interested in general non-time-homogeneous models, because transition intensities may change continuously with zebrafish embryo age. In particular, embryos grow and their internal concentrations of toxicants changes continuously with time (Siméon et al., 2020a). In that case, the matrix exponential solution does not apply, and numerical integration of the ODE system is the best solution, which we use.

3 Materials and methods

3.1 Zebrafish embryo malformation data

Spawning groups were obtained from a colony of adult wild-type zebrafish (*Danio rerio*) of the 'Westaquarium strain' maintained at the fish facilities of the Aquatic Ecology and Toxicology Group at the University of Heidelberg (licensed under no. 35-9185.64/BH). All experiments were based on OECD TG 236 and associated work (Braunbeck et al., 2015; Braunbeck and Lammer, 2006; Embry et al., 2010; Lammer et al., 2009; OECD, 2013); however, as an adjustment, for better observation of effects, exposure of the embryos was extended to 120 hours post-fertilization (hpf). According to Strähle et al. (2012), an extension of the exposure period to 120 hpf is possible without violation of the current EU animal welfare legislation (EU, 2010) and is recommended in cases of inconclusive observations after 96 hpf.

For each test, embryos were immersed in the test solutions at the 16 cell-stage at the latest (≤ 90 min; before cleavage of blastodisc). To start exposures with minimum delay, twice the number of eggs eventually needed per treatment group were picked from the same batch of eggs and transferred into 100 ml crystallization dishes with the test concentrations or negative (artificial water according to ISO 7346-3) (OECD, 2013) and positive controls (24.7 µM of 3,4-dichloroaniline). At 3 hpf at the latest, viable eggs were selected for normal development under the stereomicroscope (≥ 30-fold magnification) and transferred to a final volume of 1 ml into 24-well plates (one embryo per well). To account for

potential adsorption of the test solutions to the plastic walls of the wells, the plates had been pre-exposed to the test solutions for 24 h. Test solutions were replaced at 24, 48, 72 and 96 hpf (without changing the well plates). Embryos were not dechorionated and hatched on their own at approximately 72 hpf. VPA was tested in three independent runs at seven test concentrations ranging from 6.25 to 400 µM regularly spaced by a factor of 2. The control group comprised a total of 120 embryos, while VPA-exposed groups consisted of 60 embryos per test concentration. Note that each embryo was in its own well of the plates and was observed individually.

Prior to medium replacement and upon completion of the test at 120 hpf, embryos were analyzed for macroscopically discernable alterations including the four morphological lethal core endpoints listed by OECD TG 236 (OECD, 2013) (coagulation, lack of heart-beat, somite formation and tail detachment) as well as any additional sublethal observation including, *e.g.*, scoliosis/lordosis, eye deformation, loss of pigmentation, various types of edemata and general skeletal deformations (Braunbeck et al., 2005; Hollert et al., 2003; Nagel, 2002). Up to 55 developmental malformations could be recorded including a semi-quantitative graduation of severity for at least part of the adverse effects, *e.g.* heart-beat reduced, severely reduced or missing. All the effects recorded and their corresponding codes are summarized in Supplemental Material Table S1.

For documentation, morphological alterations were recorded with a Zeiss Axio Cam ICc1 camera mounted on a Zeiss Olympus CKX41 microscope (Carl Zeiss, Oberkochen, Germany) and analyzed using the Zeiss imaging program Zen lite 2011.

Since the addition of buffer may affect the development of the embryos, the artificial water pH was not buffered, but measured independently at the various test concentrations of VPA (Supplemental Material Table S2). Water pH decreased with VPA concentration, which was expected and taken into account for internal concentration estimation by the PBPK model (see below). The pH of the control solutions was at 7.75. According to the OECD test guideline, water pH values between 6.5 and 8.5 are not expected to induce particular toxicity and pH should be kept in that range (Busquet et al., 2014; OECD, 2013). Since the pH dropped below 6.5 for the highest VPA test concentration (800 µM), leading to rapid death of the embryos within a couple of hours by a mechanism completely independent of malformations and subsequent lethality, data obtained at 800 µM were excluded from this analysis.

3.2 Zebrafish embryo malformation multistate model

3.2.1 Model structure

The proposed multistate model of malformations in the zebrafish embryo includes five states, as shown in Figure 3: “normal”, N, *i.e.*, pre-hatched; “hatched”, H; “non-hatched with effects”, E, *i.e.*, malformed; “hatched with effects”, EH; “coagulated”, C, *i.e.*, dead. We aggregated the various malformations into non-hatched with effects and hatched with effects for simplicity.

The coagulated state is the only absorbing state (state from which you cannot return). Back transitions are possible between the normal and non-hatched with effects states, and between the hatched and hatched with effects states, since we had observations of such transitions.

The elements of the associated transition probability matrix $\boldsymbol{P}(t)$ are the transition probability $p_{i,j}(t)$. According to our model, the transition probabilities $P_{H,N}(t)$, $P_{C,N}(t)$, $P_{EH,N}(t)$, $P_{C,H}(t)$, $P_{H,E}(t)$, $P_{C,E}(t)$, $P_{EH,E}(t)$, $P_{H,C}(t)$, $P_{EH,C}(t)$, and $P_{C,EH}(t)$ will always be null, because there is no path between the first and the second state. Hence, there is no need to compute a differential equation for them. Note however, for example, that while there is no direct transition from the normal to the hatched with effects states, this transition is still possible *via* the non-hatched with effects state, so $p_{N,EH}$ is not null. The transition probability $P_{C,C}(t)$ is equal to 1, because C is absorbing state. Therefore, $\boldsymbol{P}(t)$ can be written as follow:

$$\boldsymbol{P}(t) = \begin{pmatrix} p_{N,N}(t) & p_{N,H}(t) & p_{N,E}(t) & p_{N,C}(t) & p_{N,EH}(t) \\ 0 & p_{H,H}(t) & 0 & 0 & p_{H,EH}(t) \\ p_{E,N}(t) & p_{E,H}(t) & p_{E,E}(t) & p_{E,C}(t) & p_{E,EH}(t) \\ 0 & 0 & 0 & 1 & 0 \\ 0 & p_{EH,H}(t) & 0 & 0 & p_{EH,EH}(t) \end{pmatrix} \quad (6)$$

The associated transition matrix, $\boldsymbol{Q}(t)$ only includes *direct* transitions between states (see Figure 3) and its row sums have to be equal to 0. Therefore, the diagonal transition intensities $q_{N,N}(t)$, $q_{H,H}(t)$, $q_{E,E}(t)$, $q_{C,C}(t)$, and $q_{EH,EH}(t)$ must be respectively:

$$q_{N,N}(t) = -\big(q_{N,H}(t) + q_{N,E}(t) + q_{N,C}(t)\big) \quad (7)$$

$$q_{H,H}(t) = -\, q_{H,EH}(t) \quad (8)$$

$$q_{E,E}(t) = -(q_{E,N}(t) + q_{E,C}(t) + q_{E,EH}(t)) \quad (9)$$

$$q_{C,C}(t) = 0 \quad (10)$$

$$q_{EH,EH}(t) = -q_{EH,H}(t) \quad (11)$$

The transition matrix ***Q(t)*** is therefore:

$$\boldsymbol{Q}(t) = \begin{pmatrix} -q_{N,H}(t) - q_{N,E}(t) - q_{N,C}(t) & q_{N,H}(t) & q_{N,E}(t) & q_{N,C}(t) & 0 \\ 0 & -q_{H,EH}(t) & 0 & 0 & q_{H,EH}(t) \\ q_{E,N}(t) & 0 & -q_{E,N}(t) - q_{E,C}(t) - q_{E,EH}(t) & q_{E,C}(t) & q_{E,EH}(t) \\ 0 & 0 & 0 & 0 & 0 \\ 0 & q_{EH,H}(t) & 0 & 0 & -q_{EH,H}(t) \end{pmatrix} \quad (12)$$

The model ODE system is obtained by a simple matrix multiplication of ***P(t)*** by ***Q(t)***:

$$\frac{\partial p_{N,N}(t)}{\partial t} = -\,p_{N,N}(t) \times \left(q_{N,H}(t) + q_{N,E}(t) + q_{N,C}(t)\right) + p_{N,E}(t) \times q_{E,N}(t) \quad (13)$$

$$\frac{\partial p_{N,H}(t)}{\partial t} = p_{N,N}(t) \times q_{N,H}(t) - p_{N,H}(t) \times q_{H,EH}(t) + p_{N,EH}(t) \times q_{EH,H}(t) \quad (14)$$

$$\frac{\partial p_{N,E}(t)}{\partial t} = p_{N,N}(t) \times q_{N,E}(t) - p_{N,E}(t) \times \left(q_{E,N}(t) + q_{E,C}(t) + q_{E,EH}(t)\right) \quad (15)$$

$$\frac{\partial p_{N,C}(t)}{\partial t} = p_{N,N}(t) \times q_{N,C}(t) + p_{N,E}(t) \times q_{E,C}(t) \quad (16)$$

$$\frac{\partial p_{N,EH}(t)}{\partial t} = p_{N,H}(t) \times q_{H,EH}(t) + p_{N,E}(t) \times q_{E,EH}(t) - p_{N,EH}(t) \times q_{EH,H}(t) \quad (17)$$

$$\frac{\partial p_{H,H}(t)}{\partial t} = -\,p_{H,H}(t) \times q_{H,EH}(t) + p_{H,EH}(t) \times q_{EH,H}(t) \quad (18)$$

$$\frac{\partial p_{H,EH}(t)}{\partial t} = p_{H,H}(t) \times q_{H,EH}(t) - p_{H,EH}(t) \times q_{EH,H}(t) \quad (19)$$

$$\frac{\partial p_{E,N}(t)}{\partial t} = -\,p_{E,N}(t) \times \left(q_{N,H}(t) + q_{N,E}(t) + q_{N,C}(t)\right) + p_{E,E}(t) \times q_{E,N}(t) \quad (20)$$

$$\frac{\partial p_{E,H}(t)}{\partial t} = p_{E,N}(t) \times q_{N,H}(t) - p_{E,H}(t) \times q_{H,EH}(t) + p_{E,EH}(t) \times q_{EH,H}(t) \quad (21)$$

$$\frac{\partial p_{E,E}(t)}{\partial t} = p_{E,N}(t) \times q_{N,E}(t) - p_{E,E}(t) \times \left(q_{E,N}(t) + q_{E,C}(t) + q_{E,EH}(t)\right) \quad (22)$$

$$\frac{\partial p_{E,C}(t)}{\partial t} = p_{E,N}(t) \times q_{N,C}(t) + p_{E,E}(t) \times q_{E,C}(t) \quad (23)$$

$$\frac{\partial p_{E,EH}(t)}{\partial t} = p_{E,H}(t) \times q_{H,EH}(t) + p_{E,E}(t) \times q_{E,EH}(t) - p_{E,EH}(t) \times q_{EH,H}(t) \quad (24)$$

$$\frac{\partial p_{EH,H}(t)}{\partial t} = -\,p_{EH,H}(t) \times q_{H,EH}(t) + p_{EH,EH}(t) \times q_{EH,H}(t) \quad (25)$$

$$\frac{\partial p_{EH,EH}(t)}{\partial t} = p_{EH,H}(t) \times q_{H,EH}(t) - p_{EH,EH}(t) \times q_{EH,H}(t) \quad (26)$$

