Environ. Sci. Technol. **2007**, *41,* 5888−5894

Estimation of Chemical Toxicity to Wildlife Species Using Interspecies Correlation Models

S. RAIMONDO,*,† P. MINEAU,‡ AND M. G. BARRON†

U.S. Environmental Protection Agency, National Health and Environmental Effects Laboratory, Gulf Ecology Division, 1 Sabine Island Drive, Gulf Breeze, Florida 32561, and National Wildlife Research Centre, Science and Technology Branch, Environment Canada, 1125 Colonel By Drive, Ottawa K1S 5B6, Canada

Ecological risks to wildlife are typically assessed using toxicity data for relatively few species and with limited understanding of differences in species sensitivity to contaminants. Empirical interspecies correlation models were derived from LD50 values for 49 wildlife species and 951 chemicals. The standard wildlife test species Japanese quail (*Coturnix japonica*) and mallard (*Anas platyrhynchos*) were determined to be good surrogates for many species within the database. Cross-validation of all models predicted toxicity values within 5-fold and 10-fold of the actual values with 85 and 95% certainty, respectively. Model robustness was not consistently improved by developing correlation models within modes of action (MOA); however, improved models for neurotoxicants, carbamates, and direct acting organophosphorous acetylcholenesterase inhibiting compounds indicate that toxicity estimates may improve if MOA-specific models are built with robust datasets. There was a strong relationship between taxonomic distance and cross-validation prediction success ($\chi^2 = 297$, df = 12, $p < 0.0001$), with uncertainty increasing with larger taxonomic distance between the surrogate and predicted species. Interspecies toxicity correlations provide a tool for estimating contaminant sensitivity with known levels of uncertainty for a diversity of wildlife species.

Introduction

Extrapolation between species and toxicity endpoints has remained one of the key areas of uncertainty in both ecological and human health risk assessment. For wildlife, the assessment of risks to populations of birds and mammals has typically been based on relatively limited toxicity data for only a few standard test species. A number of approaches and models have been used to extrapolate toxicity data between wildlife species, including uncertainty or safety factors, the derivation of species sensitivity distributions (*1*) with or without body size scaling (*2*), and interspecies correlation estimation (ICE) models (*3*). However, most of the strategies involving the application of safety factors can be criticized for introducing a large source of unquantifiable uncertainty in risk assessments (*4*). Strategies involving species sensitivity distributions typically require well-populated datasets and devolve, when data are limiting, to either a safety factor approach that is empirically based (*1*) or the use of pooled variances established for a large group of pesticides (*5*). The number of species and families required in species sensitivity distributions vary with both scientific and regulatory applications.

ICE models are log-linear least-square regression models that describe the relationship between the acute toxicity (LD50; mg/kg bodyweight) of a range of chemicals tested in two species. If toxicity data are available for the surrogate species, the toxicity to the predicted taxon can be estimated using the specific ICE model for that species-taxon pair. ICE models have been employed in toxicity extrapolation of aquatic invertebrates and fish (*6*), but their use in wildlife risk assessment has not been widely accepted. Earlier development of ICE models did not show good correlations between wildlife species (*3*), in part due to limited databases used in model development. Additionally, acceptance of interspecies regression models has been limited due to lack of model validation studies (*7*). While sublethal endpoints maybe more ecologically relevant, the development of ICE models for chronic toxicity is problematic because of substantially fewer toxicity values compared to LD50 databases for both aquatic and wildlife species.

This study developed a comprehensive set of ICE models for the acute toxicity to wildlife species of birds and mammals and identified key sources of uncertainty in wildlife toxicity extrapolation. ICE models were developed for 49 species from a comprehensive database compiled from open literature (*8*−*12*) and by governmental agencies of the United States (U.S. EPA Office of Pesticide Programs) and Canada (Environment Canada) (*1*, *13*). The suite of models were cross-validated and analyzed to determined if (i) ICE models with significant regressions can be used to predict acute toxicity, (ii) there are certain chemical mode of action (MOA) that should not be included in ICE models, (iii) ICE models improve when developed with data of specific chemical MOAs, (iv) ICE model robustness is related to the taxonomic distance of the predicted and surrogate species, and (v) the standard test species, mallard, northern bobwhite (*Colinus virginianus*), and Japanese quail, are adequate surrogates for predicting toxicity to other wildlife species. Results of these analyses provide user guidance for use and application of ICE models in wildlife risk assessment.

Materials and Methods

Data Collection. The wildlife dataset was comprised of 4329 single oral dose LD50 values for 156 species and 951 chemicals. The data were collected from the open literature (*8*−*12*) and from datasets compiled by governmental agencies of the United States (U.S. EPA; Office of Pesticide Programs Ecotoxicity Database) and Canada (Environment Canada; *1*, *13*). Approximately 40% of the values included in the dataset were not used in the analyses because (1) the species did not have toxicity values for three or more chemicals or (2) chemicals were not present in two or more species. Of the 2454 toxicity records used in used in the final analyses, approximately 60% were pesticides, including 41% that were carbamate or OP acetylcholinesterase inhibitors (Supporting Information 1).

Data were subjected to rigorous quality assurance guidelines and standardized by using only data for adult animals and data for chemicals of technical grade or formulations with >90% active ingredient. Open-ended toxicity values (i.e., >100 mg/kg or <100 mg/kg) and duplicate records among

* Corresponding author phone: (850) 934-2424; fax (850) 934-2402; e-mail: raimondo.sandy@epa.gov.
† U.S. Environmental Protection Agency.
‡ Environment Canada.

