Meta-analysis of genetic association studies

(Zintzaras and Lau 2008 J Clin Epidemiol)

Genetic association studies (GAS) assess the association between disease status and genetic variants (gene polymorphisms) in a population.

Often for specific disease multiple GAS are conducted for the same variant.

However, the study results are never completely homogenous and it is difficult to make inferences whether a variant is responsible for developing the disease.

Then, meta-analysis can play a role.

Example – MTHFR C677T and Breast Cancer

Eleven GAS were conducted to investigate the association between MTHFR C677T (wt=C and mt=T) variant and susceptibility to breast cancer (BC). The results were as follows (the P-value of the chi-square test can be calculated using the URL http://www.quantpsy.org/chisq/chisq.htm):

	Breast Cancer			Helathy controls			
MTHFR C677T GAS	*TT*	*TC*	*CC*	*TT*	*TC*	*CC*	P-value
1	15	23	28	11	21	25	0.70
2	43	162	110	23	92	118	<0.01
3	25	85	80	12	87	94	0.05
4	94	242	178	65	215	215	<0.01
5	32	96	58	17	80	50	0.27
6	7	38	43	24	145	173	0.73
7	343	695	274	196	577	387	<0.01
8	43	140	135	74	196	140	0.01
9	27	141	166	50	259	242	0.14
10	8	91	134	13	104	181	0.66
11	351	786	362	155	509	440	<0.01

	Breast Cancer			Helathy controls			
MTHFR C677T GAS	*TT*	*TC*	*CC*	*TT*	*TC*	*CC*	P-value
1	15	23	28	11	21	25	0.70
2	43	162	110	23	92	118	<0.01
3	25	85	80	12	87	94	0.05
4	94	242	178	65	215	215	<0.01
5	32	96	58	17	80	50	0.27
6	7	38	43	24	145	173	0.73
7	343	695	274	196	577	387	<0.01
8	43	140	135	74	196	140	0.01
9	27	141	166	50	259	242	0.14
10	8	91	134	13	104	181	0.66
11	351	786	362	155	509	440	<0.01

The GAS derived diverse results:

in 6 GAS, the association was significant (P<0.05) and

in 5 GAS, the association was not significant (P≥0.05)

	Breast Cancer			Helathy controls			
MTHFR C677T GAS	*TT*	*TC*	*CC*	*TT*	*TC*	*CC*	P-value
1	15	23	28	11	21	25	0.70
2	43	162	110	23	92	118	<0.01
3	25	85	80	12	87	94	0.05
4	94	242	178	65	215	215	<0.01
5	32	96	58	17	80	50	0.27
6	7	38	43	24	145	173	0.73
7	343	695	274	196	577	387	<0.01
8	43	140	135	74	196	140	0.01
9	27	141	166	50	259	242	0.14
10	8	91	134	13	104	181	0.66
11	351	786	362	155	509	440	<0.01

Thus, based on the current evidence it is hard to draw a safe conclusion regarding the association between MTHFR C677T variant and BC development and we need to provide an overall estimate that shows the magnitude of association.

In this instance a meta-analysis can play a role.

What is a meta-analysis?

Meta-analysis is a technique that synthesizes the results of individual GAS

However, prior to synthesis of results from individual GAS, we must specify the genetic model and the metric (or measure) for expressing the magnitude of association for each GAS.

The magnitude of association is expressed by the odds ratio (OR).

Meta-analysis allows us

i) to estimate the overall (pooled) OR after combining multiple GAS,

ii) to explore the sources of heterogeneity across studies and

iii) to investigate the existence of publication bias.

However, the synthesis of results is not just the simple sum of the data obtained from all GAS but it is a procedure that "weights" the results of each study according to its precision (which is expressed as variance).

Genetic models

In meta-analysis of GAS, we explore various genetic models of genotypes by merging genotypes. These models include:

- recessive model:
homozygous for mt (mt/mt) vs. wt-carriers

- dominant model:
mt-carriers vs. homozygous for wt (wt/wt)

- additive model:
homozygous for mt vs. homozygous for wt

- co-dominant model:
heterozygous (wt/mt) vs. all homozygotes

Recessive model

For each study, we merge the genotypes (mt/mt vs. mt/wt+wt/wt) and then, we calculate the OR with the respective 95% CI.

