SUPPLEMENTAL MATERIALS

Contents

1. **Supplementary methods**

   1) Conversion for CLoral of the geometric mean (GM) and geometric standard deviation (GSD)

   2) Markov chain Monte-Carlo (MCMC) method

   3) Evaluation of ethnic ratios

   4) The essential part of source code of error model in the present study

2. **Figures**

   Figure S1

   Figure S2

   Figure S3

   Figure S4

Supplementary Methods

1) Conversion for CLoral of the geometric mean (GM) and geometric standard deviation (GSD)

When only the arithmetic mean and standard deviation were described in the report, they were converted to GM and GSD by Eq. 5 and Eq. 6, respectively.

$$GM = \sqrt{\frac{M^4}{M^2+SD^2}} \qquad \text{Eq. 5}$$

$$GSD = \exp\sqrt{\ln\left(\frac{M^2+SD^2}{M^2}\right)} \qquad \text{Eq. 6}$$

where M and SD represent the arithmetic mean and standard deviation, respectively.

2) Markov chain Monte-Carlo (MCMC) method

We implemented Eqs.1 to 4 in the manuscript using R (Version 4.0.4) and R2OpenBUGS package (Version 3.2-3.2.1) for sampling of ln CL_{oral} in Westerners, ERs, and relating variances for each drug and each of the drug groups. For prior distributions, we used uninformative distributions; i.e., normal distribution with mean 0 and variance 1,000 for fixed-effect parameters; gamma distribution with shape and scale parameters both of 0.001 for variance parameters. The model code used for the MCMC analysis is provided below. The GS was performed for each drug group. A total of 60,000 iterations per chain (including 10,000 burn-ins) were performed with a number of chains of 3. Resulting 150,000 samples were used to calculate statistics of each parameter (Fig.2, Fig.3b and 3c). The convergence of sampling was judged by the trace plots (Supplementary data, Fig. S1) and R-hat value of almost 1.0 [1,2].

3) Evaluation of ethnic ratios

ER was judged as significant if the 99% confidence interval (CI) of ER did not cross 1.0. We chose this “conservative” CI criteria because the reputation of the test for multiple drugs could cause the multiplicity issue and consequent decline in the statistical power. We checked the appropriateness of this criteria in a simulation analysis (supplementary data, Fig. S2).

4) The essential part of source code (R script) of error model in the present study

```
model = function() {

  for (o in 1:NO) {
    Y[o] ~ dnorm(mu[o], TAU.IIV[o])
    mu[o] <- a[DRUG[o]] + er[DRUG[o]] * JAPANESE[o] + rs[STUDY[o]]
  }

  for (s in 1:NS) {
    rs[s] ~ dnorm(0, tau.isv)
  }
  for (d in 1:ND) {
    er[d] <- b + rd[d]
    rd[d] ~ dnorm(0, tau.idv)
  }

  for (d in 1:ND) {a[d] ~ dnorm(0, 0.001)}
  b ~ dnorm(0, 0.001)
  tau.idv ~ dgamma(0.001, 0.001)
  tau.isv ~ dgamma(0.001, 0.001)
  gcv.idv <- sqrt(exp(var.idv) - 1)
```

```
  gcv.isv <- sqrt(exp(var.isv) - 1)
  var.idv <- 1/tau.idv
  var.isv <- 1/tau.isv
}


model_noidv = function() {

  for (o in 1:NO) {
    Y[o] ~ dnorm(mu[o], TAU.IIV[o])
    mu[o] <- a[DRUG[o]] + er[DRUG[o]] * JAPANESE[o] + rs[STUDY[o]]
  }

  for (s in 1:NS) {
    rs[s] ~ dnorm(0, tau.isv)
  }
  for (d in 1:ND) {
    er[d] <- b
  }

  for (d in 1:ND) {a[d] ~ dnorm(0, 0.001)}
  b ~ dnorm(0, 0.001)
  tau.isv ~ dgamma(0.001, 0.001)
  gcv.isv <- sqrt(exp(var.isv) - 1)
  var.isv <- 1/tau.isv
}



# ND, number of drugs; NS; number of studies
    init = function() {
      list(a=rnorm(max(ND), 0, 0.001),
```

```
        b=rnorm(1, 0, 0.001),
        # rd=rnorm(max(ND), 0, 0),
        rs=rnorm(max(NS), 0, 0),
        # tau.idv=1/log(runif(1)**2+1),
        tau.isv=1/log(runif(1)**2+1)
    )
  }
R2OpenBUGS::bugs(
  data=list(Y=***, # observations (i.e. lnCL)
            TAU.IIV=***, # inverse of variance, 1/(SE^2/N)
            NO=***, # number of observations
            NS=***, # number of studies
            ND=***, # number of drugs
            DRUG=***, # Drug ID
            STUDY=***, # Study ID
            JAPANESE=***), # 1 if Japanese study, else 0
  init=init,
  parameters.to.save=***,
  model.file=***
  n.chains=3,
  n.iter=60000,
  n.burnin=10000,
  n.thin=5,
  working.directory=***,

 )
```