10.1021/es070359o CCC: $37.00 Published 2007 by the Am. Chem. Soc.
Published on Web 07/13/2007

multiple sources were not included in model development. When data were reported as a range (ie. 100−200 kg/mg; ref *8*) or data were collected from multiple sources for a species and chemical, the geometric mean of the values was used. In cases where the range of minimum and maximum values for a chemical and species were greater than 10-fold, all data records for that chemical were removed for that species due to their high variability.

Model Development and Validation. Models were developed using Model II least-squares regression in which variables are independent and subject to measurement error (*3*). An algorithm was written in S-plus (*14*) to pair every species with every other species by common chemical. Three or more common chemicals per pair were required for inclusion in the analysis. For each species pair, a linear model was used to calculate the regression equation Log_{10}(predicted taxa) $= a + b \times \text{Log}_{10}$(surrogate species), where a and b are the intercept and slope of the line, respectively. Only models that had a significant linear relationship (p-value < 0.05) were used for further analyses, because models that are not significant at the $p < 0.05$ level have an elevated probability of performing a Type I error.

The uncertainty of each model was assessed using leave-one-out cross-validation. In this method, each pair of LD50 values for surrogate and predicted species were systematically removed from the original model. The remaining data were used to rebuild a model and estimate the toxicity value of the removed predicted species from the respective surrogate species toxicity value. This method could only be used for models with degrees of freedom equal or greater than 2 ($N \geq 4$). To maintain uniformity among the large number and diversity of models, the N-fold difference among each estimated and actual value (nontransformed data) was calculated and used to determine the fit of the estimated toxicity value. The wildlife database was used to estimate interlaboratory variation for wildlife species from the range of multiple toxicity values for a single species and chemical. The average range of multiple toxicity measurements for a specific chemical and species was 4.0 ± 10.5 ($N = 814$) when multiple values from same source (i.e., ref *11*) were included in the analysis and 6.4 ± 15.8 ($N = 286$) when only data from different sources were compared. Thus, a 5-fold difference was deemed a good fit in the validation analysis. Each removed data point was assigned to a prediction category based on the N-fold difference of actual and predicted values. The categories were 5-fold (≤5-fold), 10-fold (>5-fold, ≤10-fold), 50-fold (>10-fold, ≤50-fold), and greater than 50-fold.

The cross-validation success rate was calculated for each model as the proportion of removed data points that were predicted within 5-fold of the actual value from models that were statistically significant. In cases where the removal of an xy data pair resulted in the development of a model that was no longer significant at the $p < 0.05$ level, these replicates were not included in the cross-validation success rate and these models were not included in further analyses. This was the case for 52 models with low degrees of freedom (≤6) and a p-value between 0.01 and 0.05 in the original model.

Determination of Sources of Model Uncertainty. To assist in the selection of robust models for toxicity estimation in risk assessment, model R^2 and mean square error (MSE) were related to cross-validation success rate. Model MSE is the unbiased sum of squared deviations of the prediction line (*15*), and is an estimate of the average model prediction error. A simple linear regression was conducted to compare cross-validation success rate with MSE for all models in which cross-validation success rate could be calculated ($N = 484$). Model R^2 is the relative prediction power of the model and describes the proportion of variation in the data explained by the model. Similarly to the MSE analysis, a linear regression was conducted to compare cross-validation success rate with model R^2 values.

The relationship between model uncertainty and chemical mode of action (MOA) was explored in two analyses. The first analysis was conducted to determine if any chemical MOA consistently failed to be predicted by ICE models. The second analysis determined if ICE models were improved by using only data from one MOA. For both analyses, each chemical in the database was assigned to a broad MOA (ie. acetylcholinesterase (AChE) inhibitors) and specific MOA (e.g., carbamate mediated AChE inhibition). Mode of action assignments were made based on an assessment of chemical structure, including major moieties and functional groups, the mechanism of acute toxicity, therapeutic category, and pesticidal activity through a review of reports and peer reviewed articles (e.g., *16*−*21*), and Internet sources (e.g., http://www.ncbi.nlm.nih.gov). Chemicals were grouped into 12 broad MOA and 26 specific MOA categories (Supporting Information 1). Organophosphorous acetylcholinesterase inhibitors (OAIs; e.g., organophosphates, phosphorothioates, phosphoramidates) were also subdivided into direct acting (containing an oxon moiety) and indirect acting chemicals requiring metabolic activation to the oxon.

For the first MOA analysis, broad and specific MOAs (and OAI subgroups where applicable) were assigned to each data point predicted in the cross-validation analysis. For every model, each removed y data point was associated with a prediction success rate (5-fold, 10-fold, etc.) and chemical identification. All predicted data points from all models were pooled (total $N = 11846$). For every broad and specific MOA category and OAI subgroup, a binomial test was conducted to determined if the MOA failed to be predicted within 5-fold of the actual value for ≥50% of the time it was used in a model. This analysis tested the null hypothesis that the linear models could not be used to predict toxicity for each MOA. The test statistic was the number of observations of an MOA within the 5-fold category of prediction, the sample size N was the total number of times the MOA was predicted in all models and the probability of failure (p) was 0.5, corresponding to the hypothesized rate of failure ≥50% of the time it was used in a model (*22*).