The data of the 1st GAS is as follows:

MTHFR C677T	Cases-BC	Controls
TT (mt/mt)	15	11
TC (mt/wt)	23	21
CC (wt/wt)	28	25

After merging the genotypes (mt/mt vs. mt/wt+wt/wt) for expressing the recessive model the data of the 1st GAS is as follows:

Outcome	**Cases-BC (T)**	**Control (C)**	***Total***
mt/mt	s_T = **15**	s_C = **11**	*s* = ***26***
wt-carrier (mt/wt+wt/wt)	f_T = **51 (23+28)**	f_C = **46 (21+25)**	*f* = ***97***
Total	n_T = ***66***	n_C = ***57***	*n* = ***123***

Odds ratio, OR

The OR is given by the following formula:

$$OR = \left(\frac{\text{"prob."\ of BC when mt / mt}}{\text{"prob."\ of BC when wt - carrier}} \right) = \left(\frac{\frac{s_T}{s_C}}{\frac{f_T}{f_C}} \right) = \left(\frac{s_T f_C}{s_C f_T} \right)$$

The ln(OR) is: $\theta = \ln\frown OR \backsimeq \ln\left(\frac{s_T f_C}{s_C f_T} \right)$

The variance of θ=ln(OR) is: $\text{var}(\theta) = \frac{1}{s_T} + \frac{1}{s_C} + \frac{1}{f_T} + \frac{1}{f_C}$

The standard error of θ is: $SE(\theta) = \sqrt{\text{var}(\theta)}$

Then, the 95% CI of θ=ln(OR) is

$$\left(\theta - 1.96 * SE,\ \theta + 1.96 * SE\right)$$

In order to estimate the 95% CI of the OR we calculate the anti-log, i.e. we calculate the exponentials of the upper and lower limits (note: $e^{\ln(x)}=x$)

$$\left(e^{\theta-1.96*SE}, e^{\theta+1.96*SE}\right) = \left(e^{\ln(OR)-1.96*SE}, e^{\ln(OR)+1.96*SE}\right)$$

Example- MTHFR C677T and Breast Cancer

Outcome	Cases-BC (T)	Control (C)	*Total*
mt/mt	$s_T = 15$	$s_C = 11$	$s = 26$
wt-carrier (mt/wt+wt/wt)	$f_T = 51$	$f_C = 46$	$f = 97$
Total	$n_T = 66$	$n_C = 57$	$n = 123$

The OR of BC when homozygous mt/mt relative to wt-carrier (mt/wt+wt/wt) is given by:

$$OR = \left(\frac{\text{"prob."of BC when mt/mt}}{\text{"prob."of BC when wt-carrier}} \right) = \left(\frac{\frac{s_T}{s_C}}{\frac{f_T}{f_C}} \right) = \left(\frac{s_T f_C}{s_C f_T} \right) = \left(\frac{15 \times 46}{11 \times 51} \right) = 1.23$$

An OR=1.23 means that there is 23% more chance of BC when mt/mt than when wt-carrier.

Outcome	Cases-BC (T)	Control (C)	*Total*
mt/mt	s_T = **15**	s_C = **11**	*s* = ***26***
wt-carrier (mt/wt+wt/wt)	f_T = **51**	f_C = **46**	*f* = ***97***
Total	n_T = ***66***	n_C = ***57***	*n* = ***123***

$$\theta = \ln(OR) = \ln(1.23) = 0.21$$

$$\text{var}(\theta) = \frac{1}{s_T} + \frac{1}{s_C} + \frac{1}{f_T} + \frac{1}{f_C} = \frac{1}{15} + \frac{1}{11} + \frac{1}{51} + \frac{1}{46} = 0.20$$

$$SE(\theta) = \sqrt{\text{var}(\theta)} = \sqrt{0.20} = 0.45$$

Then, the 95% CI of θ=ln(OR) is

$$(\theta - 1.96 * SE(\theta),\ \theta + 1.96 * SE(\theta)) =$$

$$(0.21 - 1.96 * 0.45,\ 0.21 + 1.96 * 0.45) = (-0.67,\ 1.09)$$

The 95% CI of θ=ln(OR) is $(-0.67, 1.09)$

Then, the 95% CI of OR is $\left(e^{-0.67}, e^{1.09}\right) = (0.51, 2.95)$

Thus, with 95% confidence we can claim that the OR lies from 0.51 to 2.96.

The 95% CI of OR includes "1" and therefore, the OR is not significant at P=0.05.