As mentioned above, the other differentials, *e.g.* for $P_{H,N}(t)$, are null. Similarly, $P_{C,C}(t)$ is always equal to 1 and its differential is null. To check the model equations, we can calculate the probability mass balance, which should always be equal to 5, because the initial condition for $\boldsymbol{P}(t)$ is the identity matrix of order 5:

$$Probability\ Mass\ Balance = \sum_{i=1}^{i=5}\left(\sum_{j=1}^{j=5} P_{i,j}(t)\right) = 5 \quad (27)$$

3.2.2 Transition hatching rate sub-model

The zebrafish embryo hatching rate is clearly non-constant with time: in normal conditions, hatching is never observed before 24 hpf, because the embryo is not ready for hatching (Kimmel et al., 1995). We tested five sub-models linking $q_{N,H}(t)$ and time and retained a piecewise constant model with estimated hatching time on the basic of the Akaike information criterion (see Supplemental Material Section 2). In the five-state model (Figure 3), hatching can happen to normal embryos (state N) and malformed

embryos (state E). The piecewise constant sub-model was used for both hatching rates in the final model, leading to three parameters to estimate: hatching time t_h, and the values of the intensities $q_{N,H}$ and $q_{E,EH}$ after t_h (before t_h those intensities were assumed to be null).

3.2.3 VPA concentration-dependent transition intensities

To find appropriate functional forms for the links between VPA internal concentration, *C(t)*, and transition intensities, we first estimated the remaining intensities ($q_{E,N}$, $q_{EH,H}$, $q_{N,E}$, $q_{N,C}$, $q_{E,C}$, and $q_{H,EH}$) separately for each exposure group. We then determined visually adequate sub-models: exponential for $q_{E,N}$ and $q_{N,E}$; linear for $q_{EH,H}$, $q_{N,C}$ and $q_{E,C}$; and Hill for $q_{H,EH}$ (a comparison of the final sub-models with the separate estimates is shown in Supplemental Material Figure S6). The corresponding equations are:

$$q_{E,N}(C(t)) = q_{E,N}(0) \times \mathrm{e}^{k_{q_{E,N}} \times C(t)} \tag{28}$$

$$q_{N,E}(C(t)) = q_{N,E}(0) \times \mathrm{e}^{k_{q_{N,E}} \times C(t)} \tag{29}$$

$$q_{EH,H}(C(t)) = q_{EH,H}(0) + k_{q_{EH,H}} \times C(t) \tag{30}$$

$$q_{N,C}(C(t)) = q_{N,C}(0) + k_{q_{N,C}} \times C(t) \tag{31}$$

$$q_{E,C}(C(t)) = q_{E,C}(0) + k_{q_{E,C}} \times C(t) \tag{32}$$

$$q_{H,EH}(C(t)) = \frac{q_{H,EH_{max}} \times C(t)^n}{EC_{50}{}^n + C(t)^n} \tag{33}$$

where $q_{E,N}(0)$, $q_{N,E}(0)$, $q_{EH,H}(0)$, $q_{N,C}(0)$, and $q_{E,C}(0)$ the transition intensities at VPA concentration zero; $kq_{E,N}$, $kq_{N,E}$, $kq_{EH,H}$, $kq_{N,C}$, and $kq_{E,C}$ are constants; $q_{H,EH_{max}}$ is the maximal transition intensity value of $q_{H,EH}$, n its Hill exponent, and EC_{50} the concentration for which $q_{H,EH}$ is equal to 50% of its maximal value.

3.3 Coupling to a zebrafish embryo PBPK model

The internal VPA concentration *vs.* time profiles (whole body excluding yolk) driving the above transition intensities (Eqs. 28 to 33) were computed for the various nominal exposure concentrations tested using our previously published zebrafish embryo PBPK model (Siméon et al., 2020a, 2020b). The decrease in water pH with increasing VPA test concentration was taken into account in our computation. The pre-computed internal concentration *vs.* time profiles were read at the start of the multistate calculations and interpolated linearly as needed during integration of the ODE equations. Figure S1, in Section 1 of the Supplemental Material, shows the value of internal concentration as a function of nominal concentration and time (on the Figure, discontinuities are due to the daily medium changes).

3.4 Statistical inference on multistate model parameters

The multistate parameters of the joint PBPK-multistate model were calibrated using Markov chain Monte Carlo (MCMC) simulations in a Bayesian statistical framework (Bernillon and Bois, 2000; Bois, 2012; Smith and Roberts, 1993). MCMC methods can draw samples from the joint multi-dimensional posterior distribution of the model parameters, given a statistical model, prior parameter distributions, and data for which a likelihood function can be computed (Bois, 2009).

We used the following statistical model: For most parameters ($q_{E,N}(0)$, $q_{N,E}(0)$, $q_{EH,H}(0)$, $q_{N,C}(0)$, $q_{E,C}(0)$, $kq_{E,N}$, $kq_{N,E}$, $kq_{EH,H}$, $kq_{N,C}$, $kq_{E,C}$, $q_{H,EH_{max}}$, n, and EC_{50}) we estimated one common value given the data from all experimental exposure groups. The three hatching parameters (t_h, $q_{N,H}$, and $q_{E,EH}$) were estimated at the same time, but in a multi-level framework (Bois, 2012): As indicated in the experimental protocol, because the embryos of each exposure may not have been exactly at the same morulation stage at the time of selection, and due to biological variability, average hatching time varied randomly between exposure groups. Therefore, at the higher level of the hierarchy, we have three mean hatching parameters, and for individual exposure groups we assume that group-specific values are distributed randomly around the common mean:

$$\boldsymbol{\varphi}_i \sim \text{Lognormal}(\boldsymbol{\varphi}, \boldsymbol{\sigma}) \tag{34}$$

where φ is the geometric mean in log-space of either t_h, $q_{N,H}$, or $q_{E,EH}$, and σ is the corresponding standard deviation in log-space. The lognormal distribution for t_h was truncated to fall between 0 and 5 (days). So, for hatching, three mean parameters and eight triplets of group-specific parameters were estimated.

Non-informative uniform priors were used for most transition intensities' parameters. Parameters $q_{EH,H}(0)$ and $k_{qEH,H}$ were sampled with correlated constraints to respect the fact that $q_{EH,H}$ must be positive. For the SDs of the dose-group-specific hatching parameters, very wide truncated normal distributions were used (Table 1).

The data were censored by interval. The data likelihood was therefore computed as the product of terms given by Eq. 3 over all transitions and individual embryos.

Two simulated Markov chains of 30,000 iterations each were generated, starting from a set of parameter values randomly sampled from their priors. Convergence of the last 15,000 iterations of two chains was assessed using Gelman and Rubin's Rhat convergence criterion (Gelman and Rubin, 1992).

3.5 Calculation of predicted embryos counts

The number of zebrafish embryos in state *j* at time *t*, $n_j(t)$, is conditioned by the probability to go from each initial state *i* to state *j*, according to the number of zebrafish embryos in initial state *i* at starting time *t0*, $n_i(t0)$, and is defined by:

$$n_j(t) = \sum\left(n_i(t0) \times P_{i,j}(t)\right) \tag{35}$$

3.6 Calculation of reference toxicity values

Internal concentration values (internal EC_{10}) leading to 10% change from control on intensities were simply computed by mathematical inversion of Equations 28 to 33 (see Supplemental Material, section 4). This was done for all the parameter values sampled by MCMC to obtain distributions of EC_{10} values reflecting uncertainty. External concentrations values leading to 10% effects (external EC_{10}) on probabilities were computed by running the joint PBPK-multistate model for a grid of VPA water concentrations. Estimated probabilities were recorded for a dense set of times, leading to sets of three-dimensional exposure-time-response relationships. Contour at z-values 0.1 or 0.9 (depending on

whether P increases or decreases with VPA exposure, respectively) were obtained with the "*contour*" function of the *R* statistical package for the maximum posterior parameter set and 100 random parameter sets, to estimate uncertainty.

3.7 Software

The dynamic model simulations and MCMC calibrations were performed with GNU MCSIM version 6.1.0 (https://www.gnu.org/software/mcsim) (Bois, 2009). We used the R software version 3.5.2 for plotting and miscellaneous statistical analyses (R Development Core Team, 2013). The model code and the MCMC specification file are shown in Supplemental Material, Section 6.

4 Results

4.1 Model calibration

4.1.1 Model fit to the data

Each MCMC chain of 30,000 iterations took 9 hours to complete on a Dell Precision 5530 laptop computer (with Intel Core i7-8850H CPU clocked at 2.6GHz) using the Linux subsystem. Figure 4 shows all observed transition counts plotted against the corresponding maximum posterior (best) model predictions. Note that the predictions are expected values, rather than integer counts, so there is a layer of randomness in the data that is not taken into account in this plot. Note also that the graph uses logarithmic scales for readability. Since null values (mostly for observations) cannot be shown on such a plot, we assigned them an arbitrary value of 0.1, only for plotting purposes; some of them fall far away from the identity line because of that artifact. Furthermore, the data are integers and the model predictions reals, so there will always be a mismatch. The model has a slight tendency to under-predict the data, but most of the points fall in the two-fold error interval. Table S4 (Supplemental material) gives all predicted and observed values. If we disregard the null values, the most outlying points are from the control group (three transitions observed from normal to hatched with effects on day 4, and two embryos remaining in the effect state on day 5) or correspond to observations of one or two transitions in the exposed groups while the model predicts a low probability for those events.