The second chemical MOA analysis was conducted to determine if models improved if built from only one MOA or subgroup. For all models with df ≥50 ($N = 64$, 32 species pairs), individual models were built for the most abundant broad MOAs (AChE inhibition, neurotoxicity, and anesthesia) and specific MOAs (carbamates and OAIs), and the OAI subgroups. Models built with data from a specific MOA or subgroup required a minimum of five data points to be included in the analysis to avoid including models with extremely small sample sizes that may not adequately describe the species sensitivity relationship. Values of R^2 and MSE of models comprised of all data were compared to models built from data within an MOA or subgroup using a two-tailed t test. Differences in models developed from single MOA or subgroup compared to models developed with all data were also explored qualitatively using the northern bobwhite−mallard model. MOA-specific models were built with northern bobwhite and mallard as the surrogate and predicted species, respectively, for AChE inhibition, carbamates, OAIs, direct acting OAIs, and indirect acting OAIs.

Uncertainty of models developed with all data was compared to taxonomic distance by assigning each model a distance value based on the taxonomy of the surrogate and predicted species. Models were rated such that surrogate and predicted species within the same genus = 1; family = 2; order = 3; class = 4; phylum = 5 (i.e., birds vs mammals). For all the models within each taxonomic distance category, the number of observations in each cross-validation category (5-fold, 10-fold, etc.) was determined. The observations were

TABLE 1. Species Used in ICE Models with the Total Number of Data Points and Statistically Significant Models ($p < 0.05$) for Each Species

common name	scientific name	no. of datapoints	no. of models
birds			
American crow	*Corvus brachyrhynchos*	13	12
American kestrel	*Falco sparverius*	15	6
American robin	*Turdus migratorius*	9	6
black-billed magpie	*Pica hudsonia*	14	6
boat-tailed grackle	*Quiscalus major*	10	1
brown-headed cowbird	*Molothrus ater*	15	11
budgerigar	*Melopsittacus undulatus*	9	2
california quail	*Callipepla californica*	46	15
canada goose	*Branta canadensis*	26	10
chicken	*Gallus gallus*	142	19
chukar	*Alectoris chukar*	40	21
common grackle	*Quiscalus quiscula*	63	22
eastern screech owl	*Megascops asio*	5	1
fulvous whistling duck	*Dendrocygna bicolor*	12	6
golden eagle	*Aquila chrysaetos*	15	1
golden sparrow	*Passer luteus*	5	1
golden-crowned sparrow	*Zonotrichia atricapilla*	3	1
gray partridge	*Perdix perdix*	55	13
house finch	*Carpodacus mexicanus*	54	19
house sparrow	*Passer domesticus*	174	23
Japanese quail	*Coturnix japonica*	503	28
mallard	*Anas platyrhynchos*	419	31
mourning dove	*Zenaida macroura*	29	10
northern bobwhite	*Colinus virginianus*	372	20
peking duck	*Anas domestica*	8	1
red billed quelea	*Quelea quelea*	48	22
red billed quelea	*Quelea quelea*	48	22
red-legged partridge	*Alectoris rufa*	52	12
red-winged blackbird	*Agelaius phoeniceus*	515	34
ringed turtledove	*Streptopelia risoria*	6	5
ring-necked pheasant	*Phasianus colchicus*	208	28
rock dove	*Columba livia*	142	30
sharp-tailed grouse	*Tympanuchus phasianellus*	29	17
starling	*Sturnus vulgaris*	301	21
tricolored blackbird	*Agelaius tricolor*	6	1
turkey	*Meleagris gallopavo*	13	7
white-crowned sparrow	*Zonotrichia leucophrys*	5	5
white-winged dove	*Zenaida asiatica*	10	3
yellow headed blackbird	*Xanthocephalus xanthocephalus*	17	12
mammals			
blacktailed jackrabbit	*Lepus californicus*	10	3
deer mouse	*Peromyscus maniculatus*	49	4
dog	*Canis familiariz*	17	4
guinea pig	*Cavia porcellus*	18	2
mjeadow vole	*Microtus pennsylvanicus*	8	1
mouse	*Mus musculus*	136	19
mule deer	*Odocoileus hemionus*	50	9
Norway rat	*Rattus norvegicus*	337	26
rabbit	*Oryctolagus cuniculus*	21	7
ricefield rat	*Rattus argentiventer*	8	1
roof rat	*Rattus rattus*	11	1

compared among taxonomic distance categories and cross-validation category using a χ^2 test for differences in probabilities (*22*).

Results

There were 560 significant models ($p < 0.05$) developed for 49 species from the complete dataset (Table 1). The avian surrogate species with the most models were red-winged blackbird (*Agelaius phoeniceus,* 34), mallard (31), rock dove (*Columba livia,* 30), Japanese quail (28), and ring-neck pheasant (*Phasianus colchicus,* 28). The mammal species with the most models were Norway rat (*Rattus norvegicus*, 26) and the mouse (*Mus musculus,* 19). Information on all models, including statistics and parameters necessary for toxicity estimation (slope, intercept, degrees of freedom, number of chemicals in model, mean square error, sum of squares) are provided in the Supporting Information 2.