Recessive model for all GAS

Then, the data of all studies are as follows:

MTHFR C677T GAS	Cases		Controls		RESULTS		
	mt/mt	Total	mt/mt	Total	OR	95%LL	95%UL
1	15	66	11	57	1.23	0.51	2.95
2	43	315	23	233	1.44	0.84	2.47
3	25	190	12	193	2.29	1.11	4.69
4	94	514	65	495	1.48	1.05	2.09
5	32	186	17	147	1.59	0.84	2.99
6	7	88	24	342	1.15	0.48	2.75
7	343	1312	196	1160	1.74	1.43	2.12
8	43	318	74	410	0.71	0.47	1.07
9	27	334	50	551	0.88	0.54	1.44
10	8	233	13	298	0.78	0.32	1.91
11	351	1499	155	1104	1.87	1.52	2.30

MTHFR C677T GAS	Cases		Controls		RESULTS		
	mt/mt	Total	mt/mt	Total	OR	95%LL	95%UL
1	15	66	11	57	1.23	0.51	2.95
2	43	315	23	233	1.44	0.84	2.47
3	25	190	12	193	2.29	1.11	4.69
4	94	514	65	495	1.48	1.05	2.09
5	32	186	17	147	1.59	0.84	2.99
6	7	88	24	342	1.15	0.48	2.75
7	343	1312	196	1160	1.74	1.43	2.12
8	43	318	74	410	0.71	0.47	1.07
9	27	334	50	551	0.88	0.54	1.44
10	8	233	13	298	0.78	0.32	1.91
11	351	1499	155	1104	1.87	1.52	2.30

The GAS derived diverse results:

In 8 GAS, the OR was >1, indicating that an homozygous mt/mt subject has greater chance of developing BC relative to a wt-carrier (mt/wt+wt/wt) subject

MTHFR C677T GAS	Cases		Controls		RESULTS		
	mt/mt	Total	mt/mt	Total	OR	95%LL	95%UL
1	15	66	11	57	1.23	0.51	2.95
2	43	315	23	233	1.44	0.84	2.47
3	25	190	12	193	2.29	1.11	4.69
4	94	514	65	495	1.48	1.05	2.09
5	32	186	17	147	1.59	0.84	2.99
6	7	88	24	342	1.15	0.48	2.75
7	343	1312	196	1160	1.74	1.43	2.12
8	43	318	74	410	0.71	0.47	1.07
9	27	334	50	551	0.88	0.54	1.44
10	8	233	13	298	0.78	0.32	1.91
11	351	1499	155	1104	1.87	1.52	2.30

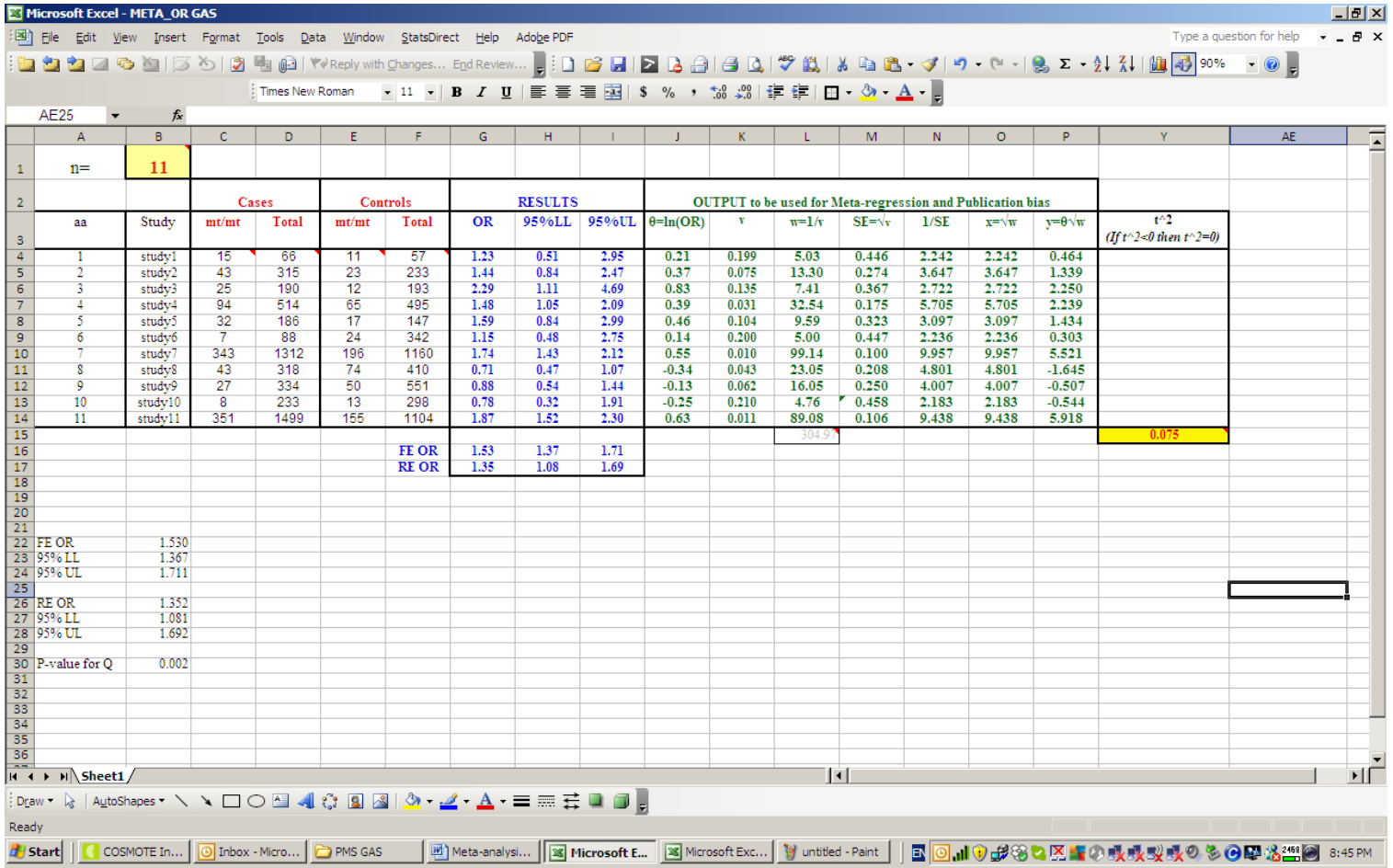
In 3 GAS, the OR was <1, indicating that an homozygous mt/mt subject has less chance of developing BC relative to a wt-carrier (mt/wt+wt/wt) subject