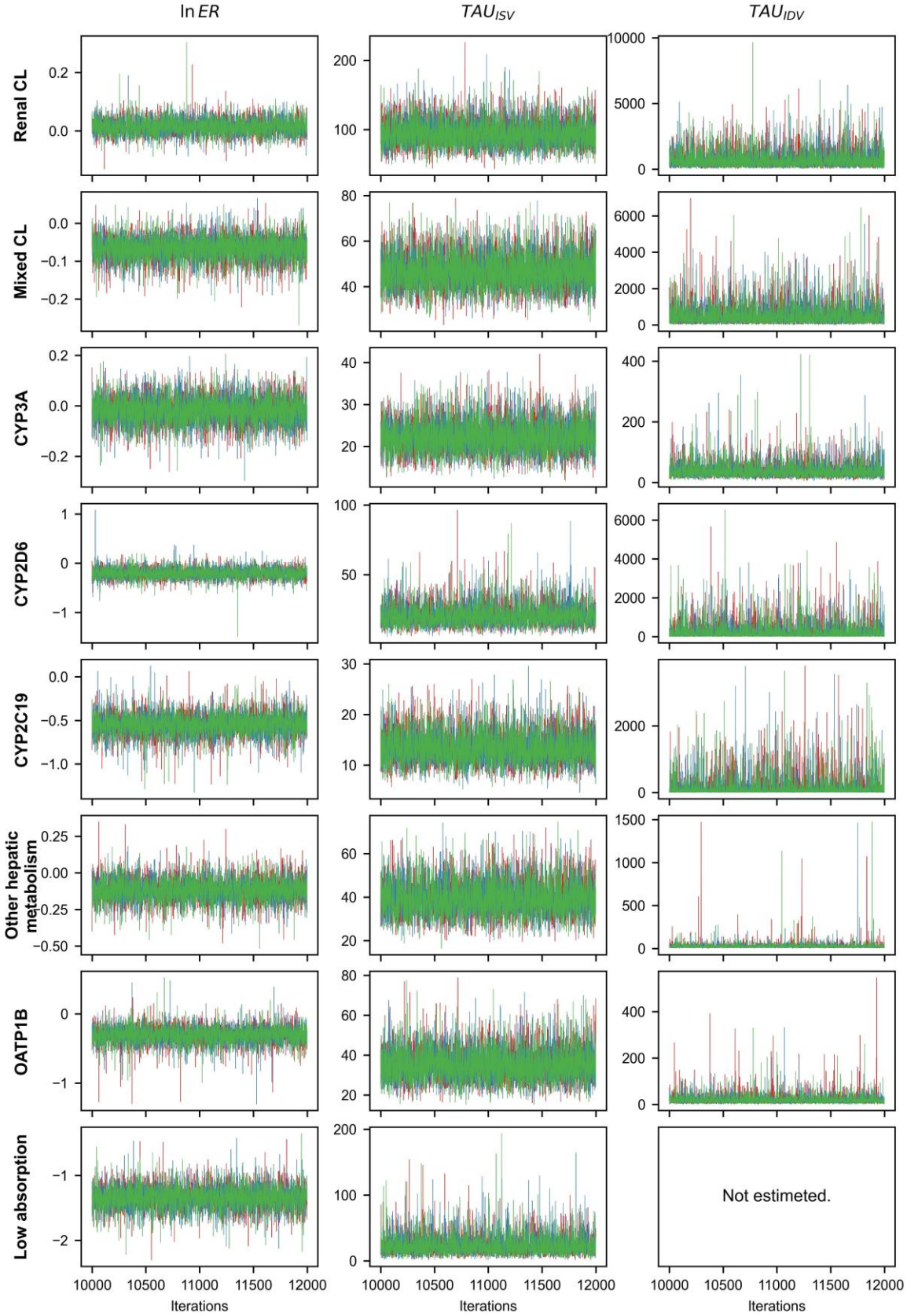
ln ER
TAU_ISV
TAU_IDV
Renal CL
Mixed CL
CYP3A
CYP2D6
CYP2C19
Other hepatic metabolism
OATP1B
Low absorption
Not estimeted.
Iterations
Iterations
Iterations