Figure 5 shows the detailed fit of the coupled PPBK-multistate model to the number of transitions between states, for the control group and the highest exposure concentration. The model is continuous in time, but we coarsely discretized its outputs when plotting them with the observations to better focus on the latter. The model predictions for all concentrations are presented in the Figure S8 of the Supplemental Material. As VPA exposure concentration increases, we observe a decrease in transitions to the hatched (H) state, and an increase in transitions to the malformed (E), hatched malformed (EH), and dead states (C). Yet, most exposure concentrations have little effect. Time has a strong effect on transitions between states, because hatching intensities are not constant in time and because of competing risk and depletion effects. For example, in the control group, the count of transitions from normal to hatched states is high at the end of day 3, because $q_{N,H}(t)$ become non-null, but it decreases at the end of days 4 and 5, simply because there are no more embryos left to hatch. At the highest VPA concentration, the count of transitions from normal to malformed (N to E) is highest at the end of day 1 because of VPA, but then decreases because of depletion of the normal fish. At the same concentration, the count of malformed to malformed transitions (E to E, *i.e.*, the count of fish staying in the malformed state) first increases and then decreases, because malformed embryos start either hatching or dying. The predicted counts of embryos malformed (E), hatched with effects (EH), or dead (C) is shown in Figure 6 as a function of the time for each exposure concentration (the predicted counts for all states are shown in Figure S9 of the Supplemental Material). The number of embryos hatched with effects or dead increases with the concentration and the time of exposure, but the number of malformed embryos decreases with time and would not be well nor coherently modeled by standard sigmoidal dose-response curves. Note that there is less information in those count data than in the transition data (less data points for the same experiments), and that is why the model is fitted to the transition data.

4.1.2 Transition intensities' estimates

The marginal posterior parameters distributions obtained by MCMC calibration with the embryo transition count data are summarized in Table 2. The parameters were estimated with a precision ranging from about 1.4% (for t_h) to 89% ($k_{qN,C}$). A parameter posterior precision depends largely on the number of observed transitions that are conditioned by that parameter. The posterior summaries of the concentration-group-specific hatching parameter estimates are given in Table S5 of the Supplemental

Material; there was no obvious correlation of those with concentration, but the hatching intensity is twice as low in malformed embryos as in normal embryos. Note, however, that MCMC sampling yields parameter vectors (sets) out of the joint posterior distribution of all parameters calibrated, that means that correlations between parameter estimates can be viewed, summarized and taken into account in predictive simulations. The correlations between parameters of the same intensity sub-models are shown in Figure S7 of the Supplemental material.

It is challenging to directly interpret the values of Table 2. More telling is Figure 7, which shows the relationships between transition intensities and internal concentration (Eqs. 28 to 33) obtained when using the above posterior parameter estimates. Those are in fact the fundamental dose-response relationships of the model. First, there are large differences between their shape and range of values. The two largest correspond to rates of malformations. Note that the toxicity to hatched embryos starts occurring at lower dose than toxicity to the non-hatched embryos, but saturates above 0.01 M. Next comes the suppression of reversions from malformed to normal or hatched malformed to hatched. Such events are sometimes observed, but become even less likely as the internal concentration of VPA increases. Finally, direct mortality is relatively low at the concentrations tested, but seems to increase linearly with internal concentration.

4.2 Calculation of reference toxicity values

Regulatory guidelines on toxicity testing typically recommend to analyze the concentration-response profiles obtained experimentally to identify points of departure from baseline (healthy) state. Such points of departure can, for example, be exposure concentrations (EC_{10}) leading to 10% effect. They can then be transposed to other species for risk assessment. Should such values be sought, they could be derived on the basis of internal or external concentrations and for effects measured by transition intensities ($\boldsymbol{q}$), or transition probabilities ($\boldsymbol{P}$). Out of the four possible combinations, we chose to present internal concentrations values leading to 10% effects (EC_{10}) for intensities, and external EC_{10} for probabilities.

The distribution of internal EC_{10} values for the reversion intensities $q_{E,N}$ and $q_{EH,H}$ are shown in Figure 8. Their maximum posterior estimates are 5.5×10^{-4} and 4.7×10^{-3} M, respectively. The EC_{10} values for the other intensities, by increasing order, are 1.3×10^{-5} M for direct death ($q_{N,C}$), 6.2×10^{-5} M for death of

malformed embryos ($q_{E,C}$), 2.7×10^{-4} M for malformations in non-hatched embryos ($q_{N,E}$), and 3.0×10^{-3} M for malformation in hatched embryos ($q_{H,EH}$). Hence, the intensities of malformations or their reversions appears to be less sensitive than death to VPA internal concentration, as judged by the sole EC_{10}.

Transition intensities are fundamental quantities in the model, but are not directly related to observations of fish counts in the different states, which may be more relevant for risk assessment. Transition probabilities condition more directly fish counts. Internal concentration EC_{10} values for probabilities could be obtained, but it may be more relevant again for risk assessment, in some context, to obtain EC_{10} values based on external concentration. This can be done, but in that case, the EC_{10} depends both on exposure concentration and time. Figure 9, top row, shows the 3D water concentration-time-transition probability relationships for transition from normal to malformed, $P_{N,E}$, normal to dead, $P_{N,C}$. and normal the healthy hatched ($P_{N,H}$). $P_{N,E}$ shows a maximum a mid-exposure concentration and mid-time, because of competition effects with other states (dead, hatched, hatched malformed). $P_{N,C}$ has a less complex shape and increases nonlinearly with time and external concentration, because death is an absorbing state from which the embryo cannot escape. $P_{N,H}$ peaks late in time, but declines quickly with concentration. The three bottom panels of Figure 9 show the corresponding EC_{10} estimates, which are functions of time. These are just contours plots of the 3D surface at 10% of the maximum probability. For $P_{N,E}$, there is clearly a sensitive period of time just before hatching, when VPA has had time to exert its effect and hatching has not depleted the pool of non-hatched embryos. At that time, the EC_{10} is around 50 μM. For $P_{N,C}$, the longer the lower the EC_{10}, but it does not go lower than 200 μM, showing that dead, here also is a less sensitive toxicity endpoint. Finally, $P_{N,H}$ appears to be a very sensitive endpoint at later times. The expected counts of fish in the various final states will look exactly like those plots, because the probability at each point would just be multiplied by the starting count of embryos (in the normal state).

Figure 10 shows Stradivarius plots comparing external concentration EC_{10} values obtained empirically (see Table S6 in Supplemental material) with the distribution of minimal EC_{10} values based on $P_{N,E}$, and $P_{N,C}$. Those minimal values were obtained by taking the minimum external concentration along each possible contour line. They fall around time 2.8 days for $P_{N,E}$, and as expected, 5 days for $P_{N,C}$. The range

of values is comparable, and the most sensitive endpoints found empirically (reduced yolk sac absorption EC_{10} at 20 μM; Pericardial edemata EC_{10} at 38 μM; Craniofacial deformation EC_{10} at 52 μM) do fall in the range of the $P_{N,E}$ based EC_{10}s we found.

5 Discussion

A five-state Markovian model was developed to simulate and predict the occurrence of developmental malformations observed in zebrafish embryo following VPA exposure. The model makes full use of time- and dose-information gathered in the time-course of an OECD 236-compliant zebrafish toxicity test (OECD, 2013). The model takes into account the possibility of reversion of malformations, concurrent transitions between states, time- and exposure-dependent transition intensities. The link to exposure is important, because it offers a semi-mechanistic explanation of the observations, whose correlation pattern is complex and forbidding time by time or endpoint by endpoint analyses. Using PBPK-predicted internal concentration makes the model even more mechanistic and offers a way to take into account the large pharmacokinetic differences between human and zebrafish embryos, if quantitative extrapolation of the results to humans is sought. However, the model has limitations, as discussed below.

5.1 Model structure and fit to the data

Our model makes large simplifications that can partly explain the relative misprediction of low transition count data. For simplicity, and because we wanted to introduce a general framework, the model aggregates in only five states some 55 malformation types which occurred with different intensities at different times. Differences in intensities could partly be explained by different organ concentrations of the chemical or different organ susceptibilities to malformations. A finer subdivision of states by organs would allow the use of organ-specific measures of internal exposure (which could be predicted by the PBPK model) and better model the effect of malformations' type and number. The total number of malformations or the number of affected organs could be computed and considered as measures of toxicity in predictive simulations, to define reference values, for example. This would still require calibration of the multistate model parameters, because some chemicals might have an affinity for an organ without causing effects to it. A finer grain model would also be more mechanistic and predictive

after extrapolation to another species. Low frequency events would be analyzed joinly, with better statistical power (power calculations could also be done to optimize experimental design for specific outcomes). The relevance of rare events in a cross-species prediction context is another matter. Extrapolation of results and correlation between different species should always be made with utmost care, not only with respect to different time-scales of development, but also with regard to differences in molecular processes relevant for drug-induced effects, namely metabolism and toxicokinetic capacities. In any case, we do not claim at all to provide the ultimate solution to toxicity assessment with only zebrafish embryo tests. Indeed, evidence from organ-specific assays and different assay systems should also be considered. As to timing, it is obvious, for example, that heart malformations will not be observed before heart formation, *etc.*, and this is not accounted for in the model at the moment. Note also that some malformation types correspond in fact to various degrees of severity of the same malformation (*e.g.*, for the heart). It is common to use multistate models to describe such progressions by a linear chain of states (Cannon et al., 2017), and those could be used here. However, one problem arises when partitioning effects into finer states: There is a combinatorial multiplication of pluri-malformation states (*i.e.*, embryos suffering from several malformations at the same time should be in specific states). We do not have a solution to that problem at the moment.

A simplified three-state model was used to find the most appropriate way to describe the time-dependence of hatching intensity. Two hatching sub-models had almost equal AIC values (a piecewise constant model with estimated hatching time and a Hill's model). We preferred the former, and data at a finer time scale (*e.g.*, every few hours) would be needed to refine the model. Note that observation times do not need at all to be equally spaced when using a multistate framework. There was an apparent random effect in hatching intensity between VPA exposure groups, and we adopted a hierarchical approach to model it. *A posteriori*, there are not huge or very significant differences in hatching between exposure groups, and only one set of hatching parameters might suffice. While there is no obvious effect of VPA on hatching time or extent, the model estimates separately hatching in normal and malformed embryos, and we found the latter to have a reduced hatching rate. Please note that one type of malformations is "lack of hatch" and could be modeled as a censored piece of data on hatching time.

The shape of the individual sub-models linking internal concentration to transition intensities was determined with a standard two-step approach: First, estimate transition intensities for each exposure group, independently; second, determine empirically the relationship of the estimates with concentration. Another route for determining those relationships would be to link them to mechanistic quantitative AOP models (Battistoni et al., 2019), but that would require much more data. We did find that all six transitions intensities, not including hatching, discussed above, were VPA concentration-dependent. A more principled method would compare systematically various sub-model fits, or use a non-parametric model able to describe all of them. For example (see Figure S6), a negative exponential function might be better suited to model $q_{EH,H}$ relationship with VPA concentration. It might not make much difference though, because reversion from malformed hatched embryos are very few. Furthermore, the computational burden would have been much higher, and we would not have had enough concentration levels to improve dramatically and validate the intensity sub-models (the four lowest VPA concentrations tested had little effects). There also seems to be a small but consistent burst of effects at low concentrations in Figure S6. We wanted to avoid overfitting and did not try to capture this pattern. We also think that this early burst is an artefact induced lumping several types of malformations, but examining that hypothesis would require a more refined model, as discussed above. Remember, though, that the points shown on Figure S6 do not represent data, but are statistical estimates of transition intensities obtained from individual concentration data analysis. It is tempting to consider Figure S6 as a visual fit check, but in fact it compares estimates from two very different models.

