Individual variability in sensitivity to environmental exposures and its implications for risk assessment

Duncan Thomas

University of Southern California
Los Angeles, USA

NAS meeting on
“Biological factors that underlie individual susceptibility to environmental stressors, and their implications for decision-making”
Washington DC, April 18-19, 2012
Take-Home Messages

• Some effects of heterogeneity on risks over time and attributable risk estimates

• Measured genetic risk modifiers:
  – Candidate gene studies
  – GWAS & GEWIS

• Unmeasured genetic risk modifiers

• Getting genetics into PBPK/PD models

• Epigenetics and trans-generational phenomena

• Policy implications: why regulate non-identifiable risk subgroups? Identifiability
An Important Distinction

- **Susceptibility**
  - Baseline risk of disease in the unexposed
  - Effects of exposure are scale-dependent

- **Sensitivity**
  - Slope of the dose-response for exposure
Effects of Heterogeneity

- **Differential survival**
  - More sensitive individuals are removed from the at-risk population at a faster rate, leaving an increasingly less sensitive population
  - Only important for common diseases

- **Nonlinear dose-response models**
  - Average dose-response of the population is not the dose-response for the average individual
  - True even for rare diseases
Differential Survival

Suppose individual dose-response is linear with random $a_i$ or $b_i$

$$\text{Risk}(Z) = a_i + b_iZ$$

Then the population average risk at age $t$ is

$$E[\text{Risk}(t,Z)] = E(a_i|t) + E(b_i|t)Z$$

E.g., suppose constant baseline risk $\lambda_0(t)$ and gamma-distributed slopes with mean $\mu$ and variance $\sigma^2$

$$R(Z, t) = 1 + \frac{\mu Z}{1 + \Lambda_0(t)\sigma^2 Z/\mu}$$

- Baseline risk = 0.01
- Slopes: mean = 1, variance = 0.25:
Nonlinear Models

- Suppose the dose-response is sigmoid (e.g., logistic or probit) with individual intercepts or slopes

\[
\text{Probit}[\text{Risk}(Z)] = a_i + b_i Z
\]

- Then if \( a_i \sim \mathcal{N}(\alpha, \sigma_a^2) \) and \( b_i = \beta \)

\[
\Pr(\gamma = 1|Z) = \Phi \left( \frac{\mu_\alpha + \beta Z}{\sqrt{1 + \sigma_\alpha^2}} \right)
\]

- Or if \( a_i = \alpha \) and \( b_i \sim \mathcal{N}(\beta, \sigma_b^2) \)

\[
\Pr(Y = 1|Z) = \Phi \left( \frac{\alpha + \mu_\beta Z}{\sqrt{1 + \sigma_\beta^2 Z^2}} \right)
\]
Linear Spline Model

- Suppose the dose-response is linear above a given threshold, with no excess risk below that level.
- Suppose individual thresholds are lognormally distributed.

```
mean 2, 5, 10
LSD = 1

mean 5
LSD = \frac{1}{4}, 1, 4
```
What if we’re interested in some upper percentile of the risk distribution?

- In principle, if we believed in linearity of the individual dose response, we could use the observed nonlinearity to estimate the variance of individual risks.
- But even if this were true, estimates would be poorly estimated.
- If not true, estimates would be hopelessly biased.
- We can estimate variability due to known modifiers.
- What about unknown modifiers?
- Heritability estimates from family data may help.
Sources of Variability

• Age / gender
• Health status
• Other exposures
• Genetic constitution
  – Candidate genes
  – Agnostic scans (GWAS)
  – Residual genetic variation (“dark matter”)
GxE in Single Gene Models

- Standard models, e.g.,

\[
\text{logit}(\text{Risk}) = \alpha + \beta E + \gamma G + \delta G \times E
\]

- Can be complex, e.g., bladder cancer in relation to smoking, red-meat, \textit{NAT2} and \textit{CYP1A2} genotypes

GEWIS Designs &Analyses

• Two-step

Murcray et al, AJE 2009; GE 2011

Mukherjee & Chatterjee, Biometrics 2008; Li & Conti, AJE 2009

\[
\hat{\beta}_{EB} = \frac{\hat{\sigma}_{CC}^2}{\hat{\sigma}_{G E}^2} \hat{\beta}_{CO} + \frac{\hat{\sigma}_{G E}^2}{\hat{\sigma}_{CC}^2} \hat{\beta}_{CC}
\]