Cross-validation was conducted for 536 models, 52 of which resulted in the development of models that were no longer significant at the $p < 0.05$. Thus, 484 models were successfully cross-validated. Cross-validation of all models showed that, overall, ICE models for wildlife species predicted within 5-fold of the actual value for 85% of all removed data points and within 10-fold for 95% of removed points ($N = 11846$). Model MSE was significantly related to cross-validation success rate, described by the relationship $y = 101 - 73.5x$ ($R^2 = 0.64$, df = 482, $p < 0.0001$; Figure 1a). Based on this relationship, MSE of 0.22 and 0.15 corresponded to cross-validation success rate of 85 and 90%, respectively. Model R^2 was also related to cross-validation success rate and is described by the relationship $y = 64.2 + 34.6x$ ($R^2 = 0.27$, df = 482, $p < 0.0001$; Figure 1b). Based on this relationship, model R^2 of 0.60 and 0.75 corresponded to cross-validation success rate of 85 and 90%, respectively. Model MSE accounted for more variation in cross-validation success rate ($R^2 = 0.64$) than the relationship described in the R^2 analyses ($R^2 = 0.27$), and should be used over model R^2 when selecting between two models.

The first MOA analysis determined that there was not an MOA category that failed to be predicted within 5-fold of the actual values more than 50% of all times the MOA was predicted by the combined set of models. On average, data points from all MOA categories were predicted within 5-fold of the actual values for 85% of all occurrences.

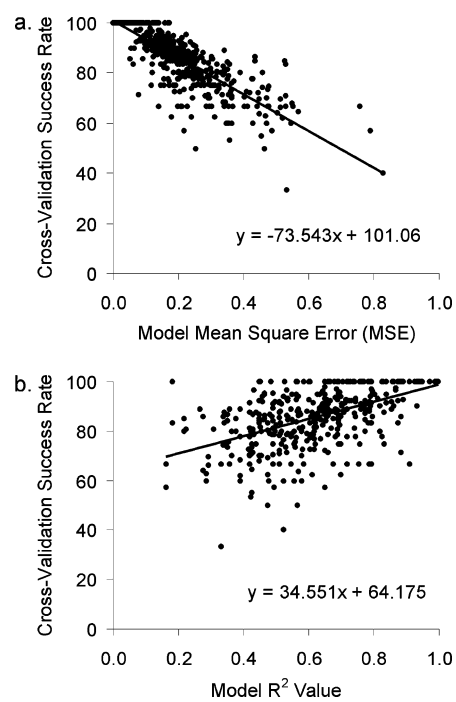

FIGURE 1. Relationship of cross-validation success to (a) model mean square error (MSE) and (b) model R^2.

In the second MOA analysis, models built with data from anesthesia, AChE inhibition, OAIs, and indirect acting OAIs MOAs were not improved over models built with all data. For these models, values of R^2 were significantly lower and MSE was not significantly different from models built with all data. Models built with neurotoxicants, carbamates, and direct acting OAIs that were significant at the $p = 0.05$ level

TABLE 2. Mean Square Error (MSE) and R^2 (±std) Compared for Models Built with Data from All MOAs with Those Built with Data from Only One MOA/chemical Class or Subgroup. Only Models That Were Significant ($p < 0.05$) When Built with One Mode of Action/Chemical Class or Subgroup Were Included in the Analyses.

mode of action/chemical class	models developed	significant models	model parameter	all data models	MOA specific models	df	t	p-value
anesthesia	64	50	R^2	0.70 (0.11)	0.62 (0.18)	98	2.54	0.01
			MSE	0.23 (0.10)	0.24 (0.15)	98	−0.55	0.58
neurotoxicity	38	32	R^2	0.71 (0.11)	0.85 (0.10)	62	−5.18	< 0.01
			MSE	0.21 (0.06)	0.15 (0.15)	62	2.11	0.04
AChE[a] inhibition	64	64	R^2	0.64 (0.15)	0.56 (0.15)	126	3.30	< 0.01
			MSE	0.25 (0.11)	0.22 (0.11)	126	1.50	0.14
carbamates	58	46	R^2	0.65 (0.14)	0.65 (0.15)	90	−0.24	0.81
			MSE	0.24 (0.10)	0.18 (0.11)	90	2.60	0.01
organophosphorous[b] compounds	64	64	R^2	0.64 (0.15)	0.57 (0.17)	126	2.82	< 0.01
			MSE	0.25 (0.11)	0.21 (0.12)	126	1.66	0.10
direct acting[c]	62	52	R^2	0.64 (0.15)	0.70 (0.15)	102	−2.24	0.03
			MSE	0.25 (0.12)	0.14 (0.09)	102	5.31	<0.01
indirect acting[d]	64	54	R^2	0.66 (0.13)	0.58 (0.19)	106	2.77	< 0.01
			MSE	0.24 (0.11)	0.23 (0.15)	106	0.30	0.76

[a] AChE is acetylcholinesterase. [b] Organophosphorous-type acetylcholinesterase inhibitors, including organophosphates, phosphorothioates, phosphonates, phosphoroamidates, etc. [c] Direct inhibition of AChE. [d] Indirect acting requires metabolic transformation to oxon form for AChE inhibition.