Figure S1. Trace plots of key parameters.

The graph shows the first 2000 samples after burn-in period per chain. Each color represents a different chain. Since there were only two drugs in the low absorption group, IDV was not sampled (i.e., ethnic ratio was assumed to be common across drugs). ERs are shown on a logarithmic scale, and ISVs and IDVs are shown on the precision scale (i.e., the inverse of variance, tau).

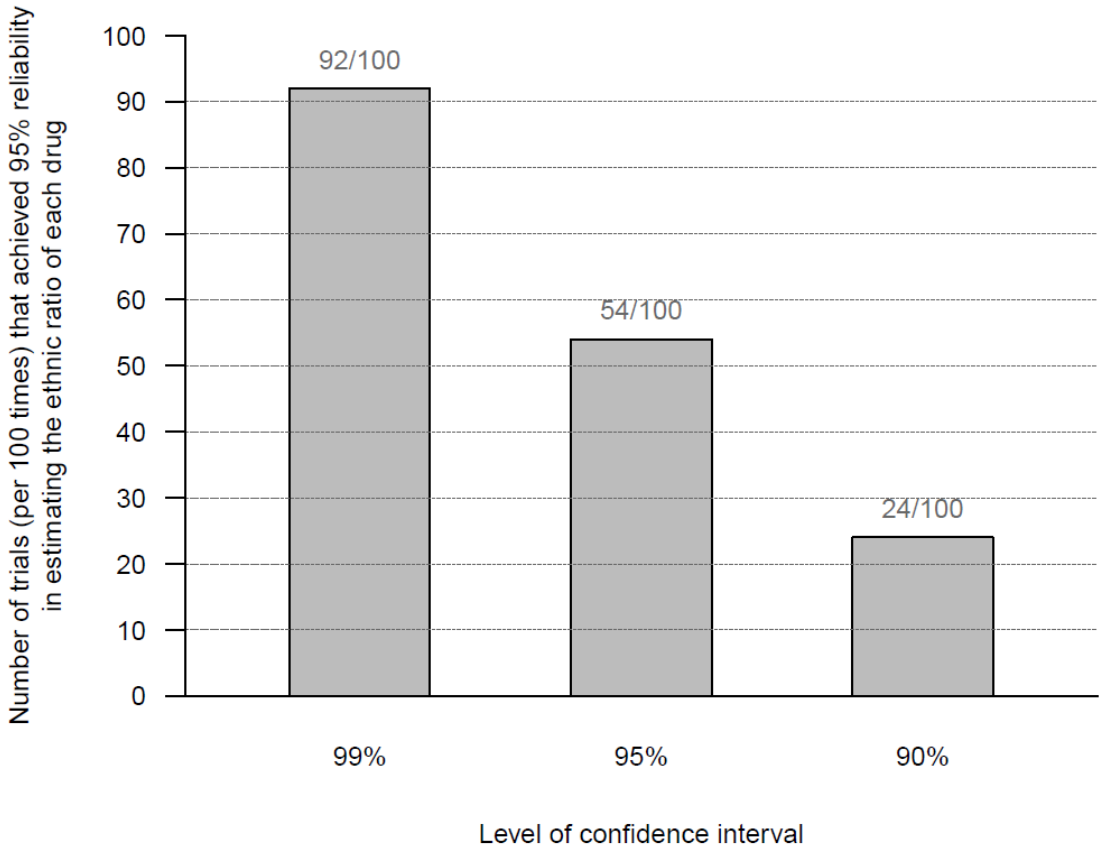

Fig. S2. Comparison of reliability of ethnic ratio estimation at different confidence interval levels.

To quantify the effect of multiplicity in detecting the ethnic ratio of multiple drugs, 3 confidence interval levels (90%, 95%, 99%) were tested in 100 virtual datasets. Each dataset includes 25 drugs, and 6 studies (4 Western and 2 Japanese) were generated for each drug. ERs were sampled from a uniform distribution from 0.5 to 2.0. IIV, ISV, and IDV were sampled randomly from a uniform distribution from 10 