Chiu et al.:

A Tiered, Bayesian Approach to Estimating Population Variability for Regulatory Decision-Making

Supplementary Data

Materials and methods

Tab. S1: Chemicals considered to have less than reliable individual EC_{10} estimates (due to poor convergence, poor model fit, or low potency)

Number	CASRN	Rationale
7	50-02-2	Low potency
22	72-14-0	Low potency
23	91-53-2	Low potency
33	25152-84-5	Not fully converged
36	126-98-7	Low potency
41	4376-20-9	Low potency
44	116-06-3	Low potency
59	102-36-3	Low potency
70	100-42-5	Low potency
71	110-57-6	Low potency
86	81-55-0	Not fully converged
90	118-75-2	High uncertainty in one batch
94	87-68-3	Low potency
95	393-75-9	Low potency
97	93-05-0	Not fully converged
99	84-69-5	Low potency
110	598-91-4	Not fully converged
113	117-81-7	Low potency
117	6338-41-6	Low potency
124	6112-76-1	Some individuals high uncertainty
127	762-75-4	Low potency
134	106-89-8	Low potency
138	101-72-4	Not fully converged
141	97-32-5	Low potency
143	7779-30-8	Low potency
144	103-33-3	Low potency
154	64-86-8	Poor fit
158	137-30-4	Poor fit
161	52645-53-1	Low potency
163	23541-50-6	Poor fit
166	20830-75-5	Poor fit
168	67-97-0	Not fully converged

doi:10.14573/altex.1608251s

Tab. S2: Chemicals considered to have reliable individual-based EC_{10} estimates but less than reliable population-based EC_{10} estimates (due to correlation, departure from normal distribution)

Number	CASRN	Rationale	Individual-based $TDVF_{01}$ estimate
17	396-01-0	Correlation	4.25
19	55-56-1	Correlation	7.14
25	130-26-7	Correlation	2.91
34	148-24-3	Correlation	5.45
39	127-47-9	Multi-modal	1.81
54	154-42-7	Correlation	30.4
76	548-62-9	Multi-modal	3.14
92	4252-78-2	Correlation	2.51
98	60-57-1	Correlation	3.76
100	117-10-2	Multi-modal	2.79
109	98-29-3	Correlation	2.06
119	137-26-8	Correlation	9.29
140	55-86-7	Correlation	7.26
145	26530-20-1	Correlation	5.49
147	309-00-2	Correlation	3.75
148	7789-12-0	Correlation	5.95
151	2437-29-8	Multi-modal	2.81
162	39025-24-6	Correlation	2.56
170	379-79-3	Correlation	5.22

Tab. S3: Chemicals considered to have reliable individual-based EC_{10} estimates and reliable population-based EC_{10} estimates

Number	CASRN	Individual-based $TDVF_{01}$ estimate	Population-based $TDVF_{01}$ estimate
1	66-71-7	1.15	1.14
2	1948-33-0	2.03	1.93
3	446-86-6	6.89	10.88
4	2016-88-8	2.45	2.43
5	549-18-8	1.48	1.55
6	106-51-4	4.52	3.49
8	305-03-3	2.79	2.87
9	120-80-9	3.2	2.96
10	50-28-2	3.13	2.47
11	51-21-8	11.2	10.24
12	13311-84-7	1.83	1.79
13	123-31-9	2.4	2.43
14	73-31-4	2.98	2.71
15	57-83-0	1.41	1.43
16	50-55-5	1.45	1.44
18	152-11-4	3.63	3.50
20	58-14-0	6.25	6.61
21	58-54-8	2.05	2.48
24	121-54-0	1.9	1.98
26	834-28-6	3.01	3.14
27	458-37-7	1.89	1.90
28	97-18-7	3.47	2.70
29	478-43-3	3.77	3.94
30	54965-24-1	1.45	1.47
31	1675-54-3	1.67	1.68
32	533-74-4	3.34	4.33
35	77439-76-0	2.2	1.98
37	34256-82-1	1.4	1.44
38	7778-50-9	5.46	4.92
40	15663-27-1	3.23	2.55
42	55981-09-4	1.98	1.81
43	142-83-6	1.99	2.09
45	123-30-8	1.69	1.62
46	472-86-6	1.57	1.68
47	116-31-4	2.29	2.02
48	879-39-0	1.92	1.86
49	2243-76-7	2.63	2.65
50	961-11-5	2.12	1.96