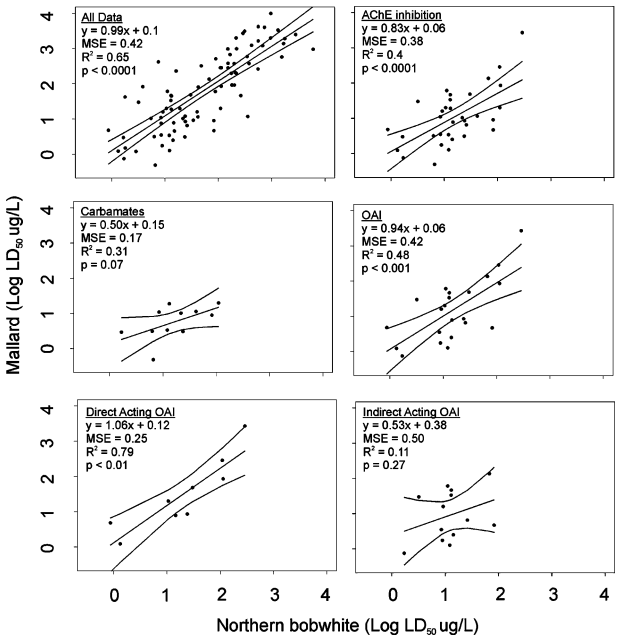

FIGURE 2. MOA/chemical class-specific models for northern bobwhite and mallard.

were improved over models built with all data based on significantly lower MSE (Table 2). For all MOA or subgroups except AChE inhibition and OAIs, building models with MOA-specific data resulted in the loss of a significant relationship for some models. For example, based on the criteria that MOA-specific models contain $N \geq 5$, there were 58 carbamate models developed from a possible 64 models. Of those 58 carbamate models, 46 were significant at the $p = 0.05$ level (Table 2).

These results are supported by the qualitative analysis of the northern bobwhite/mallard model built for all data compared to AChE inhibition, carbamates, OAIs, direct acting OAIs, and indirect acting OAIs models (Figure 2). The models built with carbamates and direct acting OAIs had an MSE 2.5 and 1.7 times smaller than the model built with all data, respectively; however, the carbamate model was not significant at the $p = 0.05$ level. The MSE for models built with AChE inhibition, OAI, and indirect acting OAIs data were not markedly different than the MSE for the model built with all data and all of these models showed a decrease in R^2 relative to the all data model. Additionally, the Northern bobwhite/mallard model was not significant when built solely with indirect acting OAIs (Figure 2).

TABLE 3. Number of Observations within Each Prediction Category for Models Grouped by Closest Shared Taxon[a]

shared taxonomic level	total observations	cross-validation prediction category: 5-fold	10-fold	50-fold	>50-fold
genus	48	48 (100)	0 (0)	0 (0)	0 (0)
family	1452	1332 (92)	89 (6)	31 (2)	0 (0)
order	2238	2014 (90)	154 (7)	63 (3)	7 (0.3)
class	5706	4861 (85)	564 (10)	267 (5)	14 (0.2)
phylum	2402	1836 (76)	322 (13)	208 (9)	36 (1.5)

[a] Numbers in parentheses are the percentage of observations in each prediction category for each shared taxonomic level.

There was a strong relationship between taxonomic distance and cross-validation prediction category ($\chi^2 = 297$, df = 12, $p < 0.0001$). There was a decrease in the percentage of data points in the 5-fold prediction category with increasing taxonomic distance; models built for species within the same genus predicted within 5-fold of the actual value for 100% of all data points, where models built for two species within the same phylum (bird vs mammal) predicted within 5-fold of the actual value for 76% of all data points. There was an increase in the percentage of data points in the other prediction categories (10-fold, 50-fold, >50-fold) with increasing taxonomic distance (Table 3).

Based on the results of model uncertainty analyses, surrogacy was evaluated for the avian and mammal species for which the most models were developed (Table 4). Rock dove, Japanese quail, and red-winged blackbird all had greater than 20 models with MSE ≤ 0.22 or $R^2 \geq 0.6$, indicating that these three species have the potential to provide good surrogacy for acute toxicity estimation. Mallard, ring-necked pheasant, and house sparrow (*Passer domesticus*) are potentially good surrogates for more than 15 models based on low MSE or high R^2. Northern bobwhite, common grackle (*Quiscalus quiscula*), chukar (*Alectoris chukar*), and red-billed quelea (*Quelea quelea*) provided good surrogacy for over 10 species each based on low MSE. For mammal species, both the Norway rat and mouse had 10 models with MSE ≤ 0.22 or $R^2 \geq 0.6$ (Table 4).

Discussion

Interspecies correlation estimations for wildlife species were significant for 560 models built for 49 species. Model MSE

TABLE 4. Summary of the Surrogate for Which the Most ICE Models Were Developed and the Number of Models of Each Surrogate Fitting Each Model Parameter (i.e., MSE, R^2, etc.) Criteria[a]

surrogate species	*N*	average MSE	MSE ≤ 0.22	$R^2 \geq 0.6$	cross-validation success rate ≥ 85%	taxonomic distance ≤ 3
chukar	21 (19)	0.22	11	9	9	8
common grackle	22 (22)	0.18	16	6	18	9
house sparrow	23 (20)	0.16	16	18	16	8
japanese quail	28 (26)	0.21	20	22	18	9
mallard	31 (28)	0.23	17	17	13	2
northern bobwhite	20 (20)	0.26	13	10	13	8
red-billed quelea	22 (18)	0.14	17	14	12	9
red-winged blackbird	34 (31)	0.20	21	20	17	12
ring-necked pheasant	28 (25)	0.26	17	16	13	9
rock dove	30 (26)	0.19	23	21	19	2
starling	21 (19)	0.26	7	6	8	8
mouse	19 (19)	0.22	10	11	10	3
Norway rat	26 (26)	0.29	10	11	14	6

