

Dealing with confounder imbalance using CIMBAL v0.6 in R when meta-analyzing cohorts

Debashree Ray, PhD, MStat

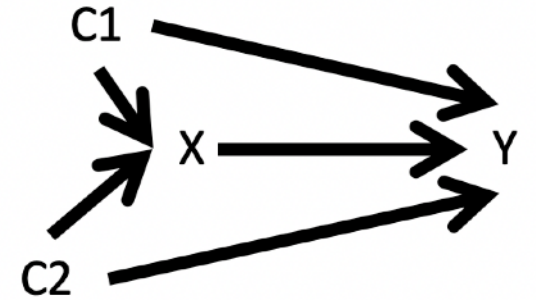
ECHO DAC

Assistant Professor
Department of Epidemiology
Department of Biostatistics
Johns Hopkins University

Contact: dray@jhu.edu

Challenges in meta-analyzing cohorts in ECHO

- Consider outcome Y, exposure X, two possible confounders C1 and C2



- Some cohorts may not have measured any confounder

$$Y = \alpha + \beta_{unadj}X$$

- Some cohorts may have measured both C1 and C2

$$Y = \alpha + \beta_{adj}X + \gamma_1C_1 + \gamma_2C_2$$

- How to handle when one or more cohorts do not measure a set of confounders?
 - meta-analyze only unadjusted estimates?
 - meta-analyze only complete adjusted estimates?
 - meta-analyze a mix of unadjusted and adjusted estimates?

1. Obtain adjusted estimates for incomplete cohorts using CIMBAL

- Assume Cohort 1 has no confounder information
 - reports only unadjusted estimates $(\hat{\beta}_{1,unadj}, \hat{se}_{1,unadj})$
- Independent Cohort 2 has all confounder information
 - reports adjusted estimates $(\hat{\beta}_{2,adj}, \hat{se}_{2,adj})$
 - can also report unadjusted estimates $(\hat{\beta}_{2,unadj}, \hat{se}_{2,unadj})$
- Obtain Cohort 1 adjusted (“corrected”) estimates as:

$$\tilde{\beta}_{1,adj} = \hat{\beta}_{2,adj} - \hat{\beta}_{2,unadj} + \hat{\beta}_{1,unadj}$$

$$\widetilde{se}_{1,adj} = \frac{\hat{se}_{2,adj}}{\hat{se}_{2,unadj}} \times \hat{se}_{1,unadj}$$

2. Collective meta-analysis using CIMBAL

- Meta-analyze cohorts 1 and 2
(accounting for dependence between estimates):

$$\hat{\beta}_{adj} = \frac{\tilde{w}_1 \tilde{\beta}_{1,adj} + \hat{w}_2 \hat{\beta}_{2,adj}}{\tilde{w}_1 + \hat{w}_2}$$

$$\hat{se}_{adj} = \frac{\sqrt{3\tilde{w}_1 + \hat{w}_2 - 2\tilde{w}_1\hat{w}_2 \text{Cov}(\hat{\beta}_{unadj}, \hat{\beta}_{adj})}}{\tilde{w}_1 + \hat{w}_2}$$

where $\tilde{w}_1 = \frac{\text{Cov}(\hat{\beta}_{unadj}, \hat{\beta}_{adj})}{2\text{Cov}(\hat{\beta}_{unadj}, \hat{\beta}_{adj}) + \tilde{se}_1^2 - \hat{se}_2^2}$ and $\hat{w}_2 = 1 - \tilde{w}_1$

- Estimate $\text{Cov}(\hat{\beta}_{unadj}, \hat{\beta}_{adj})$ from cohorts with adjusted estimates

Implementing CIMBAL v0.6

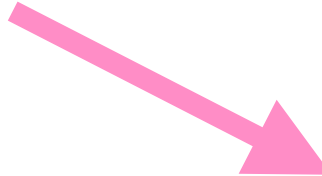
- Created an R program that
 - obtains adjusted estimates for all incomplete cohorts combined
 - meta-analyzes all cohorts
 - current version v0.6 is for ECHO internal use only

- Users need to
 - know how to use R
 - input the collection of estimates as a data frame →

cohort	samplesize	b.unadj	se.unadj	b.adj	se.adj
1	102	-0.54637661	0.5069429	-0.4693723	0.6008678
2	102	-0.67372909	0.6180420	NA	NA
3	102	-1.05209227	0.5539443	-0.2925080	0.6789556
4	102	-0.95885035	0.6165002	-0.4562840	0.7384806
5	102	-0.41629945	0.5322192	0.1851238	0.6648366
6	102	-0.95551136	0.6920797	-0.8136245	0.8445270
7	102	-0.26706279	0.5156975	-0.5095491	0.6073732
8	102	-1.30833282	0.5495699	NA	NA
9	102	-0.11332869	0.4931676	0.4249399	0.6092826
10	102	-1.28865582	0.6011629	-0.6328755	0.7482955
11	102	0.57981850	0.4715728	NA	NA
12	102	-0.09097178	0.5103033	-0.1489653	0.6472801
13	102	0.03636764	0.5137012	0.9364682	0.7646166
14	102	-0.94504015	0.4806770	0.5541546	0.6847592
15	102	-0.80760243	0.4931024	NA	NA
16	102	-0.71562004	0.4965247	-0.9293616	0.6147366
17	102	0.48492928	0.5487416	0.8487228	0.7455959
18	102	-0.03352269	0.4777256	0.1885899	0.5968010
19	102	-0.39641527	0.4366502	-0.4949824	0.5503682
20	102	0.09976453	0.4645534	-0.1327242	0.6467754