MTHFR C677T GAS	Cases		Controls		RESULTS		
	mt/mt	Total	mt/mt	Total	OR	95%LL	95%UL
1	15	66	11	57	1.23	0.51	2.95
2	43	315	23	233	1.44	0.84	2.47
3	25	190	12	193	2.29	1.11	4.69
4	94	514	65	495	1.48	1.05	2.09
5	32	186	17	147	1.59	0.84	2.99
6	7	88	24	342	1.15	0.48	2.75
7	343	1312	196	1160	1.74	1.43	2.12
8	43	318	74	410	0.71	0.47	1.07
9	27	334	50	551	0.88	0.54	1.44
10	8	233	13	298	0.78	0.32	1.91
11	351	1499	155	1104	1.87	1.52	2.30

In 4 GAS, the OR was significant and in 7 GAS, the OR was non-significant.

Practice in XL

Enter the data of the 11 GAS in XL file META_OR_GAS (yellow color). Then, the OR and the respective 95% CI of each GAS is calculated. Note that this file contains all the results of the subsequent meta-analysis. Also, data to be used in SPSS for analysis later are shown (blue and green color).



Microsoft Excel - META_OR GAS

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Y	AE
1	n=	11																
2			Cases		Controls		RESULTS			OUTPUT to be used for Meta-regression and Publication bias								
3	aa	Study	mt/mt	Total	mt/mt	Total	OR	95%LL	95%UL	θ=ln(OR)	v	w=1/v	SE=√v	1/SE	x=√w	y=θ√w	t^2 (If t^2<0 then t^2=0)	
4	1	study1	15	66	11	57	1.23	0.51	2.95	0.21	0.199	5.03	0.446	2.242	2.242	0.464		
5	2	study2	43	315	23	233	1.44	0.84	2.47	0.37	0.075	13.30	0.274	3.647	3.647	1.339		
6	3	study3	25	190	12	193	2.29	1.11	4.69	0.83	0.135	7.41	0.367	2.722	2.722	2.250		
7	4	study4	94	514	65	495	1.48	1.05	2.09	0.39	0.031	32.54	0.175	5.705	5.705	2.239		
8	5	study5	32	186	17	147	1.59	0.84	2.99	0.46	0.104	9.59	0.323	3.097	3.097	1.434		
9	6	study6	7	88	24	342	1.15	0.48	2.75	0.14	0.200	5.00	0.447	2.236	2.236	0.303		
10	7	study7	343	1312	196	1160	1.74	1.43	2.12	0.55	0.010	99.14	0.100	9.957	9.957	5.521		
11	8	study8	43	318	74	410	0.71	0.47	1.07	-0.34	0.043	23.05	0.208	4.801	4.801	-1.645		
12	9	study9	27	334	50	551	0.88	0.54	1.44	-0.13	0.062	16.05	0.250	4.007	4.007	-0.507		
13	10	study10	8	233	13	298	0.78	0.32	1.91	-0.25	0.210	4.76	0.458	2.183	2.183	-0.544		
14	11	study11	351	1499	155	1104	1.87	1.52	2.30	0.63	0.011	89.08	0.106	9.438	9.438	5.918		
15												304.97					0.075	
16						FE OR	1.53	1.37	1.71									
17						RE OR	1.35	1.08	1.69									
18																		
19																		
20																		
21																		
22	FE OR	1.530																
23	95% LL	1.367																
24	95% UL	1.711																
25																		
26	RE OR	1.352																
27	95% LL	1.081																
28	95% UL	1.692																
29																		
30	P-value for Q	0.002																
31																		
32																		
33																		
34																		
35																		
36																		