to 60%CV. The GS model that was the same as used in the main analysis was applied on each of dataset, and ER, ISV, and IDV were inferred. Since a number of drugs in a dataset is 25, if the true ERs fell within the confidence intervals of estimates with only one or no exception, it was considered as "achieved 95% reliability". When 99% was used as the confidence interval level, 92 datasets out of 100 achieved 95% reliability. Based on this simulation analysis, we selected 99% CI as the appropriate criteria.

Number of subjects	ER[1*]	CYP3A substrates group	
		ISV(%CV)	
		0%	20%
10	1.2	0.33	0.58
	1.5	0.029	0.22
	2.0	<0.01	0.038
20	1.2	<0.01	0.55
	1.5	<0.01	0.19
	2.0	<0.01	0.024
50	1.2	<0.01	0.52
	1.5	<0.01	0.16
	2.0	<0.01	0.017

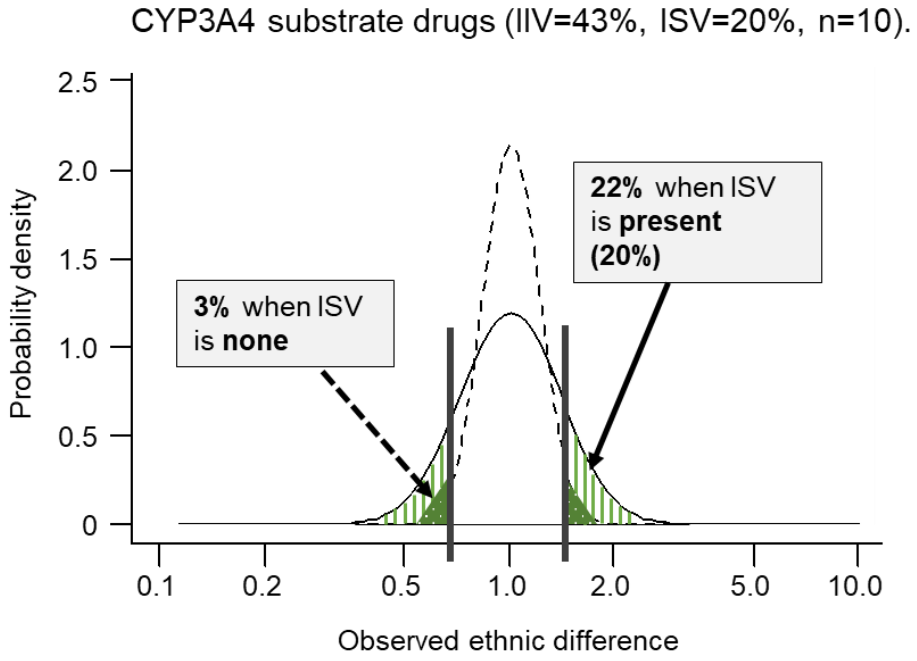

Fig. S3. Estimated probability of erroneously observing ethnic ratios (ER) in Japanese Phase I trials in the case of CYP3A4 substrate drugs even when there is no actual ethnic difference (alpha error).