[a] The total number of models developed is indicated as *N*, with the number of models for which cross-validation success rate could be calculated in parenthesis.

and R^2 are linearly related to model robustness and should be used to guide risk assessors in selection of appropriate models for toxicity estimation. Based on the strength of their linear relationships, model MSE explained more variation in cross-validation success rate than R^2 and should be used as the primary criterion in model selection. High cross-validation success rate (≥90%) for models built for species within the same genus, family, and order demonstrate that these models can be used to estimate toxicity of an unknown chemical with limited and known uncertainty. These results provide guidance on selecting ICE models for risk assessment; ICE models are generally most robust when they possess close taxonomic distance (within order), small model MSE (approximately 0.22 or less), and large model R^2 value (approximately 0.6 or greater), and high cross-validation success rate (approximately 85% or greater).

Interspecies extrapolation models use surrogate species, often standard test organisms, to estimate toxicity of a chemical to wildlife species for which no toxicity data exists. In North America, tests are usually conducted on mallards and northern bobwhite, whereas European countries often use Japanese quail as their standard test species (*13*). As a surrogate species, the number of significant models was 31 for mallard, 20 for northern bobwhite, and 28 for Japanese quail. Of the models developed for these species, a relatively large proportion of them possessed combinations of low MSE, high R^2, and/or high cross-validation success rate, demonstrating their potential to be used as avian surrogates for estimating toxicity to wildlife species with limited or no toxicity data for a specific chemical. Several other species, including red-winged blackbird, rock dove, and ring-necked pheasant, were surrogate species for more than 20 models and offer potential to be used as surrogates for toxicity extrapolation (Table 4). Because model uncertainty increases with increasing taxonomic distance, estimating toxicity to mammals from surrogate bird species is not recommended. The Norway rat and house mouse yielded 9 and 5 robust ICE models, respectively, for mammal species and are recommended for estimating toxicity to mammals in wildlife risk assessment. Fewer ICE models for mammal species is a reflection of a bias toward birds within the dataset and does not necessarily reflect poor surrogacy of mammal-based models. As new LD50 values for mammals become available, ICE models can be updated to include new chemicals and species in the suite of ICE wildlife models. Future research should also evaluate other species as surrogates as additional test data become available.

Model uncertainty analyses using leave-one-out cross-validation provides only an estimate of generalization error and is not a true validation of model fitness. Particularly for models built from small datasets, a small change in the data can cause a large change in the model (*23*). Although bootstrap validation or leave-v-out cross-validation may provide more robust estimates of error for models built with larger datasets (*24*), leave-one-out cross-validation was used here to maintain uniformity among all 484 models validated. Additionally, because the model fitness criterion (5-fold) is based on interlaboratory/inter-test variation of acute toxicity tests rather than model prediction error, the leave-one-out method provided a vehicle for comparing estimated toxicity values among a diversity of different models.

We developed MOA-specific models to determine if there would be a significant reduction in variability and uncertainty attributed to chemical class/MOA-specific toxicity, which can exceed 2 orders of magnitude (*1*). Reduced model MSE and increased model R^2 were used to determine if MOA-specific models were improved over models built from all data. Based on these criteria, only models built from neurotoxicants, carbamates, and direct acting OAIs were improved over models built with all data. However, many MOA-specific models, including those developed for MOAs that showed model improvement, were not significant at the $p = 0.05$ level (Table 2). Although increased model p-value is concomitant with reduced degrees of freedom, it also signifies a reduction in model robustness and confidence. While there is evidence to suggest that models may be improved when built with MOA-specific data, abundant data is necessary to ensure model robustness is preserved. Chemical class and MOA-specific ICE models may provide significant improvement in sublethal toxicity prediction; however, the development of ICE models for chronic toxicity is problematic because of relatively limited datasets compared to acute toxicity.

The comparison of MOA-specific models with models built from all data was only conducted for a small group of MOA and subgroups; however, more than half of the MOA-specific models did not demonstrate marked improvement over models built with all data. The ICE models show that chemical classes and MOAs with high toxicity in one wildlife species generally have high degree of toxicity in another species. Also, ICE models describe interspecies variation on a logarithmic scale, in which a large degree of change in actual toxicity values is necessary for a small change in the linear model. This comparison might yield different results if the compared models were of equal sample sizes (i.e., by selecting a random subsample of all data that was equal in size to the MOA-specific dataset); however, this analysis was conducted to identify and provide users with the most robust

ICE models for risk assessment. While it is recognized that MOA and chemical class are important determinants of acute toxicity (*25*), it does not appear necessary to develop MOA-specific models to achieve confident ICE model estimates for use in risk assessment. The results of these analyses suggest that ICE models built from all available data can be used to estimate acute toxicity in wildlife species over a broad range of MOA and chemical classes with known uncertainty.