Figure 1 | Schematic representation of the two-step ge

\[ n_1 = n_0 = 10,000 \exp(\theta_{GE}) = 0.8 \]

\[ n_1 = n_0 = 10,000 \exp(\theta_{GE}) = 1.0 \]

\[ n_1 = n_0 = 10,000 \exp(\theta_{GE}) = 1.1 \]
Sample Size Requirements for GxE studies

Candidate GxE
($\alpha = 0.05$, $1 - \beta = 0.8$)

GWAS GxE Scan
($\alpha = 10^{-8}$, $1 - \beta = 0.8$)

- $\text{Sample Size (N)}$
- $\text{Interaction Odds Ratio}$

- $\text{MAF}=0.1$, $p_E = 0.1$
- $\text{MAF}=0.4$, $p_E = 0.1$
- or $\text{MAF}=0.1$, $p_E = 0.4$
- $\text{MAF}=0.1$, $p_E = 0.4$
Unmeasured Genetic Variability

- Polygenic models
- Using family studies to estimate “dark matter”, e.g., height & psychoses

Yang et al, Nat Genet, 2010
Longo Allen et al, Nat Genet, 2010
and Purcell et al, Nature, 2009

Figure 2 | Replication of the ISC-derived polygenic component in independent schizophrenia and bipolar disorder samples. Variance explained in the target samples on the basis of scores derived in the entire ISC for five significance thresholds ($P_T < 0.1, 0.2, 0.3, 0.4$ and $0.5$). Plotted left...
Polygenic Models

- Familial RRss decline exponentially with degree of relationship.
- Polygenic model is based on assumption that a trait is determined by a sum of many genes, each with small effects.
- Family’s trait vector is multivariate normal with covariance $\sigma^2 I + \tau^2 K$, where $K$ is the kinship matrix.
- Heritability $h^2$ is the percent of variance due to additive genetic effects: $h^2 = \tau^2 / (\sigma^2 + \tau^2)$.
- For binary traits, we assume risk is a probit function of unobserved “liability”.
- Given $\sigma^2$ and $\tau^2$ estimates, one could predict percentiles of risk distribution.

Figure 1. The Liability Threshold Model for a Disease Prevalence of $K$
An underlying continuous random variable determines disease status. If liability exceeds the threshold $t$, then individuals are affected.
But what about exposure?

• Given family data on risk and exposure, one could fit a model

\[
\text{probit}(\text{Risk}) = a_i + b_i Z
\]

where

\[
a_i \sim \text{MVN}(\alpha, \sigma^2 I + \tau^2 K) \quad \text{and} \quad b_i \sim \text{MVN}(\beta, \nu^2 I + \nu^2 K)
\]

• E.g., family study within A-bomb survivors, but to my knowledge, never been done
Identifiability of Individual Risks
(Greenland & Robins, 1988 – 91)

• Suppose exposure causes every case to occur one day earlier than otherwise

• All cases were contributed to by exposure, so true PAR = 100%

• But population RR is very small, so estimated PAR is also small

• PC cannot be estimated without making unverifiable assumptions about mechanisms and heterogeneity

• Even mean PC not estimable

• Average LLE can be estimated (but not individually, conditional on age at death)
Putting Genetics into PBPK-PD Models

• Physiologically-Based Pharmaco-Kinetic and Pharmaco-Dynamic models are typically deterministic, based on systems of ordinary differential equations

• But there is individual variability in underlying rate parameters, which could depend upon genes

• Bayesian population PBPK models
Consider a single intermediate metabolite $M$, created from exposure $E$ at rate $\lambda_i E$, removed at rate $\mu_i M$, with linear kinetics.

Finally, suppose disease risk depends upon the long-term average of $M$ through a logistic model.

Suppose personal kinetic rate parameters $\lambda_i$ and $\mu_i$ are randomly distributed with means $\lambda_G$ and $\mu_G'$ that depend upon genotypes $G$ and $G'$ at the encoding loci.
Epigenetics

• Changes that affect expression of genes without altering the coding sequence, e.g., DNA methylation, histone modifications, etc.

• Well studied in relation to various diseases (e.g., cancer)

• Less well studied in terms of environmental determinants of epigenetic state

• Potential for epigenetic state to mediate or modify exposure-response relationships

• Hints of possible mechanism for trans-generational exposure effects, but evidence mainly from experimental animal studies so far
DAG for Transgenerational Inheritance

Grandparental genotypes and exposures

Grandparental germline epigenotypes

Maternal inherited epigenotype and exposure

maternal sources

Offspring inherited epigenotype

Offspring somatic epigenotype

Offspring phenotype
Policy Implications

• Why should we regulate risks to non-identifiable subgroups?

• Is the distribution of individual risk even identifiable?