Top table; Bilateral probability of ER being greater than or equal to *1 and less than or equal to the reciprocal of *1 for CYP3A4 substrate drugs (IIV=43%, ISV=20%). Cases with and without between inter study differences are shown. The figure below is an example of 10 subjects. Solid line is for 20% ISV, dashed line is for 0% ISV. Vertical lines indicate the location of 1.5-fold or its reciprocal fold ER.

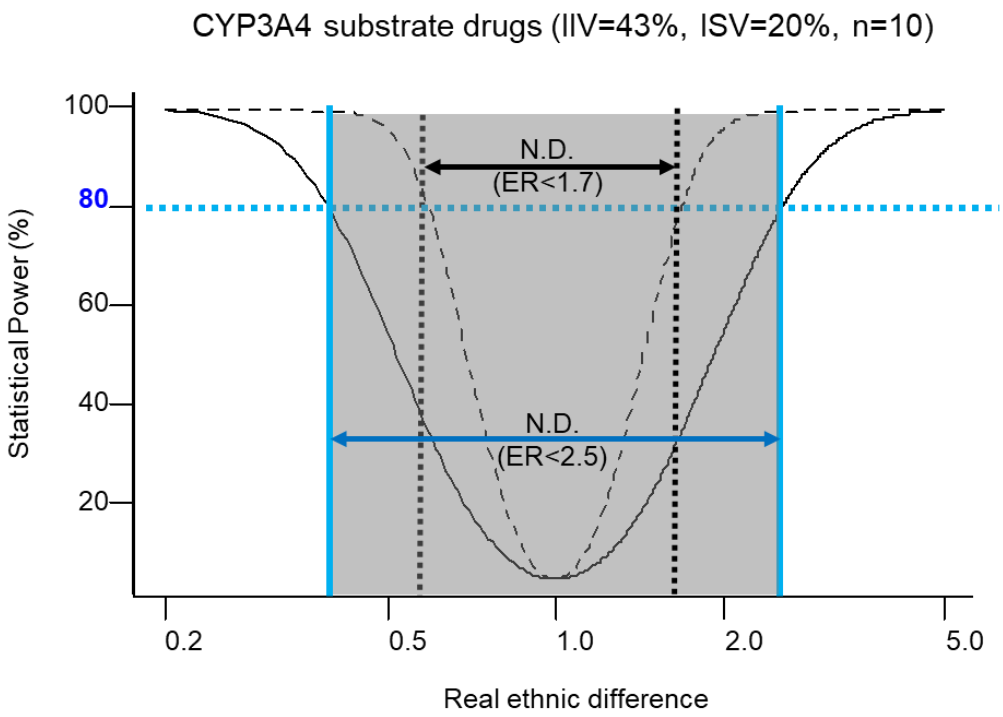

Fig. S4. Estimated power relative to true ethnic ratio (ER) for Japanese Phase I trials in the case of CYP3A4 substrate drugs.

An example of 10 subjects. Solid line is for 20% ISV, dashed line is for 0% ISV. N.D.; ER is not detected even at 80% Power in the shaded areas, respectively.

References

1. Bolker, B. M. et al. Generalized linear mixed models: a practical guide for ecology and evolution. Trends Ecol Evol 24, 127-135, (2009).

2. Annis, J., Miller, B. J. & Palmeri, T. J. Bayesian inference with Stan: A tutorial on adding custom distributions. Behav Res Methods 49, 863-886, (2017).