Safety factors, the generation of species sensitivity distributions (SSDs) with or without body size scaling, or within-compound extrapolation models based on body size alone have been the most commonly used methods for extrapolating toxicity between wildlife species in ecological risk assessment. Safety factors and within-chemical weight-based models have been typically used to extrapolate toxicity from a test species to untested but identified ecological receptors, while SSDs have been used to identify generic receptors of putative sensitivity. Fairbrother (*26*) and others have recommended greater reliance on empirical data and robust statistical approaches because of the high variation in organism responses and the sometimes arbitrary nature of safety factors. Body size scaling in birds and mammals is an empirical approach that extrapolates from an LD50 (mg/kg) based on organism weight. Scaling factors can vary greatly for different chemicals and where chemical-specific factors are unknown, an average is applied to extrapolate toxicity among wildlife species (*1*−*2*). ICE models provide an additional tool for estimating the acute toxicity of chemicals to a diversity of wildlife species with known uncertainty across MOA and taxonomic level. There may be grounds to incorporate scaling in future ICE models, isolating the proportion of the overall inter-species variance that is the result of chemical-specific scaling from the variance resulting from phylogeny and other scale-independent factors. In theory, this should yield ICE models with lower overall MSE. Unfortunately, such an approach will prove data-intensive and will be of minimal usefulness for new chemicals for which few species have been tested.

The application of ICE models requires selection of the most appropriate surrogate species based on a review of model robustness (R^2, MSE, degrees of freedom (df), p-value), cross-validation success rate, and taxonomic distance. In a hypothetical example, the acute toxicity of a chemical to the red-winged blackbird needs to be estimated in light of available data on the toxicity of the chemical to five species: northern bobwhite, mallard, Japanese quail, fulvous whistling duck, or common grackle (see Supporting Information 2). The grackle has the closest taxonomic distance (*2*), low MSE (0.11), highest cross-validation success rate (98), high R^2 (0.65), and high degrees of freedom (53), and is the best potential surrogate in this example. Japanese quail (MSE = 0.19, R^2 = 0.79, df = 134, cross-validation success rate = 92, taxonomic distance = 4) would be the next best surrogate based on low MSE and high cross-validation success rate, followed by northern bobwhite (MSE = 0.22, R^2 = 0.63, df = 45, cross-validation success rate = 87, taxonomic distance = 4) and mallard (MSE = 0.27, R^2 = 0.48, df = 80, cross-validation success rate = 87, taxonomic distance = 4). Although fulvous whistling duck has the highest model R^2 (0.91) and lowest MSE (0.05), low df (df = 2) and high p-value (0.047) do not make it as good of a surrogate as other species because the range of applicability of the model is limited by the small dataset and the significance of the relationship is weak.

In screening level assessments, high confidence in interspecies toxicity estimation is necessary. To increase confidence in ICE model predictions, confidence limits may be calculated for the expected toxicity value (*15*). Narrow confidence limits represent high confidence that the model fits through the range of datapoints for the specified surrogate species toxicity. ICE model have the most confidence, and can only be cross-validated, within the range of data points that were used to estimate the model. As such, using ICE models to predict a toxicity value from surrogate species toxicity outside the range of values used to develop the model increases model uncertainty. Because model certainty is also a function of model attributes, further analyses should include a mechanistic interpretation of model attributes such as slope and intercept.

ICE Model information (including minimum, maximum, and average values of the surrogate species used to derive each model) and equations are provided in the Supporting Information 2 to facilitate use of these models in risk assessment. Additionally, the models developed here are available as a predictive modeling tool (Web-ICE) through the EPA Center for Exposure Assessment Modeling website (http://www.epa.gov/ceampubl/fchain/index.htm), where updates to these models will be provided as new data becomes available. This Internet tool provides model information and user guidance, and calculates toxicity values and confidence intervals for selected models. ICE models can be used when toxicity data for a specific chemical are available for a selected surrogate species, and there is an existing model between the species pair of interest. Models should be selected appropriately (close taxonomic distance, low MSE, high R^2, high df, high cross-validation success rate) for application in wildlife risk assessment to reduce uncertainty in determining species sensitivity. The application of ICE models should be considered as an additional approach for extrapolating across species and for assessing relative species sensitivity.

Acknowledgments

For assistance with database development, we thank Sonny Mayer (U.S. EPA, retired), and Thomas Steeger and Brian Montague (U.S. EPA, Office of Pesticide Programs), and Alain Baril and Brian Collins (Environment Canada). Rick Bennet (U.S. EPA) provided a valuable critique of the manuscript. Also, thanks to our support personnel: Deborah Vivian, Marion Marchetto, Anthony DiGirolamo, Christel Chancy, Nathan Lemoine, Nicole Allard, and Cheryl McGill (U.S. EPA, Gulf Ecology Division).

Supporting Information Available

Additional information on chemicals used in model development are available in Supporting Information 1. ICE models described in this study are available in Supporting Information 2. This material is available free of charge via the Internet at http://pubs.acs.org.

Literature Cited

(1) Mineau, P.; Baril, A.; Collins, B. T.; Duffe, J.; Joerman, G.; Luttik, R. Pesticide acute toxicity reference values for birds. *Rev. Environ. Contam. Toxicol.* **2001**, *170*, 13−74.

(2) Mineau, P.; Collins, B. T.; Baril, A. On the use of scaling factors to improve interspecies extrapolation of acute toxicity in birds. *Reg. Toxicol. Pharmacol.* **1996**, *24*, 24−29.

(3) Asfaw, A.; Ellersieck, M. R.; Mayer, F. L. *Interspecies Correlation Estimations (ICE) for Acute Toxicity to Aquatic Organisms and Wildlife. II. User Manual and Software*, EPA/600/R-03/106; United States Environmental Protection Agency: Washington, DC, 2003.