The actual comparison of data and model predictions (Figures 4 and 5) shows a good agreement. Note that at the highest VPA concentration actually tested (800 µM) all embryos had died within the first day, potentially through a direct effect on water pH. We focused on malformations and excluded the 800 µM concentration and associated data from our analysis. The question remains of the effect of pH at the lower VPA concentrations. Andrade *et al*. (2017) showed that the survival rate of fish embryos depends on pH. In their study, at pH 3.5 all zebrafish embryos died within 24 hours; under pH 5, survival was definitely reduced, but not at pH 6. The sublethal effects of pH concerned mostly cardiac function and appeared at concentrations close to the lethal ones. Understanding which part of VPA embryo toxicity is due to its acidity could be important in the context of inter-species extrapolation, if the buffering

capacity of the embryo environment was different from that of water. We account for the action of pH on VPA kinetics (whose description is much improved when considering pH) and indirectly on its effects. However, we did not model the action of pH itself and that may have biased somewhat our dose-response sub-models. We could have modeled this effect by adding a separate term to the death intensity sub-model. However, clearly differentiating deaths due to pH from those due to VPA would require doing experiments where only pH would change.

There are seven *reverse* transitions going from malformed to normal-looking embryos, and 19 transitions on day 5 from malformed hatched embryos to normal-looking ones. The malformations observed were different in hatched *versus* unhatched states: Malformations "Delayed body development", "Reduced eye development", "Missing heart beat" and "Severely reduced blood flow" were more frequently observed in the unhatched states. They were not really organ-specific, but early development stages-specific. The hatched state coincided with further organ differentiation and growth, and there we observed specific organs' malformation such as "Pericardial edema", "Reduced blood flow with pooling at yolk", "Scoliosis", *etc*. For reversions from the malformed unhatched state back to normal (E-N) were: Blood pooling, Pericardial edema and Yolk edema. Reversions from the malformed hatched state back to hatched (EH-H) were: Tremor, Tailfin development impacted, Scoliosis, Craniofacial deformation, Side lying, Blood pooling, Pericardial edema and Yolk edema. So definitely, the reversions from un-hatched malformed and hatch malformed were different. We estimated the corresponding intensities and found them to be negatively affected by VPA. They could have been treated as miss-classified events by using a Hidden Markov extension of the model (Williams et al., 2020), but it seems unlikely that observation errors would be affected by VPA exposure. Furthermore, among the observed malformations (Table S1), some, like edema, are known to be reversible with the embryo's growth and its development its immune system (van der Vaart et al., 2012; von Hellfeld et al., 2020). An advantage of our global model is that "small" events such as these are identifiable, because all time and dose data are analyzed jointly. We would argue that if all 55 malformation types were analyzed, each of them would have low counts, and only such a model would be able to quantify them jointly and adequately.

5.2 Model solving and calibration

We used the transition count data, as is done in multistate modeling, rather than population count data in the various states, as would be typically done in toxicology. Population counts evolve with time and concentration, but they carry less information than transition counts, because they lose information about the trajectory of individual fish through states in time. The standard OECD protocol calls for observations once per day, but indeed, more frequent observations would be more informative. The model takes into account the fact that an observed transition, say from normal to dead, may have occurred through passing (with the day) by the malformed state, without direct observation of it. Making sure that most states are actually observed would bring more precise predictions. Note also that only the three highest doses are strongly informative about effects. Obviously, statisticians always beg for more data, but one application of the model could be to estimate the value of added information and help planning further experiments (Bois et al., 1999; Overstall et al., 2020).

General multistate models can use time-dependent transition intensities. However, in all multistate model applications and software we have seen, the matrix exponential solution is always used and cannot accommodate time-dependent intensities. The traditional solution is to approximate time dependent parameters by piecewise constant functions of time (Williams et al., 2020). However, assuming a piecewise constant process is a strong limitation which makes it difficult and inaccurate to link multistate and continuous pharmacokinetic models, and even more to make inference about the joint pharmacokinetic-pharmacodynamic model. It also increases the number of parameters to estimate. Finally, the matrix exponential solution is clumsy and slow to compute. A much more general solution is to integrate numerically the set of differential equations. Numerical integration imposes no constraints on linearity of the transition intensities or their time independence. Modern integration algorithms are also very fast and precise. Surprisingly, we seem to be the first to have used direct numerical integration of the differential system specifying P in the published literature. Curiously, the statistical community has so far preferred approximating non-linear general models with piecewise-linear models, despite the ensuing lack of flexibility and increased number of parameters to estimate. We compared the solution of the same multistate model using the well-known *msm* R package (Jackson, 2011), which uses in this case the analytical solution (much faster than matrix exponentials) and maximum-likelihood, and our

integration method in a Bayesian framework, using *GNU MCSim*. We obtain the same or more precise estimates (see Supplemental material, section 5 and Figure S10), but *GNU MCSim* took 20 minutes to obtain 2000 MCMC samples, and *msm* took 1 hour to obtain 1000 bootstrap samples.

5.3 Reference values and extrapolation

We explored ways to derive different types of EC_{10} values using our joint PBPK-multistate model, because the model (after further validation) could be used in various risk assessment contexts. For direct assessment of risk to fish in ecotoxicology, it is possible to use external (water) concentration EC_{10}s, taking fish population counts (for example, successfully hatched fish) as endpoints (Figure 9). However, for extrapolation of the results to other species, and in particular humans, endpoints corresponding to specific organs' malformation would probably be of interest for mechanistic reasons. In that case, it might be better to use internal concentration EC_{10}s for specific transition rates.

The zebrafish PBPK model allows us to obtain organ concentration values and link them to specific toxic event rates. Using a human PBPK model (Abduljalil et al., 2019), fetal concentrations could be obtained and the ensuing malformation risk to the embryo would be computed using these internal EC_{10}s. Alternatively, using reverse dosimetry (Brochot et al., 2019), the mother exposure resulting in an acceptable risk level could be computed. A more detailed model could factor in malformation severity or reversibility and fine-tune reference values and risk assessment, making them more specific of particular effects, the extrapolation of which would be warranted by mechanistic considerations. For example, VPA prenatal exposure in humans has been associated with neurodevelopmental disorders such as autism (Kozma, 2001; Ornoy, 2009; Christensen et al., 2013). The zebrafish embryo has been proposed as a model species for such effects *via* the observation of physiological impairment and neurological effects causing behavioral alterations (Zimmermann et al., 2015; Chen et al., 2018; Dwivedi et al., 2019). Relevant physiological impairments are any effect (*e.g.*, small eyes, chordal deformations, head and craniofacial deformations) which alters sensory perception and the animal's ability to react. Direct neurological effects include for example jitter and tremor, or lack of movement. Such effects, and their potential reversibility, could be included in the multistate framework we propose, and their probability of appearance could be linked to specific organ's concentration (eye, brain …). However, the actual relevance of such malformations for future behavioral changes depends on their

severity, reversibility, short- and long-term effects on the organism *etc*., and should be discussed for each substance using all the relevant evidence from multiple tests systems (including organ-specific assays), in an AOP or weight of evidence framework, for example. So far, however, the predictivity of our zebrafish embryo PBPK model for specific organ concentrations has not been cross-checked with experimental data for organ-specific concentrations. This should be done before applying it for predictions in a regulatory context. It may be reassuring that we found EC_{10} values in the range of those found with simpler methods.

Extrapolating EC_{10} values between species, with safety factors for example, is a standard procedure. But we could also take the more mechanistic approach that would transpose elements of the model itself to the target species. Assumptions would have to be made about the states to conserve or change. For example, a hatched state might or might not be relevant to humans, depending on whether human birth can be assimilated to hatching. We do not have experience with the extrapolation of transition rates. To understand better how it could work, it would be very interesting to compare rates obtained in zebrafish and rat for the same chemicals, for example. In general, some transition rates could be set to zero in the target species if they are determined to be irrelevant. This could also be motivated by health protective reasons (for example, turning off all reversion rates in human). The diagonal terms of the transition matrix $\boldsymbol{Q}$ (Eq. 12) would also have to be adjusted to respect the constraint of null row sums. To account for differences in developmental timescales between test and target species, note that transition rates are defined as probabilities per unit time. It should be possible to normalize or transform the development time scale between species. For example, assume that time to hatching is about 3 days in zebrafish and time to birth in human is about 270 days. The ratio is 90, so all zebrafish transition rates could be divided by 90 to get rough human estimates. This would "slow" time by a factor 90. All these ideas should be validated with further data and case studies.

6 Conclusion

The multistate malformation model we propose is very general and versatile. It could be applied to analyze data on many chemicals in the zebrafish and in other species, in particular rodents. There is nothing specific to the zebrafish about the general framework; the definition of the states, the possible

transitions between them would need to be adapted, and a specific PBPK model should be used, but the method would stay the same. For different chemicals, the sub-models linking malformation rates and exposure concentration would simply need to be fine-tuned with chemical-specific data. We showed how to do that using Bayesian inference with *GNU MCSim*, but a maximum likelihood approach and other software like R packages could also be used, at the eventual cost of some loss of flexibility in setting up hierarchical models for variability for example.

For the zebrafish embryo test, our model fits our VPA data well and is an obvious refinement, allowing a better use of data. We could differentiate effects of VPA on various malformation steps to a level that would be inaccessible to simpler dose-response analyses. Our framework could also lead to potential saving of animal lives through optimal design or statistical sample size calculations.

This model only sets up a framework and should be extended by considering more states and differentiating the severity of malformations. If that were done, its coupling to a PBPK would become even more relevant as specific organ concentrations could be linked to specific organ malformations for better data analyses and predictions of specific risk to humans, for example. We are aware that our approach is relatively complex and needs to be further verified with other chemicals before it could be used routinely or in any regulatory context. It would also be possible to simplify it, by using water concentration as a measure of exposure, for example. That could be justified assuming proportionality between water and body concentration. In any case, our model should help the scientific and risk assessment communities in improving toxicity data analyses.

Corresponding author

Email: frederic.bois@certara.com

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 681002 (*EU-ToxRisk*). The authors would like to thank Dr. Rachel Rose and the two anonymous reviewers for useful comments and discussion.

