An Efficient Gatekeeper Method for Detecting GxE

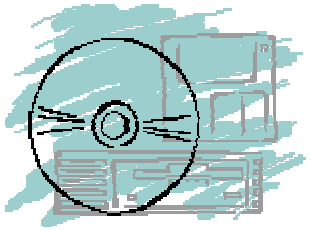
Jimmy T. Efird, Ph.D., M.Sc

Center for Health of Vulnerable Populations
University of North Carolina at Greensboro

Jimmy.efird@stanfordalumni.org
650.248.8282 (cell)

UNCG

Gene-Environment Interaction (GxE)



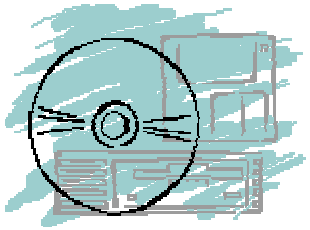
Definition

- GxE denotes the increased risk for disease that occurs when genetic and environmental factors are present in combination.
- Individual factors alone convey little or no risk for disease

UNCG

Methodology

- **Gatekeeper Approach**

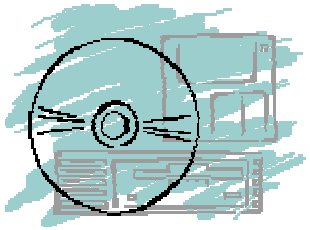


  - **Step 1:** Compute multiplicity adjusted indirect estimates for OR(GE|D).
  - **Step 2:** Use LCI for indirect estimates to screen for interaction effects in a direct association study for GxE.

Methodology

Step 1

- The OR(E|D) in the population may be expressed as

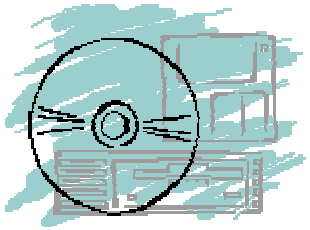


$$OR(E|D)=\frac{P(E|D)/P(\bar{E}|D)}{P(E|\bar{D})/P(\bar{E}|\bar{D})}.$$

UNCG

Methodology

- **Step 1 (cont.)**
  - Assuming that D is relatively rare in both the exposed and unexposed pop. and that genotype freq is independent of E, e.g., P(G|E)=g, OR(E|D) may be written as



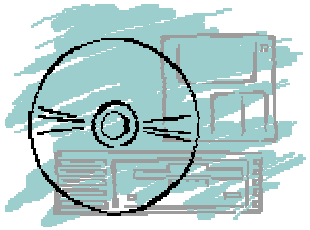
$$\frac{\left[P(D|GE)g \Big/ P(D|\bar{G}\bar{E})\right] + \left[P(D|\bar{G}E)(1-g) \Big/ P(D|\bar{G}\bar{E})\right]}{\left[P(D|G\bar{E})g \Big/ P(D|\bar{G}\bar{E})\right] + (1-g)}.$$

UNCG

Methodology

Step 1 (cont.)

- Considering the simple case



$$OR(G\bar{E}|D)=OR(\bar{G}E|D)=1$$

and rearranging terms, we see that

$$OR(GE|D)=[OR(E|D)-1+g]/g.$$

Methodology

Step 1 (cont.)

- Treating (g) as fixed, it follows that the 95% CI for OR(GE|D) is approximately equal to

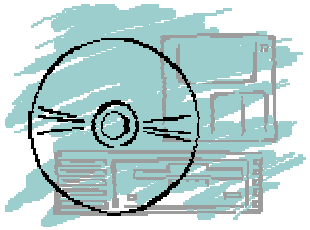
$$\exp\left\{\log\left\{[OR(E|D)-1+g]/g\right\}\pm 1.96\bullet\frac{OR(E|D)\sqrt{\left[\operatorname{var}\left(OR(E|D)\right)\right]}}{\left\{OR(E|D)-1+g\right\}}\right\}.$$

UNCG

Methodology

Step 1 (cont.)

- Substituting the Hochberg step-up P value, denoted by "*", in to the above equation and rearranging, the multiplicity adjusted 95% CI for a set of OR(GE|D) is given as

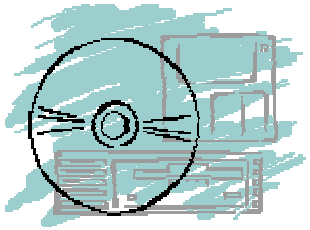


$$CI^{*}_{(1-\alpha/2)} = e^{\left\{\log\left(OR(GE|D)_i\right) \pm z(1-\alpha/2)\sqrt{\left\{\dfrac{\log\left(OR(GE|D)_i\right)}{\Phi^{-1}\left(1-\dfrac{p^{*}_{(j)}}{2}\right)}\right\}^2}\right\}}.$$

UNCG

Methodology

Step 2



- Identify SNPs in which the indirect multiplicity adjusted LCI for OR(GE|D) is greater than a pre-specified value.
- Only consider the latter SNPs when "directly" computing OR(GE|D) in a new study.