Implementing CIMBAL v0.6

```
> source("cimbal_v0.6.R")
> cimbal
function(dat, mute.msgs=FALSE, ncohort.thresh=25){
```



No. of complete cohorts to be used to determine covariance

$$\text{Cov}(\hat{\beta}_{unadj}, \hat{\beta}_{adj})$$

cohort	samplesize	b.unadj	se.unadj	b.adj	se.adj
1	102	-0.54637661	0.5069429	-0.4693723	0.6008678
2	102	-0.67372909	0.6180420	NA	NA
3	102	-1.05209227	0.5539443	-0.2925080	0.6789556
4	102	-0.95885035	0.6165002	-0.4562840	0.7384806
5	102	-0.41629945	0.5322192	0.1851238	0.6648366
6	102	-0.95551136	0.6920797	-0.8136245	0.8445270
7	102	-0.26706279	0.5156975	-0.5095491	0.6073732
8	102	-1.30833282	0.5495699	NA	NA
9	102	-0.11332869	0.4931676	0.4249399	0.6092826
10	102	-1.28865582	0.6011629	-0.6328755	0.7482955
11	102	0.57981850	0.4715728	NA	NA
12	102	-0.09097178	0.5103033	-0.1489653	0.6472801
13	102	0.03636764	0.5137012	0.9364682	0.7646166
14	102	-0.94504015	0.4806770	0.5541546	0.6847592
15	102	-0.80760243	0.4931024	NA	NA
16	102	-0.71562004	0.4965247	-0.9293616	0.6147366
17	102	0.48492928	0.5487416	0.8487228	0.7455959
18	102	-0.03352269	0.4777256	0.1885899	0.5968010
19	102	-0.39641527	0.4366502	-0.4949824	0.5503682
20	102	0.09976453	0.4645534	-0.1327242	0.6467754

Implementing CIMBAL v0.6

```
> cimbal(dat=mydat.inc, mute.msgs=FALSE, ncohort.thresh=25)
```

```
$dat.corrected
```

	cohort	samplesize	b.unadj	se.unadj	b.adj	se.adj
2	1	102	-0.54637661	0.5069429	-0.46937230	0.6008678
8	3	102	-1.05209227	0.5539443	-0.29250797	0.6789556
11	4	102	-0.95885035	0.6165002	-0.45628396	0.7384806
14	5	102	-0.41629945	0.5322192	0.18512376	0.6648366
17	6	102	-0.95551136	0.6920797	-0.81362453	0.8445270
20	7	102	-0.26706279	0.5156975	-0.50954914	0.6073732
26	9	102	-0.11332869	0.4931676	0.42493991	0.6092826
29	10	102	-1.28865582	0.6011629	-0.63287549	0.7482955
35	12	102	-0.09097178	0.5103033	-0.14896527	0.6472801
38	13	102	0.03636764	0.5137012	0.93646824	0.7646166
41	14	102	-0.94504015	0.4806770	0.55415464	0.6847592
47	16	102	-0.71562004	0.4965247	-0.92936159	0.6147366
50	17	102	0.48492928	0.5487416	0.84872284	0.7455959
53	18	102	-0.03352269	0.4777256	0.18858986	0.5968010
56	19	102	-0.39641527	0.4366502	-0.49498237	0.5503682
59	20	102	0.09976453	0.4645534	-0.13272425	0.6467754
65	22	102	-1.20019208	0.6628141	-1.39772706	0.8204493
71	24	102	-0.71562004	0.4965247	0.87936186	0.7316906
74	25	102	0.51879379	0.5472877	0.42532527	0.8156624
77	26	102	-0.03593201	0.4514444	0.09526628	0.5989131
80	27	102	1.38629436	0.5254511	2.06333909	0.9284734
83	28	102	-2.69011250	1.0519918	-2.03622927	1.1901544
86	29	102	-0.99039870	0.4946601	-0.51483850	0.6606504
89	30	102	-1.23125676	0.5920043	-1.08949318	0.7532382
92	31	102	-1.05220898	0.4810754	-0.83001470	0.6090956
95	32	102	-0.08768500	0.5151978	0.80016224	0.6941748
107	36	102	-0.69314718	0.4886533	-0.09816650	0.6614800
110	37	102	-0.62371867	0.4857896	-0.81444836	0.6389800
113	38	102	0.02643326	0.4737275	0.11912143	0.5457148
119	40	102	-0.43891304	0.5888327	-0.06447643	0.7677157
31	2+8+11+15+21+23+33+34+35+39	1020	NA	NA	-0.18116804	0.2116642

→ cohorts with complete information

```
$beta.meta.cimbal  
[1] -0.1466263
```

```
$SE.meta.cimbal  
[1] 0.1161116
```

→ meta-analyzed estimate
and its SE from all cohorts

“corrected” adjusted
estimates for the
combined incomplete cohorts

Final Notes

- CIMBAL: a new approach
 - to borrow information from complete cohorts
 - to get adjusted estimates for cohorts reporting only unadjusted estimates
- Crucial assumption: All cohorts come from the same underlying population
- Advantage: better than commonly used meta-analysis approaches in terms of MSE and/or type I error
- Limitation: requires at least 20 cohorts to appropriately estimate CIMBAL meta-analyzed SE
- Contact dray@jhu.edu to report any issue or question on CIMBAL.