7 References

Abduljalil, K., Pan, X., Pansari, A., Jamei, M., Johnson, T.N., 2019. A preterm physiologically based pharmacokinetic model. Part I: physiological parameters and model building. Clinical Pharmacokinetics. https://doi.org/10.1007/s40262-019-00825-6

Andersen, P.K., Geskus, R.B., de Witte, T., Putter, H., 2012. Competing risks in epidemiology: possibilities and pitfalls. International Journal of Epidemiology 41, 861–870. https://doi.org/10.1093/ije/dyr213

Andrade, T.S., Henriques, J.F., Almeida, A.R., Soares, A.M.V.M., Scholz, S., Domingues, I., 2017. Zebrafish embryo tolerance to environmental stress factors - Concentration-dose response analysis of oxygen limitation, pH, and UV-light irradiation. Environmental Toxicology and Chemistry 36, 682–690. https://doi.org/10.1002/etc.3579

Battistoni, M., Di Renzo, F., Menegola, E., Bois, F.Y., 2019. Quantitative AOP based teratogenicity prediction for mixtures of azole fungicides. Computational Toxicology 11, 72–81. https://doi.org/10.1016/j.comtox.2019.03.004

Bernillon, P., Bois, F.Y., 2000. Statistical issues in toxicokinetic modeling: a Bayesian perspective. Environmental Health Perspectives 108 (suppl. 5), 883–893.

Beyersmann, J., Allignol, A., Schumacher, M., 2012. Competing Risks and Multistate Models with R. Springer New York, New York, NY. https://doi.org/10.1007/978-1-4614-2035-4

Bois, F.Y., 2012. Bayesian inference, in: Reisfeld, B., Mayeno, A.N. (Eds.), Computational Toxicology Vol. II. Humana Press, New-York, pp. 597–636.

Bois, F.Y., 2009. GNU MCSim: Bayesian statistical inference for SBML-coded systems biology models. Bioinformatics 25, 1453–1454. https://doi.org/10.1093/bioinformatics/btp162

Bois, F.Y., Smith, T.J., Gelman, A., Chang, H.Y., Smith, A.E., 1999. Optimal design for a study of butadiene toxicokinetics in humans. Toxicological Sciences 49, 213–224.

Bradburn, M.J., Clark, T.G., Love, S.B., Altman, D.G., 2003a. Survival Analysis Part II: Multivariate data analysis – an introduction to concepts and methods. British Journal of Cancer 89, 431–436. https://doi.org/10.1038/sj.bjc.6601119

Bradburn, M.J., Clark, T.G., Love, S.B., Altman, D.G., 2003b. Survival Analysis Part III: Multivariate data analysis – choosing a model and assessing its adequacy and fit. British Journal of Cancer 89, 605–611. https://doi.org/10.1038/sj.bjc.6601120

Braunbeck, T., Boettcher, M., Hollert, H., Kosmehl, T., Lammer, E., Leist, E., Rudolf, M., Seitz, N., 2005. Towards an alternative for the acute fish LC(50) test in chemical assessment: the fish embryo toxicity test goes multi-species - an update. ALTEX 22, 87–102.

Braunbeck, T., Kais, B., Lammer, E., Otte, J., Schneider, K., Stengel, D., Strecker, R., 2015. The fish embryo test (FET): origin, applications, and future. Environmental Science and Pollution Research 22, 16247–16261. https://doi.org/10.1007/s11356-014-3814-7

Braunbeck, T., Lammer, E., 2006. Draft detailed review paper on fish embryo toxicity assays. UBA Contract Number 203 85 422. Umweltbundesamt – German Federal Environment Agency, Dessau, Germany.

Brinks, R., Hoyer, A., 2018. Illness-death model: statistical perspective and differential equations. Lifetime Data Anal 24, 743–754. https://doi.org/10.1007/s10985-018-9419-6

Brochot, C., Casas, M., Manzano-Salgado, C., Zeman, F.A., Schettgen, T., Vrijheid, M., Bois, F.Y., 2019. Prediction of maternal and foetal exposures to perfluoroalkyl compounds in a Spanish birth cohort using toxicokinetic modelling. Toxicology and Applied Pharmacology 379, 114640. https://doi.org/10.1016/j.taap.2019.114640

Brotzmann, K., Wolterbeek, A., Kroese, D., Braunbeck, T., 2020. Neurotoxic effects in zebrafish embryos by valproic acid and nine of its analogues: the fish-mouse connection? Archives of Toxicology. https://doi.org/10.1007/s00204-020-02928-7

Busquet, F., Strecker, R., Rawlings, J.M., Belanger, S.E., Braunbeck, T., Carr, G.J., Cenijn, P., Fochtman, P., Gourmelon, A., Hübler, N., Kleensang, A., Knöbel, M., Kussatz, C., Legler, J., Lillicrap, A., Martínez-Jerónimo, F., Polleichtner, C., Rzodeczko, H., Salinas, E., Schneider, K.E., Scholz, S., van den Brandhof, E.-J., van der Ven, L.T.M., Walter-Rohde, S., Weigt, S., Witters, H., Halder, M., 2014. OECD validation study to assess intra- and inter-laboratory reproducibility of the zebrafish embryo toxicity test for acute aquatic toxicity testing. Regulatory Toxicology and Pharmacology 69, 496–511. https://doi.org/10.1016/j.yrtph.2014.05.018

Cannon, J., Roberts, K., Milne, C., Carapetis, J.R., 2017. Rheumatic heart disease severity, progression and outcomes: a multi-state model. Journal of the American Heart Association 6. https://doi.org/10.1161/JAHA.116.003498

Chen, J., Lei, L., Tian, L., Hou, F., Roper, C., Ge, X., Zhao, Y., Chen, Y., Dong, Q., Tanguay, R.L., Huang, C., 2018. Developmental and behavioral alterations in zebrafish embryonically exposed to valproic acid (VPA): an aquatic model for autism. Neurotoxicology and Teratology 66, 8–16. https://doi.org/10.1016/j.ntt.2018.01.002

Christensen, J., Grønborg, T.K., Sørensen, M.J., Schendel, D., Parner, E.T., Pedersen, L.H., Vestergaard, M., 2013. Prenatal valproate exposure and risk of autism spectrum disorders and childhood autism. JAMA 309, 1696. https://doi.org/10.1001/jama.2013.2270

Chuang, C.-M., Chang, C.-H., Wang, H.-E., Chen, K.-C., Peng, C.-C., Hsieh, C.-L., Peng, R.Y., 2012. Valproic Acid Downregulates RBP4 and Elicits Hypervitaminosis A-Teratogenesis—A Kinetic Analysis on Retinol/Retinoic Acid Homeostatic System. PLoS ONE 7, e43692. https://doi.org/10.1371/journal.pone.0043692

Clark, T.G., Bradburn, M.J., Love, S.B., Altman, D.G., 2003a. Survival Analysis Part I: Basic concepts and first analyses. British Journal of Cancer 89, 232–238. https://doi.org/10.1038/sj.bjc.6601118

Clark, T.G., Bradburn, M.J., Love, S.B., Altman, D.G., 2003b. Survival Analysis Part IV: Further concepts and methods in survival analysis. British Journal of Cancer 89, 781–786. https://doi.org/10.1038/sj.bjc.6601117

d'Amora, M., Giordani, S., 2018. The utility of zebrafish as a model for screening developmental neurotoxicity. Frontiers in Neuroscience 12, 976. https://doi.org/10.3389/fnins.2018.00976

Driessen, M., Kienhuis, A.S., Pennings, J.L.A., Pronk, T.E., van de Brandhof, E.-J., Roodbergen, M., Spaink, H.P., van de Water, B., van der Ven, L.T.M., 2013. Exploring the zebrafish embryo as an alternative model for the evaluation of liver toxicity by histopathology and expression profiling. Archives of Toxicology 87, 807–823. https://doi.org/10.1007/s00204-013-1039-z

Dwivedi, S., Medishetti, R., Rani, R., Sevilimedu, A., Kulkarni, P., Yogeeswari, P., 2019. Larval zebrafish model for studying the effects of valproic acid on neurodevelopment: an approach towards modeling autism. Journal of Pharmacological and Toxicological Methods 95, 56–65. https://doi.org/10.1016/j.vascn.2018.11.006

Embry, M.R., Belanger, S.E., Braunbeck, T.A., Galay-Burgos, M., Halder, M., Hinton, D.E., Léonard, M.A., Lillicrap, A., Norberg-King, T., Whale, G., 2010. The fish embryo toxicity test as an animal alternative method in hazard and risk assessment and scientific research. Aquatic Toxicology 97, 79–87. https://doi.org/10.1016/j.aquatox.2009.12.008

EU, 2010. Directive 2010/63/EU of the European parliament and of the council of 22 September 2010 on the protection of animals used for scientific purposes. Official Journal of the European Union L 276, 33–79.

Farewell, V.T., Tom, B.D.M., 2014. The versatility of multi-state models for the analysis of longitudinal data with unobservable features. Lifetime Data Analysis 20, 51–75. https://doi.org/10.1007/s10985-012-9236-2

Fathe, K., Palacios, A., Finnell, R.H., 2014. Brief report novel mechanism for valproate-induced teratogenicity: Novel Mechanism for Valproate-Induced Teratogenicity. Birth Defects Research Part A: Clinical and Molecular Teratology 100, 592–597. https://doi.org/10.1002/bdra.23277

Fisz, M., 1976. Probability Theory and Mathematical Statistics, 3rd ed. Wiley, New York.

Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences (with discussion). Statistical Science 7, 457–511.

Gould, A.L., Boye, M.E., Crowther, M.J., Ibrahim, J.G., Quartey, G., Micallef, S., Bois, F.Y., 2015. Responses to discussants of 'Joint modeling of survival and longitudinal non-survival data: current methods and issues. report of the DIA Bayesian joint modeling working group.' Statistics in Medicine 34, 2202–2203. https://doi.org/10.1002/sim.6502

Hill, A.J., Teraoka, H., Heideman, W., Peterson, R.E., 2005. Zebrafish as a Model Vertebrate for Investigating Chemical Toxicity. Toxicological Sciences 86, 6–19. https://doi.org/10.1093/toxsci/kfi110

Hodgson, P., Ireland, J., Grunow, B., 2018. Fish, the better model in human heart research? Zebrafish Heart aggregates as a 3D spontaneously cardiomyogenic in vitro model system. Progress in Biophysics and Molecular Biology 138, 132–141. https://doi.org/10.1016/j.pbiomolbio.2018.04.009

Hollert, H., Keiter, S., König, N., Rudolf, M., Ulrich, M., Braunbeck, T., 2003. A new sediment contact assay to assess particle-bound pollutants using zebrafish (<i>Danio rerio<\i>) embryos. Journal of Soils and Sediments 3, 197–207. https://doi.org/10.1065/jss2003.09.085

ISO, 1996. International Organization for Standardization. Water quality - Determination of the 28 acute lethal toxicity of substances to a freshwater fish [Brachydanio rerio Hamilton-Buchanan 29 (Teleostei, Cyprinidae)]. ISO 7346-3: Flow-through method. Available: [http://www.iso.org].