Sheet1

Ready

Start | COSMOTE In... | Inbox - Micro... | PMS GAS | Meta-analysi... | Microsoft E... | Microsoft Exc... | untitled - Paint | 8:45 PM

Graphical presentation of data

In meta-analysis, the data are graphical displayed using a CI plot (forest plot).

This plot provide information on the magnitude of the individual study estimates of treatment difference, an indication of the precision of these estimates and a means of assessing consistency amongst the studies.

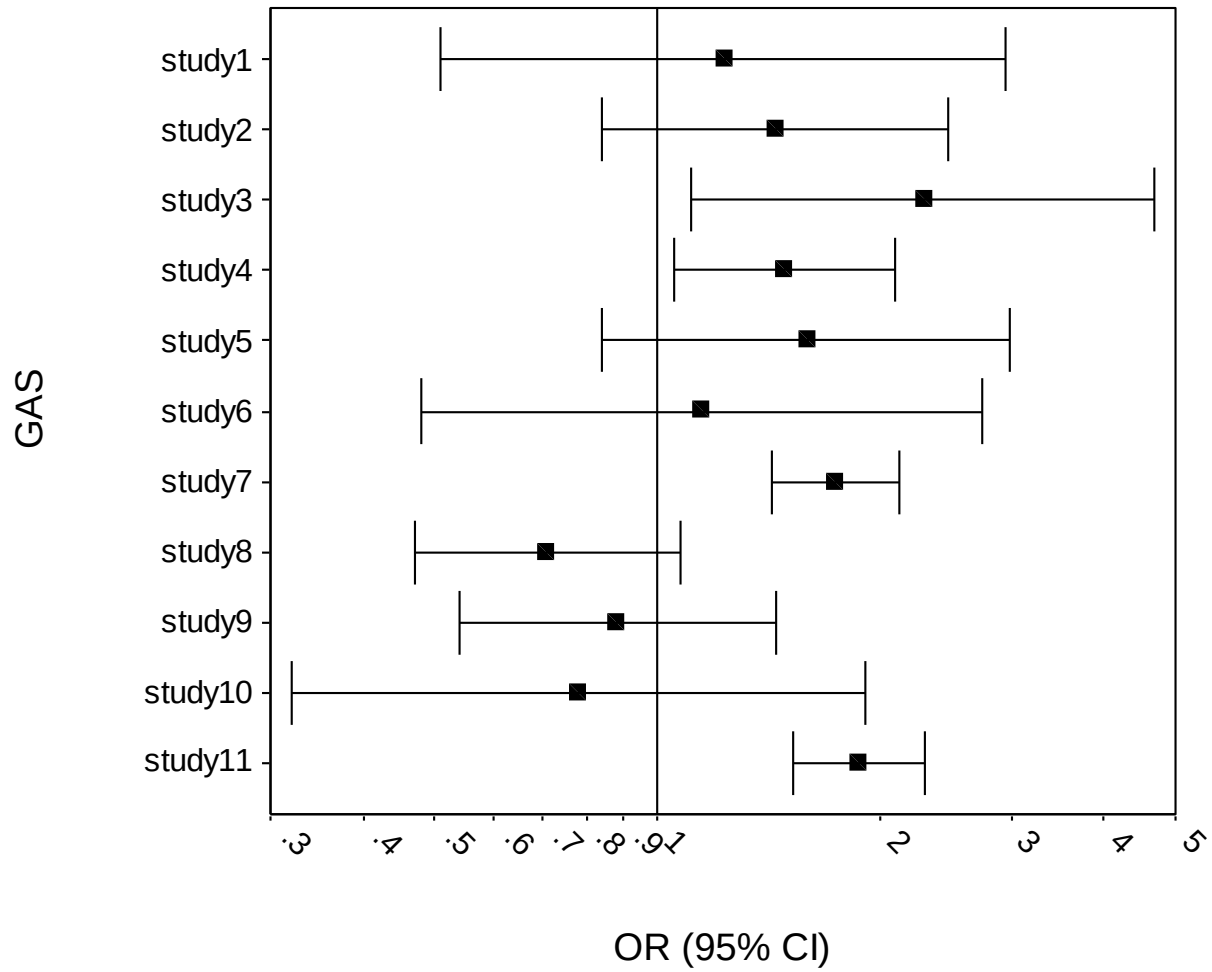
When an overall estimate has been calculated, this can be included.