Number	CASRN	Individual-based $TDVF_{01}$ estimate	Population-based $TDVF_{01}$ estimate
51	115-09-3	14.4	16.03
52	83-26-1	1.45	1.98
53	74-31-7	1.83	1.84
55	95-54-5	3.15	3.43
56	104-40-5	2.06	2.05
57	10108-64-2	2.03	2.22
58	65558-69-2	2.06	2.34
60	10563-29-8	3.22	3.08
61	446-35-5	3.06	2.98
62	62-73-7	1.64	1.67
63	143-50-0	1.55	1.56
64	538-71-6	2.59	2.23
65	14866-33-2	4.42	5.39
66	140-72-7	3.23	3.10
67	17831-71-9	2.1	1.94
68	133-06-2	2.57	2.04
69	62-38-4	8.45	71.04
72	1191-41-9	1.65	1.90
73	134-32-7	3.38	3.24
74	5153-67-3	2.79	2.81
75	111-30-8	1.58	1.64
77	485-47-2	2.6	2.72
78	3018-12-0	1.72	1.66
79	90-41-5	2.92	2.94
80	100-22-1	2.81	2.67
81	2451-62-9	4.4	4.50
82	95-55-6	4.5	4.65
83	70-30-4	2.59	2.47
84	532-27-4	1.83	1.87
85	4074-88-8	2.56	2.17
87	610-39-9	3.03	2.76
88	764-42-1	2.25	2.41
89	33229-34-4	1.49	1.42
91	13048-33-4	1.45	1.41
93	496-72-0	3.72	4.29
96	156-10-5	4.49	6.44
101	55-55-0	3.24	3.53
102	140-49-8	2.1	2.04

Number	CASRN	Individual-based $TDVF_{01}$ estimate	Population-based $TDVF_{01}$ estimate
103	1239-45-8	4.73	4.69
104	3252-43-5	2.02	2.45
105	3524-68-3	4.17	4.01
106	12789-03-6	1.85	1.93
107	70-25-7	3.7	3.03
108	630-16-0	2.61	2.18
111	538-75-0	2.65	3.13
112	115-29-7	1.58	1.60
114	96-69-5	2.44	2.33
115	57-97-6	2.12	2.08
116	6317-18-6	3.01	2.93
118	793-24-8	1.56	1.50
120	95-85-2	7.86	7.43
121	7487-94-7	17.1	34.94
122	95-84-1	1.6	1.76
123	1271-19-8	2.42	3.24
125	100-27-6	2.23	2.20
126	95-83-0	3.63	3.84
128	138-89-6	7.8	4.88
129	21829-25-4	5.65	5.12
130	619-23-8	1.98	2.01
131	612-23-7	5.75	5.51
132	12083-48-6	3.67	2.45
133	933-78-8	1.85	2.70
135	101-90-6	3.74	3.42
136	101-96-2	2.59	2.32
137	6219-89-2	3.64	3.66
139	4901-51-3	1.26	1.31
142	1421-63-2	1.94	1.98
146	8001-35-2	1.58	1.57
149	1465-25-4	1.42	1.40
150	113-92-8	4.37	3.86
152	4/4/5743	2.65	2.58
153	52417-22-8	3.62	3.75
155	517-28-2	1.97	1.99
156	60207-90-1	1.41	1.50
157	88671-89-0	1.28	1.31
159	66-81-9	3.39	4.41

Number	CASRN	Individual-based $TDVF_{01}$ estimate	Population-based $TDVF_{01}$ estimate
160	1162-65-8	2.7	2.57
164	149845-06-7	1.76	1.64
165	434-07-1	2.29	1.99
167	105-11-3	1.78	1.68
169	13463-41-7	8.83	11.67