Jackson, C.H., 2011. Multi-State Models for Panel Data: The **msm** Package for *R*. J. Stat. Soft. 38. https://doi.org/10.18637/jss.v038.i08

Jiang, H., Fine, J.P., 2007. Survival Analysis, in: Ambrosius, W.T. (Ed.), Topics in Biostatistics. Humana Press, Totowa, NJ, pp. 303–318. https://doi.org/10.1007/978-1-59745-530-5_15

Jones, E., Epstein, D., García-Mochón, L., 2017. A procedure for deriving formulas to convert transition rates to probabilities for multistate markov models. Medical Decision Making 37, 779–789. https://doi.org/10.1177/0272989X17696997

Kantae, V., Krekels, E.H.J., Ordas, A., González, O., van Wijk, R.C., Harms, A.C., Racz, P.I., van der Graaf, P.H., Spaink, H.P., Hankemeier, T., 2016. Pharmacokinetic modeling of paracetamol uptake and clearance in zebrafish larvae: expanding the allometric scale in vertebrates with five orders of magnitude. Zebrafish 13, 504–510. https://doi.org/10.1089/zeb.2016.1313

Kanungo, J., Cuevas, E., Ali, S., Paule, M., 2014. Zebrafish model in drug safety assessment. Current Pharmaceutical Design 20, 5416–5429. https://doi.org/10.2174/1381612820666140205145658

Karmen, C., Gietzelt, M., Knaup-Gregori, P., Ganzinger, M., 2019. Methods for a similarity measure for clinical attributes based on survival data analysis. BMC Medical Informatics and Decision Making 19, 195. https://doi.org/10.1186/s12911-019-0917-6

Kimmel, C.B., Ballard, W.W., Kimmel, S.R., Ullmann, B., Schilling, T.F., 1995. Stages of embryonic development of the zebrafish. Developmental Dynamics 203, 253–310. https://doi.org/10.1002/aja.1002030302

Kozma, C., 2001. Valproic acid embryopathy: report of two siblings with further expansion of the phenotypic abnormalities and a review of the literature. American Journal of Medical Genetics 98, 168–175.

Lammer, E., Carr, G.J., Wendler, K., Rawlings, J.M., Belanger, S.E., Braunbeck, Th., 2009. Is the fish embryo toxicity test (FET) with the zebrafish (Danio rerio) a potential alternative for the fish acute

toxicity test? Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology 149, 196–209. https://doi.org/10.1016/j.cbpc.2008.11.006

Larisch, W., Brown, T.N., Goss, K.-U., 2017. A toxicokinetic model for fish including multiphase sorption features: A high-detailed, physiologically based toxicokinetic model. Environmental Toxicology and Chemistry 36, 1538–1546. https://doi.org/10.1002/etc.3677

Lee, E.T., Go, O.T., 1997. Survival analysis in public health research. Annual Review of Public Health 18, 105–134. https://doi.org/10.1146/annurev.publhealth.18.1.105

MacRae, C.A., Peterson, R.T., 2015. Zebrafish as tools for drug discovery. Nature Reviews Drug Discovery 14, 721–731. https://doi.org/10.1038/nrd4627

Meira-Machado, L., de Uña-Álvarez, J., Cadarso-Suárez, C., Andersen, P.K., 2009. Multi-state models for the analysis of time-to-event data. Statistical Methods in Medical Research 18, 195–222. https://doi.org/10.1177/0962280208092301

Nagel, R., 2002. DarT: the embryo testa with the zebrafish <i>Danio rerio<\i> - A general model in ecotoxicology and toxicology. Alternatives to Laboratory Animals 19 Suppl 1, 38–48.

OECD, 2013. OECD Guidelines for the Testing of Chemicals. Section 2: Effects on Biotic Systems - Test No. 236: Fish Embryo Acute Toxicity (FET) Test. Organization for Economic Cooperation and Development, Paris, France. https://doi.org/10.1787/9789264203709-en

Ornoy, A., 2009. Valproic acid in pregnancy: how much are we endangering the embryo and fetus? Reproductive Toxicology 28, 1–10. https://doi.org/10.1016/j.reprotox.2009.02.014

Overstall, A.M., Woods, D.C., Parker, B.M., 2020. Bayesian optimal design for ordinary differential equation models with application in biological science. Journal of the American Statistical Association 115, 583–598. https://doi.org/10.1080/01621459.2019.1617154

Phiel, C.J., Zhang, F., Huang, E.Y., Guenther, M.G., Lazar, M.A., Klein, P.S., 2001. Histone Deacetylase Is a Direct Target of Valproic Acid, a Potent Anticonvulsant, Mood Stabilizer, and Teratogen. Journal of Biological Chemistry 276, 36734–36741. https://doi.org/10.1074/jbc.M101287200

Putter, H., Fiocco, M., Geskus, R.B., 2007. Tutorial in biostatistics: competing risks and multi-state models. Statistics in Medicine 26, 2389–2430. https://doi.org/10.1002/sim.2712

Quignot, N., Hamon, J., Bois, F.Y., 2014. Extrapolating <i>in vitro<\i> results to predict human toxicity, in: Bal-Price, A., Jennings, P. (Eds.), In Vitro Toxicology Systems, Methods in Pharmacology and Toxicology. Springer Science, New-York, pp. 531–550.

R Development Core Team, 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Russell, W.M.S., Burch, R.L., 1959. The principles of humane experimental technique. Methuen, London.

Siméon, S., Brotzmann, K., Fisher, C., Gardner, I., Silvester, S., Maclennan, R., Walker, P., Braunbeck, T., Bois, F.Y., 2020a. Development of a generic zebrafish embryo PBPK model and application to the developmental toxicity assessment of valproic acid analogs. Reproductive Toxicology 93, 219–229. https://doi.org/10.1016/j.reprotox.2020.02.010

Siméon, S., Brotzmann, K., Fisher, C., Gardner, I., Silvester, S., Maclennan, R., Walker, P., Braunbeck, T., Bois, F.Y., 2020b. Corrigendum to “Development of a generic zebrafish embryo PBPK model and application to the developmental toxicity assessment of valproic acid analogs” [Reprod. Toxicol. 93 (2020) 219-229]. Reproductive Toxicology S089062382030232X. https://doi.org/10.1016/j.reprotox.2020.10.006

Sipes, N.S., Padilla, S., Knudsen, T.B., 2011. Zebrafish as an integrative model for twenty-first century toxicity testing. Birth Defects Research Part C: Embryo Today: Reviews 93, 256–267. https://doi.org/10.1002/bdrc.20214

Smith, A.F.M., Roberts, G.O., 1993. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. Journal of the Royal Statistical Society Series B 55, 3–23.

Strähle, U., Scholz, S., Geisler, R., Greiner, P., Hollert, H., Rastegar, S., Schumacher, A., Selderslaghs, I., Weiss, C., Witters, H., Braunbeck, T., 2012. Zebrafish embryos as an alternative to animal experiments - A commentary on the definition of the onset of protected life stages in animal welfare regulations. Reproductive Toxicology 33, 128–132. https://doi.org/10.1016/j.reprotox.2011.06.121

Tsiros, P., Bois, F.Y., Dokoumetzidis, A., Tsiliki, G., Sarimveis, H., 2019. Population pharmacokinetic reanalysis of a Diazepam PBPK model: a comparison of Stan and GNU MCSim. Journal of Pharmacokinetics and Pharmacodynamics 46, 173–192. https://doi.org/10.1007/s10928-019-09630-x

van der Vaart, M., Spaink, H.P., Meijer, A.H., 2012. Pathogen Recognition and Activation of the Innate Immune Response in Zebrafish. Advances in Hematology 2012, 1–19. https://doi.org/10.1155/2012/159807

von Hellfeld, R., Brotzmann, K., Baumann, L., Strecker, R., Braunbeck, T., 2020. Adverse effects in the fish embryo acute toxicity (FET) test: a catalogue of unspecific morphological changes versus more specific effects in zebrafish (Danio rerio) embryos. Environmental Sciences Europe 32, 122. https://doi.org/10.1186/s12302-020-00398-3

Welton, N.J., Ades, A.E., 2005. Estimation of Markov chain transition probabilities and rates from fully and partially observed data: uncertainty propagation, evidence synthesis, and model calibration. Medical Decision Making 25, 633–645. https://doi.org/10.1177/0272989X05282637

Williams, J.P., Storlie, C.B., Therneau, T.M., Jr, C.R.J., Hannig, J., 2020. A Bayesian approach to multistate hidden Markov models: application to dementia progression. Journal of the American Statistical Association 115, 16–31. https://doi.org/10.1080/01621459.2019.1594831

Zare, A., Mahmoodi, M., Mohammad, K., Zeraati, H., Hosseini, M., Naieni, K.H., 2014. Assessing Markov and Time Homogeneity Assumptions in Multi-state Models: Application in Patients with Gastric Cancer Undergoing Surgery in the Iran Cancer Institute. Asian Pacific Journal of Cancer Prevention 15, 441–447. https://doi.org/10.7314/APJCP.2014.15.1.441

Zimmermann, F.F., Gaspary, K.V., Leite, C.E., De Paula Cognato, G., Bonan, C.D., 2015. Embryological exposure to valproic acid induces social interaction deficits in zebrafish (Danio rerio): a developmental behavior analysis. Neurotoxicology and Teratology 52, 36–41. https://doi.org/10.1016/j.ntt.2015.10.002

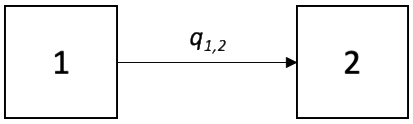

Figure 1. Schematic representation of a basic two-state survival model (state 1: alive; state 2: dead).

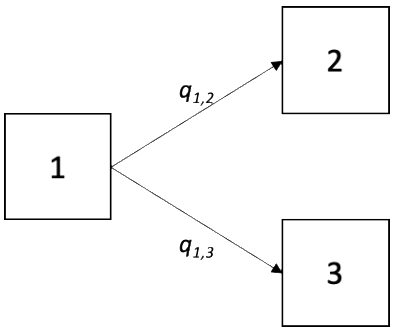

Figure 2. Schematic representation of a simple competing risk model. After a transition from states 1 (e.g., healthy) to 2 (e.g., dead), state 3 (e.g., sick) cannot be reached.