Example - MTHFR C677T and Breast Cancer

Eleven GAS were conducted to investigate the association between MTHFR C677T variant and BC. The results were as follows:

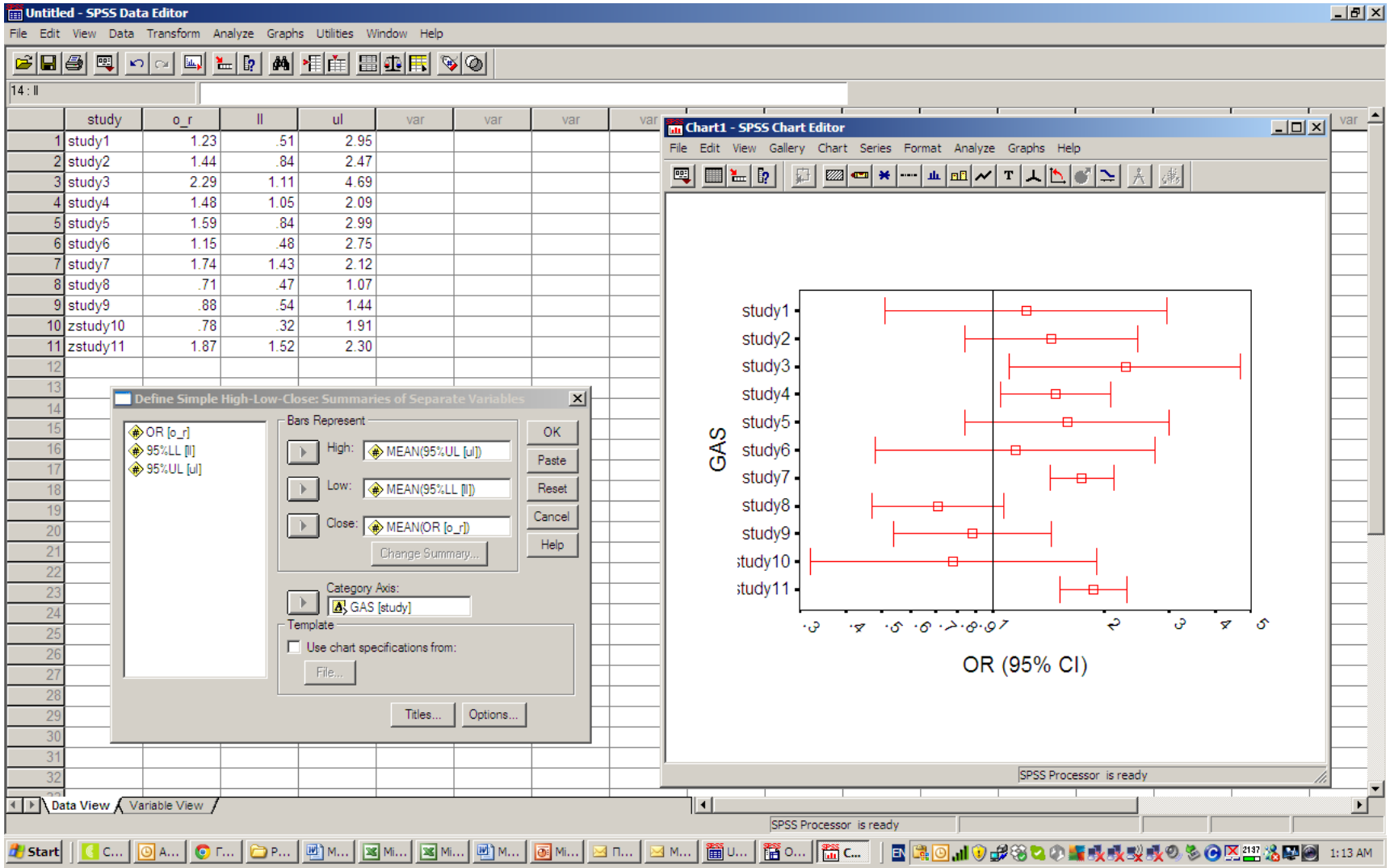
MTHFR C677T GAS	Cases		Controls		RESULTS		
	mt/mt	Total	mt/mt	Total	OR	95%LL	95%UL
1	15	66	11	57	1.23	0.51	2.95
2	43	315	23	233	1.44	0.84	2.47
3	25	190	12	193	2.29	1.11	4.69
4	94	514	65	495	1.48	1.05	2.09
5	32	186	17	147	1.59	0.84	2.99
6	7	88	24	342	1.15	0.48	2.75
7	343	1312	196	1160	1.74	1.43	2.12
8	43	318	74	410	0.71	0.47	1.07
9	27	334	50	551	0.88	0.54	1.44
10	8	233	13	298	0.78	0.32	1.91
11	351	1499	155	1104	1.87	1.52	2.30