Stan (version 2.6.2) statistical model code

```
data {
   int<lower=0> Nconc;       // Number of concentrations in each dataset
   vector[Nconc] x_conc;                  // Chemical concentrations
   int<lower=0> Nindiv;                   // Number of individuals
   int<lower=0> Nsets;       // Number of data sets
   vector[Nconc] y[Nsets];                // Responses
   int<lower=1,upper=Nindiv> indiv[Nsets];     // Which individual for each data set
   int<lower=1> Nquants;    // Number of quantiles of EC10 to calculate
   vector[Nquants] quants;   // Quantiles (e.g., c(0.01,0.025,0.5,0.975,0.99))
   int<lower=1> Nbatch; // Number of batches
   int<lower=1,upper=Nbatch> batch[Nsets];    // Which batch for each dataset
   }
parameters {
   real m_0;                              // pop mean for beta_0
   real<lower=0> m_1;                            // pop mean for beta_1 (non-negative)
   real<lower=0> sd_0;                           // pop sd for beta_0
   real<lower=0> sd_1;                           // pop sd for beta_1
   real<lower=0> sigma_y[Nbatch]; // batch-specific error scale on y
   real m_theta_0;
   real<lower=0> sd_theta_0;
   real z_0[Nindiv];                      // z-score for individual beta_0
   real z_1[Nindiv];                      // z-score for individual beta_1
   real z_theta_0[Nsets];       // z-score for dataset theta_0
}
transformed parameters {
   real beta_0[Nindiv];
   real beta_1[Nindiv];
   real theta_0[Nsets];
   for (i in 1:Nindiv) {
            beta_0[i] <- m_0 + sd_0 * z_0[i];
            beta_1[i] <- m_1 + sd_1 * z_1[i];
   }
   for (i in 1:Nsets) {
            theta_0[i] <- m_theta_0 + sd_theta_0 * z_theta_0[i];
   }
}
model {
   // Local scope temporary variable
   vector[Nconc] eta;
   // Prior distributions
   m_0 ~ normal(-2,10);
   m_1 ~ normal(0,5);         // Half-normal
   sd_0 ~ normal(0,1);        // Half-normal
   sd_1 ~ normal(0,1);        // Half-normal
```

```
  sigma_y ~ normal(0,10);  // Half-normal
  m_theta_0 ~ normal(0,5);
  sd_theta_0 ~ normal(0,1);
  z_0 ~ normal(0,1);
  z_1 ~ normal(0,1);
  z_theta_0 ~ normal(0,1);
  for (s in 1:Nsets) {
        for (c in 1:Nconc)
              eta[c] <- theta_0[s] + (-100-theta_0[s]) * inv_logit(beta_0[indiv[s]] + beta_1[indiv[s]] * log(x_conc[c]));
        y[s] ~ student_t(5, eta, sigma_y[batch[s]]); // Use student-t with dof=5 to address outliers
  }
}
generated quantities {
  vector[Nindiv] ec10;              // Concentration at which the relative response is 10%
  real ec10_GM;                            // Geometric mean
  real ec10_GSD;                           // Geometric standard deviation
  real ec10_GSD_approx;             // Approximate GSD derived from m_0,m_1,sd_0, and sd_1
  real ec10_median;                        // Median
  vector[Nquants] ec10_quants;      // Quantiles
  vector[Nquants] ec10_quant_ratios;       // Ratios of median to quantiles

  for (i in 1:Nindiv)
        ec10[i] <- exp((log(0.1/(1-0.1)) - beta_0[i])/beta_1[i]);
  ec10_GM <- exp(mean(log(ec10)));
  ec10_GSD <- exp(sd(log(ec10)));
  ec10_GSD_approx <- exp(sqrt((sd_0 / m_1)^2 + (sd_1 / m_1)^2 * ((log(0.1/(1-0.1)) - m_0)/m_1)^2));
  {
  int ec10_indx[Nindiv];
  real p;
  real m;
  real g;
  int i;
  int itmp;
  ec10_indx <- sort_indices_asc(ec10);
  p<- 0.5;
  m<-1-p;
  itmp<-1;
  while (itmp <(Nindiv*p + m))
        itmp<-itmp+1;
  i <- itmp-1;
  g <- Nindiv*p + m - i;
  ec10_median <- (1-g)*ec10[ec10_indx[i]]+g*ec10[ec10_indx[i+1]];
  for (q in 1:Nquants) {
        p<- quants[q];
        m<- 1-p;
        itmp<-1;
        while (itmp <(Nindiv*p + m))
              itmp<-itmp+1;
        i <- itmp-1;
        g <- Nindiv*p + m - i;
        ec10_quants[q] <- (1-g)*ec10[ec10_indx[i]]+g*ec10[ec10_indx[i+1]];
  }
  for (q in 1:Nquants) {
        ec10_quant_ratios[q] <- ec10_median/ec10_quants[q];
  }
  }
}
```