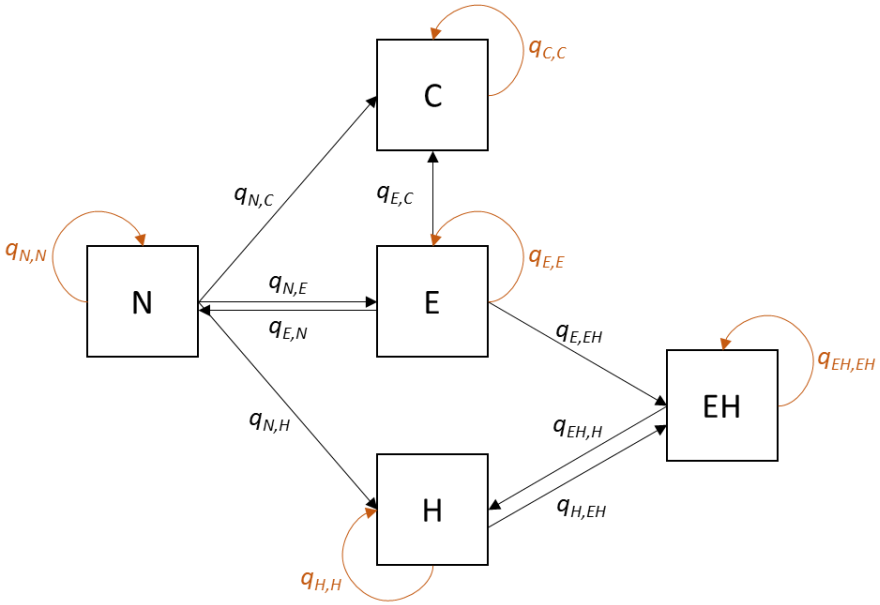

Figure 3. Schematic representation of our five-state competing risk model for the zebrafish embryo: "normal" (N), "hatched" (H), "non-hatched with effects" (E), "coagulated" (C), and "hatched with effects" (EH). The transition intensities from one state to itself (instantaneous probability of staying in a state) have been made explicit.

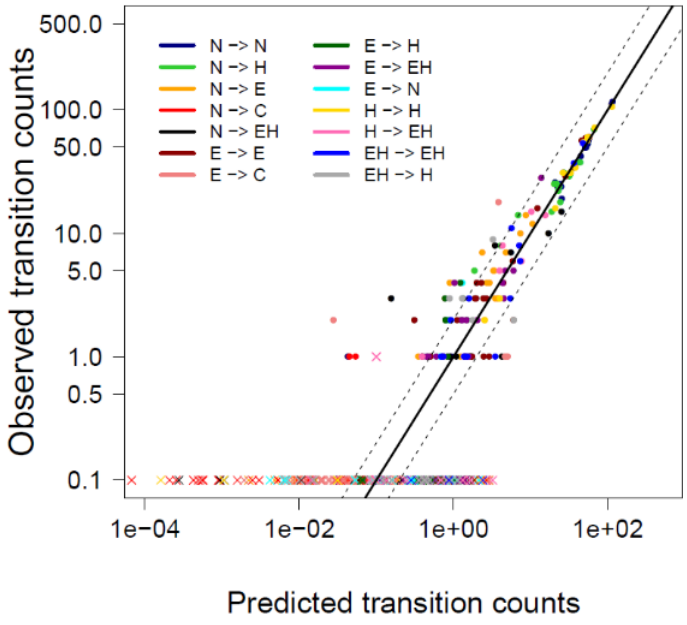

Figure 4. Observed versus best model-predicted transition counts of transitions between states normal (N), hatched (H), malformed (E), hatched with effects (EH) and dead (C). The black line is the perfect fit. The dashed lines give the two-fold-error band. The crosses are null observations or predictions assigned an arbitrary value of 0.1 to appear on the graph.

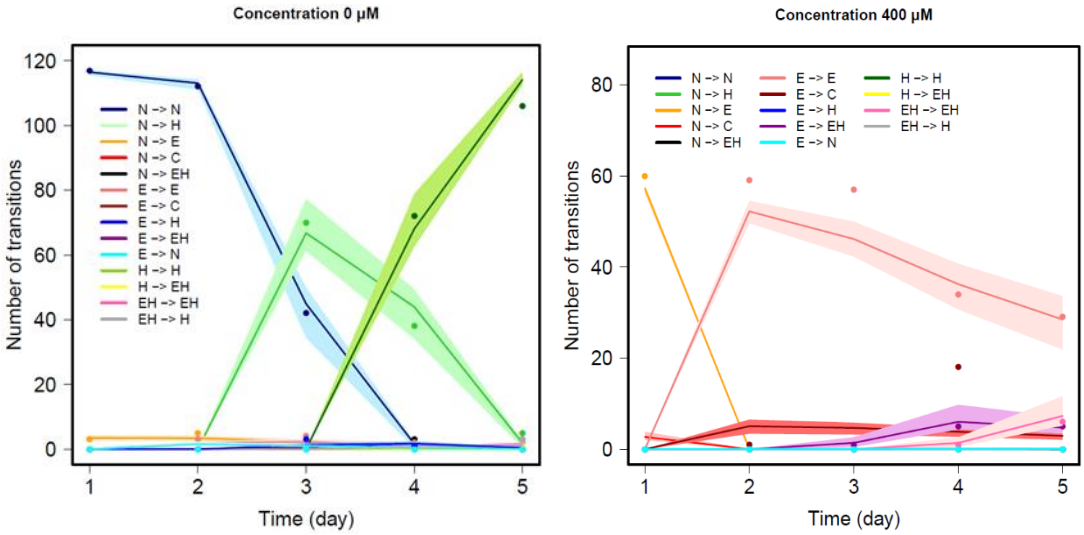

Figure 5. Observed (dots) and predicted (lines) number of transitions between states normal (N), hatched (H), malformed (E), hatched with effects (EH) and dead (C), as a function of time for the exposure concentrations 0 and 400 µM. Hatching rate is modeled by a piecewise constant intensity with estimated hatching time. The colored bands give the predictions' 95% confidence intervals.

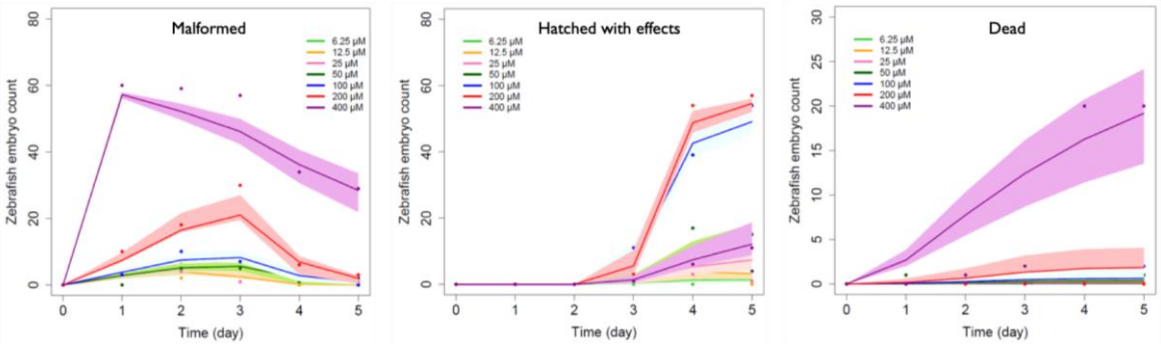

Figure 6. Counts of embryos in the malformed, hatched with effects and dead states, as a function of time and VPA concentrations. The dots are the data, the lines are model predictions. The colored bands give the predictions' 95% confidence intervals. The control group is omitted because very few embryos developed effects or died at concentration zero.

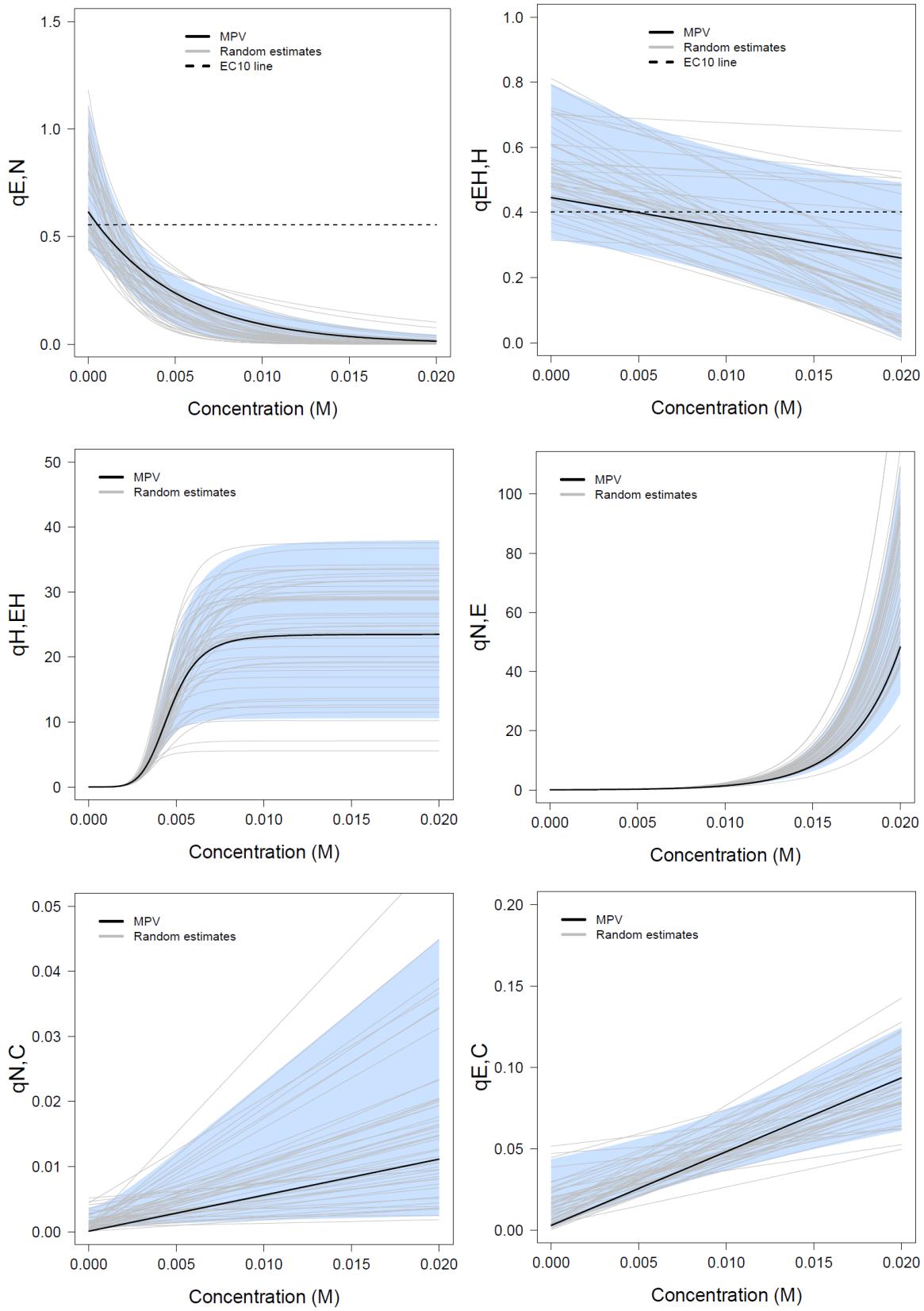

Figure 7. Transition rates as a function of VPA internal concentration. The black line is the best estimate. The grey lines are 50 random estimates. On the top panels, the dashed line corresponds to a 10% change from control (for the best estimate). The blue areas show the 95% confidence interval.