In plotting the results, the y-axis depicts the GAS and the x-axis the ORs with the respective 95% CIs, the x-axis is always shown in logarithmic scale for making the visual presentation easy.



Practice in SPSS

Copy-paste the data from XL file META_OR_GAS to SPSS. Then, in SPSS, the graph is contracted as follows:



Combination of estimates of a treatment difference across trials – pooled estimates

In combining the results from multiple GAS, the pooled estimate of OR can be estimated using two approaches: the fixed effects (FE) model and the random effects (RE) model.

The FE model assumes that the GAS are homogeneous in terms of magnitude of association (ie the differences in OR across studies are due to chance).

Then the pooled estimate of OR is given by the weighted average of the ORs of the GAS included in the meta-analysis. The weight is the precision of each GAS (i.e. the variance).

The RE model assumes a genuine diversity in the ORs of various GAS, and it incorporates to the calculations a between study variance.

Hence, when there is significant heterogeneity between GAS, the pooled estimate of the OR is calculating using the RE model.

Heterogeneity

Heterogeneity is a consequence of different populations, sampling strategies and methodological (in genotypying, clinical setting, blindness of the laboratory personnel, etc) diversity across studies.

Fixed effects model

Lets consider the magnitude of association for GAS i, $\theta_i = \ln(OR_i)$.

Then, the pooled estimate of treatment difference $\theta_p = \ln(OR_p)$ (ie the meta-analysis' global outcome or pooled effect) is a weighted mean of θ_i's:

$$\theta_p = \frac{\sum_{i=1}^{n} w_i \theta_i}{\sum_{i=1}^{n} w_i} = \frac{w_1\theta_1 + \ldots + w_n\theta_n}{w_1 + \ldots + w_n}$$

where n is the number of GAS involved in the meta-analysis and w_i is the weight for trial i.

$$\theta_p = \frac{\sum_{i=1}^{n} w_i \theta_i}{\sum_{i=1}^{n} w_i} = \frac{w_1\theta_1 + ... + w_n\theta_n}{w_1 + ... + w_n}$$

The weight of each trial i is given by the inverse of the variance of θ_i (i,e, an estimate of the precision):

$$w_i = \frac{1}{v_i}$$

where v_i the variance of θ_i for the GAS i, $\mathbf{var}(\theta_i) = v_i$

The standard error of θp = ln(ORp) is given by

$$SE(\theta_p) = \sqrt{\frac{1}{\sum_{i=1}^{n} w_i}}$$

The 95% CI for θp = ln(ORp) is given by

$$\left(\theta_p - 1.96 * SE(\theta_p),\ \theta_p + 1.96 * SE(\theta_p)\right)$$

In order to estimate the FE pooled ORp, we calculate the anti-log of θp = ln(ORp),

i.e. we calculate the exponential of θp = ln(ORp) (note: $e^{\ln(x)}=x$)

$$OR_p = e^{\theta_p} = e^{\ln(OR_p)}$$

The 95% CI of the FE ORp, is calculated by taking the anti-log of the limits of the CI,

i.e. we calculate the exponentials of the upper and lower limits (note: $e^{\ln(x)}=x$)

$$\left(e^{\theta_p - 1.96*SE(\theta_p)},\ e^{\theta_p + 1.96*SE(\theta_p)} \right)$$