Example

- Meta-analysis, childhood obesity
- Environmental OR=1.5
- 95% CI=1.1676 – 1.9270
- Pre-specified threshold value, OR=3.0

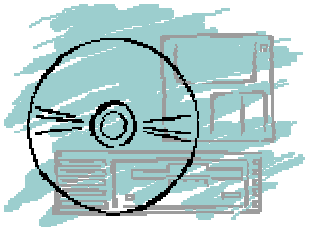
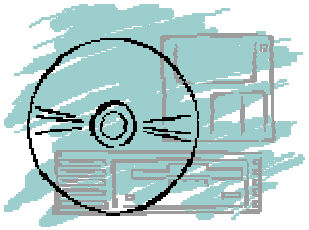


Table 1. Indirect multiplicity adjusted 95% lower confidence intervals for OR(GE|D) given the genotype frequency (g) for 100 obesity related SNPs and OR(E|D)=1.5 (95% LCI=1.1676)

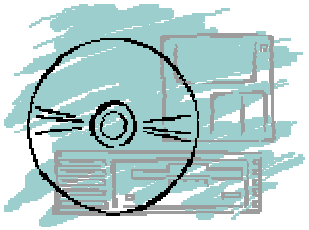
g	OR (GE|D)	Multiplicity Adjusted 95% LCI	g	OR (GE|D)	Multiplicity Adjusted 95% LCI	g	OR (GE|D)	Multiplicity Adjusted 95% LCI	g	OR (GE|D)	Multiplicity Adjusted 95% LCI
74.0	1.68	1.24	49.4	2.01	1.21	39.4	2.27	1.25	26.8	2.87	1.48
73.0	1.68	1.22	49.0	2.02	1.21	39.0	2.28	1.26	25.8	2.94	1.51
72.0	1.69	1.21	48.6	2.03	1.21	38.6	2.30	1.26	24.8	3.02	1.54
71.0	1.70	1.20	48.2	2.04	1.21	38.2	2.31	1.27	23.8	3.10	1.58
70.0	1.71	1.19	47.8	2.05	1.21	37.8	2.32	1.27	22.8	3.19	1.62
69.0	1.72	1.19	47.4	2.05	1.21	37.4	2.34	1.27	21.8	3.29	1.66
68.0	1.74	1.18	47.0	2.06	1.21	37.0	2.35	1.28	20.8	3.40	1.71
67.0	1.75	1.18	46.6	2.07	1.21	36.6	2.37	1.28	19.8	3.53	1.76
66.0	1.76	1.18	46.2	2.08	1.21	36.2	2.38	1.29	18.8	3.66	1.82
65.0	1.77	1.18	45.8	2.09	1.21	35.8	2.40	1.29	17.8	3.81	1.89
64.0	1.78	1.17	45.4	2.10	1.21	35.4	2.41	1.30	16.8	3.98	1.96
63.0	1.79	1.17	45.0	2.11	1.22	35.0	2.43	1.30	15.8	4.16	2.04
62.0	1.81	1.17	44.6	2.12	1.22	34.6	2.45	1.31	14.8	4.38	2.14
61.0	1.82	1.17	44.2	2.13	1.22	34.2	2.46	1.32	13.8	4.62	2.25
60.0	1.83	1.17	43.8	2.14	1.22	33.8	2.48	1.32	12.8	4.91	2.38
59.0	1.85	1.18	43.4	2.15	1.22	33.4	2.50	1.33	11.8	5.24	2.53
58.0	1.86	1.18	43.0	2.16	1.23	33.0	2.52	1.33	10.8	5.63	2.70
57.0	1.88	1.18	42.6	2.17	1.23	32.6	2.53	1.34	9.8	6.10	2.92
56.0	1.89	1.18	42.2	2.18	1.23	32.2	2.55	1.35	8.8	6.68	3.18
55.0	1.91	1.18	41.8	2.20	1.23	31.8	2.57	1.35	7.8	7.41	3.51
54.0	1.93	1.19	41.4	2.21	1.24	31.4	2.59	1.36	6.8	8.35	3.94
53.0	1.94	1.19	41.0	2.22	1.24	31.0	2.61	1.37	5.8	9.62	4.52
52.0	1.96	1.19	40.6	2.23	1.24	29.8	2.68	1.40	4.8	11.42	5.34
51.0	1.98	1.20	40.2	2.24	1.25	28.8	2.74	1.42	3.8	14.16	6.60
49.8	2.00	1.20	39.8	2.26	1.25	27.8	2.80	1.45	2.8	18.86	8.80

Overall Significance Level

- Let α_1 and α_2 denote the family-wise type 1 error for the indirect and direct tests
- The overall procedure will be protected a $\alpha \leq \alpha_1 + \alpha_2$



Overall Significance Level

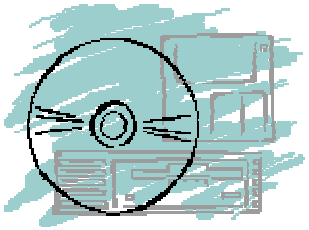


- The significance level of the likelihood ratio test for the global null hypothesis $\beta=0$, where β denotes the vector of beta coefficients in a direct multivariable logistic regression model, may be set to a nominal value, e.g., $\alpha_2 \leq 0.001$.

- Thus, the overall procedure will be protected at $\alpha=0.051$.

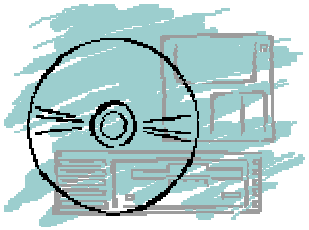
Sample Size Computation

- Sample size may be computed using standard maximum likelihood methods for the logistic regression model and setting the α-level equal to α_2.



Sample Size Computation

- In the example, approximately n_1=260 cases and n_2=260 controls are needed in a direct study to have at least 80% power to detect an OR(GE|D)$\geq$3.18 (corresponding to the LCI of the minimum SNP passing the threshold for entrance into the direct model), given that the squared multiple correction coefficient=0.2, p(E)=0.10 and P(GE)=(0.88)(0.10)=0.088. The overall test procedure is protected at α<0.051 (i.e., $\alpha_1+\alpha_2$=0.05+0.001=0.051)



UNCG