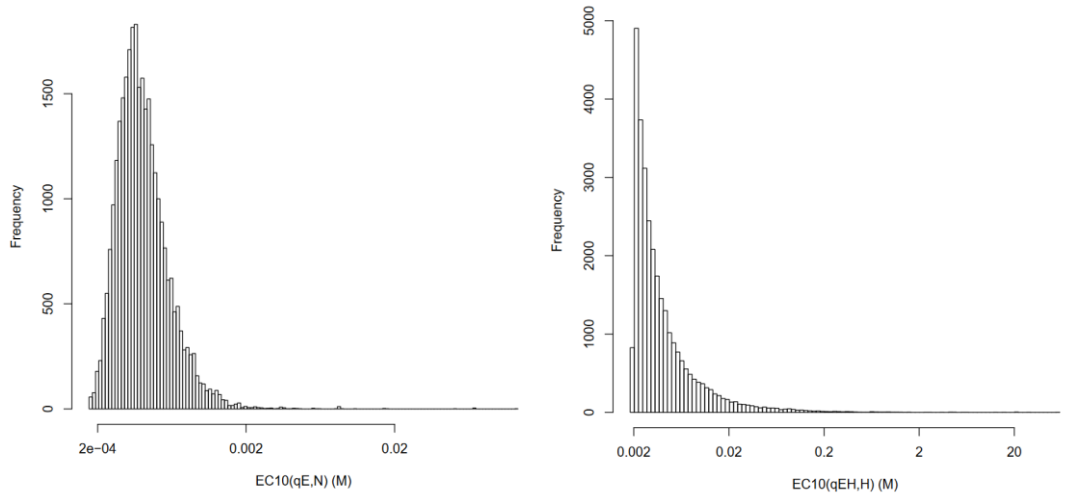

Figure 8. Distribution of EC10 values for the transition rates $q_{E,N}$ and $q_{EH,H}$. Those EC10 are defined on a scale of internal embryo concentrations.

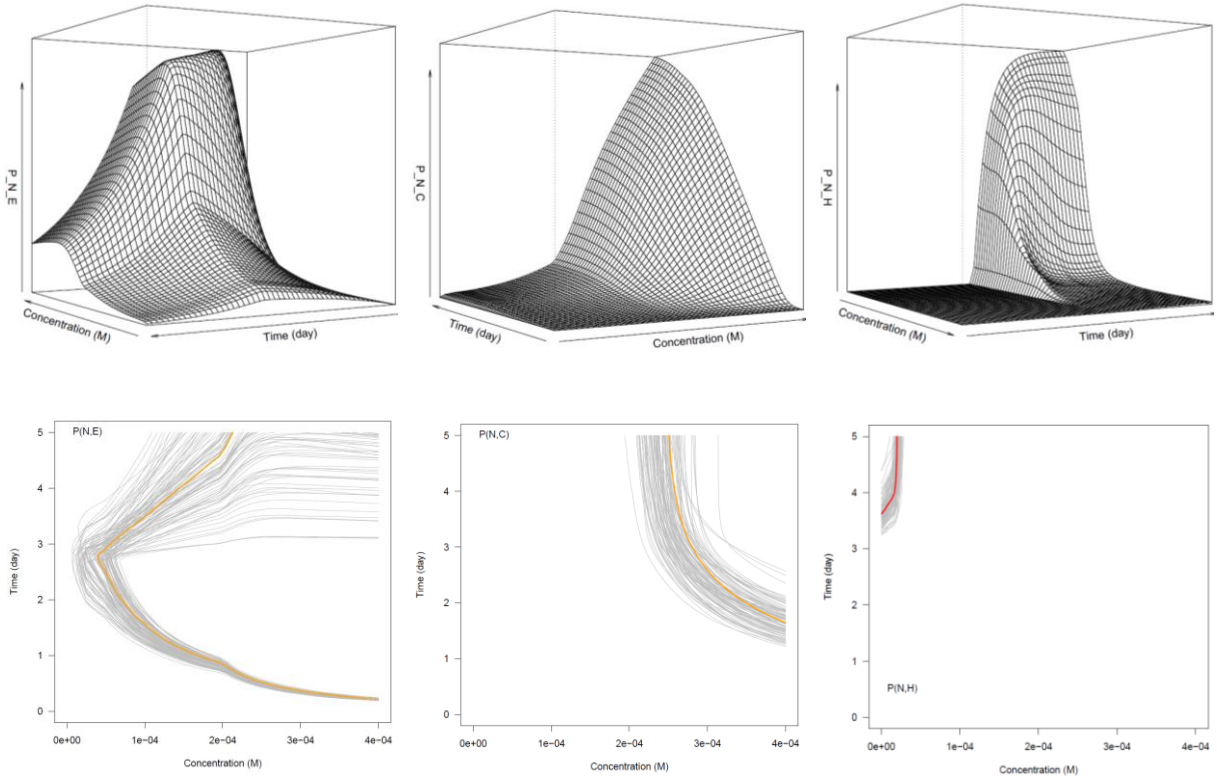

Figure 9. Top row: Probability for an embryo normal at time zero to be in state malformed ($P_{N,E}$, left), or to be dead ($P_{N,C}$, center), or to be in the healthy hatched state ($P_{N,H}$, right), as a function of time and VPA water concentration. The parameter best estimates were used to compute the surfaces. Bottom row: Contour lines (EC_{10}) at 10% probability for $P_{N,E}$ (left) and $P_{N,C}$ (center), and at 90% for $P_{N,H}$ (right) for the best parameter estimates and 100 random parameter sets.

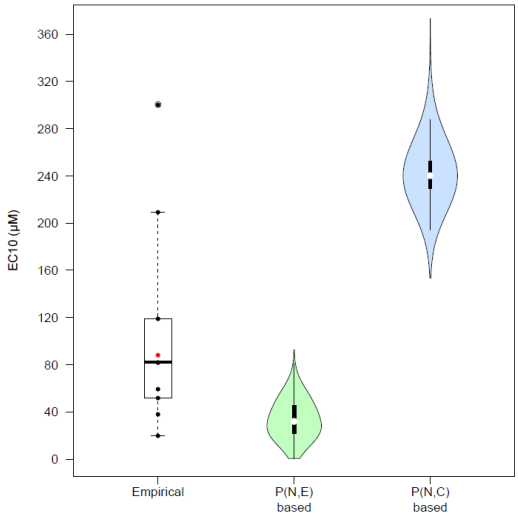

Figure 10. Distributions of the empirically determined external concentration EC_{10} values (the red dot corresponds to death) and of the minimal EC_{10} values obtained for $P_{N,E}$ (malformations) and $P_{N,C}$ (death) transition probabilities.

Table 1. Prior parameters distribution used for MCMC calibration of the five-state model with embryo transition count data.

Parameter	Prior distribution [a]
$q_{E,N}(0)$	$U(0, 2)$
$k_{qE,N}$	$U(-6\times10^2, 0)$
$q_{EH,H}(0)$	$U(0, 4)$
$k_{qEH,H}$	$U(-50\times q_{EH,H}(0), 0)$ [b]
$q_{N,E}(0)$	$U(0, 0.06)$
$k_{qN,E}$	$U(2\times10^2, 4.5\times10^2)$
$q_{N,C}(0)$	$U(0, 0.01)$
$k_{qN,C}$	$U(0, 10)$
$q_{E,C}(0)$	$U(0, 0.1)$
$k_{qE,C}$	$U(0, 10)$
$q_{H,EHmax}$	$U(0, 40)$
n	$U(0, 15)$
EC_{50}	$U(0, 0.01)$
$q_{N,H}$ [c]	$U(0, 20)$
$q_{E,EH}$ [c]	$U(0, 10)$
t_h [c]	$U(1, 5)$
$\sigma_{q_{N,H}}$	$N_t(1.1, 0.5, 1.01, 10)$
$\sigma_{q_{E,EH}}$	$N_t(1.1, 0.5, 1.01, 10)$
σ_{th}	$N_t(1.1, 0.5, 1.01, 10)$

[a] Distribution shapes: $U(a, b)$ for uniform between a and b; $N_t(m, s, a, b)$ for truncated normal with mean m, SD s, and bounds a and b.

[b] Correlated sampling.

[c] Prior distribution of the inter-group mean.

Table 2. Mean, standard deviation (SD), 95% confidence intervals (IC95%) and maximum posterior (MP) value of the MCMC-sampled multistate model parameters.

Parameter	Mean ± SD	IC 95%	MP
$q_{E,N}(0)$	0.75 ± 0.21	[0.40; 1.2]	0.61
$k_{qE,N}$	-280 ± 97.1	[-479; -103]	-190
$q_{EH,H}(0)$	0.53 ± 0.15	[0.29; 0.86]	0.45
$k_{qEH,H}$	-16.5 ± 0.182	[-35; -1.15]	-9.5
$q_{N,E}(0)$	0.042 ± 0.0059	[0.032; 0.055]	0.040
$k_{qN,E}$	371 ± 23	[318; 404]	355
$q_{N,C}(0)$	0.0014 ± 0.0012	[5.9×10^{-5}; 0.0043]	7.1×10^{-5}
$k_{qN,C}$	0.78 ± 0.70	[0.027; 2.7]	0.55
$q_{E,C}(0)$	0.017 ± 0.014	[0.00065; 0.050]	0.0028
$k_{qE,C}$	3.7 ± 1.2	[1.3; 6.1]	4.5
$q_{H,EHmax}$	24 ± 8.6	[9.2; 39]	23
n	5.8 ± 0.62	[4.6; 7.1]	5.4
EC_{50}	0.0044 ± 0.00056	[0.00345; 0.0056]	0.0046
$q_{N,H}$ [a]	3.8 ± 0.61	[2.9; 5.2]	3.3
$q_{E,EH}$ [a]	0.84 ± 0.29	[0.41; 1.5]	0.57
t_h [a]	2.8 ± 0.042	[2.7; 2.9]	2.8
$\sigma_{q_{N,H}}$	1.2 ± 0.19	[1.0; 1.7]	1.0
$\sigma_{q_{E,EH}}$	2.0 ± 0.26	[1.6; 2.6]	2.2
σ_{th}	1.0 ± 0.015	[1.0; 1.1]	1.0

[a] Inter-group mean.