(4) Luttik, R.; Mineau, P.; Roelofs, W. A review of interspecies toxicity extrapolation in birds and mammals and a proposal for long-term toxicity data. *Ecotoxicology* **2005**, *14*, 817−832.

(5) Luttik, R.; Aldenberg, T. Extrapolation factors for small samples of pesticide toxicity data: special focus on LD50 values for birds and mammals. *Environ. Toxicol. Chem.* **1997**, *16*, 1785−88.

(6) Dyer, S. D.; Versteeg, D. J.; Belanger, S. E.; Chaney, J. G.; Mayer, F. L. Interspecies correlation estimates predict protective environmental concentrations. *Environ Sci. Technol.* **2006**, *40*, 3102−3111.

(7) Colton, J. A. Toxicity-extrapolation models. In *Ecological Modeling in Risk Assessment: Chemical Effects on Populations,*

Ecosystems, and Landscapes; Pastorok, R. A., Bartell, S. M., Ferson, S., Ginzbury, L. R., Eds.; Lewis Publishers: Boca Raton, FL, 2002; p 181, Chapter 12.

(8) Hudson, R. H.; Tucker, R. K.; Haegele, M. A. *Handbook of Toxicity of Pesticides to Wildlife*; Resource Publ. 153; United States Fish and Wildlife Service: Washington, DC, 1984. p 90.

(9) Shafer, E. W., Jr.; Bowles, W. A., Jr.; Hurlbut, J. The acute oral toxicity, repellency and hazard potential of 998 chemicals to one or more species of wild and domestic birds. *Arch. Environ. Contam Toxicol.* **1983**, *12*, 355−382.

(10) Shafer, E. W., Jr.; Bowles, W. A., Jr. Acute oral toxicity and repellency of 933 chemicals to house and deer mice. *Arch. Environ. Contam Toxicol.* **1985**, *14*, 111−129.

(11) Shafer, E. W., Jr.; Bowles, W. A., Jr. *Toxicity, Repellency or Phototoxicity of 979 Chemicals to Birds, Mammals and Plants*; Research Report No. 04−01; United States Department of Agriculture: Fort Collins, CO, 2004; p 118.

(12) Smith, G. J. *Pesticide Use and Toxicology in Relation to Wildlife: Organophosphorus and Carbamate Compounds*; Resource Publication 170; United States Department of the Interior: Washington, DC, 1987; p 171.

(13) Baril, A.; Jobin, B.; Mineau, P.; Collins, B. T. *A Consideration of Inter-Species Variability in the Use of the Median Lethal Dose (LD_{50}) in Avian Risk Assessment*; Technical Report No. 216; Canada Wildlife Service: Ottawa, Canada, 1994; p 12 and Appendices.

(14) Insightful. *S-plus Programmer's Guide*; Insightful Corporation: Seattle, WA, 2001.

(15) Dowdy, S.; Weardon, S. *Statistics for Research*; John Wiley and Sons: New York, 1991.

(16) Leigh, G. J.; Favre, H. A.; Metanomski, W. V. *Principles of Chemical Nomenclature. A Guide to IUPAC Recommendations*; Blackwell Science: Oxford, UK 1998.

(17) Barron, M. G.; Hansen, J.; Lipton, J. Association between contaminant tissue residues and adverse effects in aquatic organisms. *Rev. Environ. Contamin. Toxicol.* **2002**, *173*, 1−37.

(18) Barron, M. G. Environmental Contaminants Altering Behavior. In *Behavioural Ecotoxicology*; Dell'Omo, G., Ed.; John Wiley and Sons: West Sussex, UK, 2002; pp 167−186, Chapter 3.

(19) Budavari, S. *The MERCK Index,* 11th ed. MERCK and Company: Rahway, NJ, 1989.

(20) Russom, C. I.; Bradbury, S. P.; Broderius, S. J.; Hammermeister, D. E.; Drummond, R. A. Predicting modes of toxic action from chemical structure: acute toxicity in the fathead minnow (pimephales promelas). *Environ. Toxicol. Chem.* **1997**, *16*, 948−967.

(21) Wishart, D. S.; Knox, C.; Guo, A. C.; Shrivastava, S.; Hassanali, M.; Stothard, P.; Chang, Z.; Woolsey, J. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **2006**, *34*, D668−D672.

(22) Conover, W. J. *Practical Nonparametric Statistics,* 3rd ed; John Wiley and Sons.: New York, 1999.

(23) Breiman, L. Heuristics of instability and stabilization in model selection. *Ann. Stat.s* **1996**, *24*, 2350−2383.

(24) Efron, B. Estimating the error rate of a prediction rule: improvement on cross-validation. *J Am. Stat. Assoc* **1983**, *78*, 316−331.

(25) Mineau, P.; Whiteside, M. Lethal risk to birds from insecticide use in the United States: a spatial and temporal analysis. *Environ. Toxicol. Chem.* **2006**, *25*, 1214−1222.

(26) Fairbrother, A. Ecotoxicology: Use of safety factors in ecological risk assessment. In *Encyclopedia of Ecology*; Jørgensen, S. E. Ed.; Elsevier: Amsterdam, The Netherlands, 2007, in print.

Received for review February 12, 2007. Revised manuscript received April 23, 2007. Accepted April 26, 2007.

ES070359O