Example - MTHFR C677T and Breast Cancer

MTHFR C677T GAS	Cases		Controls		RESULTS		
	mt/mt	Total	mt/mt	Total	OR	95%LL	95%UL
1	15	66	11	57	1.23	0.51	2.95
2	43	315	23	233	1.44	0.84	2.47
3	25	190	12	193	2.29	1.11	4.69
4	94	514	65	495	1.48	1.05	2.09
5	32	186	17	147	1.59	0.84	2.99
6	7	88	24	342	1.15	0.48	2.75
7	343	1312	196	1160	1.74	1.43	2.12
8	43	318	74	410	0.71	0.47	1.07
9	27	334	50	551	0.88	0.54	1.44
10	8	233	13	298	0.78	0.32	1.91
11	351	1499	155	1104	1.87	1.52	2.30

The pooled estimate of treatment difference θp = ln(ORp) is

$$\theta_p = \frac{\sum_{i=1}^{8} w_i\theta_i}{\sum_{i=1}^{8} w_i} = \frac{w_1\theta_1 + \ldots + w_{11}\theta_{11}}{w_1 + \ldots + w_{11}} = \frac{5.03\times 0.21 + \ldots + 89.08\times 0.63}{5.03 + \ldots + 89.08} = 0.43$$

Example - MTHFR C677T and Breast Cancer

MTHFR C677T GAS	Cases		Controls		RESULTS		
	mt/mt	Total	mt/mt	Total	OR	95%LL	95%UL
1	15	66	11	57	1.23	0.51	2.95
2	43	315	23	233	1.44	0.84	2.47
3	25	190	12	193	2.29	1.11	4.69
4	94	514	65	495	1.48	1.05	2.09
5	32	186	17	147	1.59	0.84	2.99
6	7	88	24	342	1.15	0.48	2.75
7	343	1312	196	1160	1.74	1.43	2.12
8	43	318	74	410	0.71	0.47	1.07
9	27	334	50	551	0.88	0.54	1.44
10	8	233	13	298	0.78	0.32	1.91
11	351	1499	155	1104	1.87	1.52	2.30

The pooled estimate of treatment difference θp = ln(ORp) is

$$\theta_p = \frac{\sum_{i=1}^{8} w_i\theta_i}{\sum_{i=1}^{8} w_i} = \frac{w_1\theta_1 + \ldots + w_{11}\theta_{11}}{w_1 + \ldots + w_{11}} = \frac{5.03\times 0.21 + \ldots + 89.08\times 0.63}{5.03 + \ldots + 89.08} = 0.43$$

Example - MTHFR C677T and Breast Cancer

MTHFR C677T GAS	Cases		Controls		RESULTS		
	mt/mt	Total	mt/mt	Total	OR	95%LL	95%UL
1	15	66	11	57	1.23	0.51	2.95
2	43	315	23	233	1.44	0.84	2.47
3	25	190	12	193	2.29	1.11	4.69
4	94	514	65	495	1.48	1.05	2.09
5	32	186	17	147	1.59	0.84	2.99
6	7	88	24	342	1.15	0.48	2.75
7	343	1312	196	1160	1.74	1.43	2.12
8	43	318	74	410	0.71	0.47	1.07
9	27	334	50	551	0.88	0.54	1.44
10	8	233	13	298	0.78	0.32	1.91
11	351	1499	155	1104	1.87	1.52	2.30

The SE of θp = ln(ORp) is

$$se(\theta_p \cong \sqrt{\frac{1}{\sum_{i=1}^{11} w_i}} = \sqrt{\frac{1}{w_1 + \ldots + w_{11}}} = \sqrt{\frac{1}{5.03 + \ldots + 89.08}} = 0.057$$

The 95% CI of θp = ln(ORp) is

$$\left(\theta_p - 1.96 se\widehat{\theta_p}, \ \theta_p + 1.96 se\widehat{\theta_p}\right) =$$

$$= \left(0.429 - 1.96 \times 0.057, \ 0.429 + 1.96 \times 0.057\right) =$$

$$= \left(0.32, 0.54\right)$$

The FE pooled ORp is calculated by taking the anti-log of θp = ln(ORp), i.e. we calculate the exponential of θp = ln(ORp) (note: $e^{\ln(x)}=x$):

$$OR_p = e^{\theta_p} = e^{\ln(OR_p)} = e^{0.43} = 1.53$$

In order to estimate the 95% CI of the pooled OR, ORp, we calculate the anti-log of the limits, i.e. we calculate the exponentials of the upper and lower limits of the 95% CI of θp = ln(ORp) (note: $e^{\ln(x)}=x$):

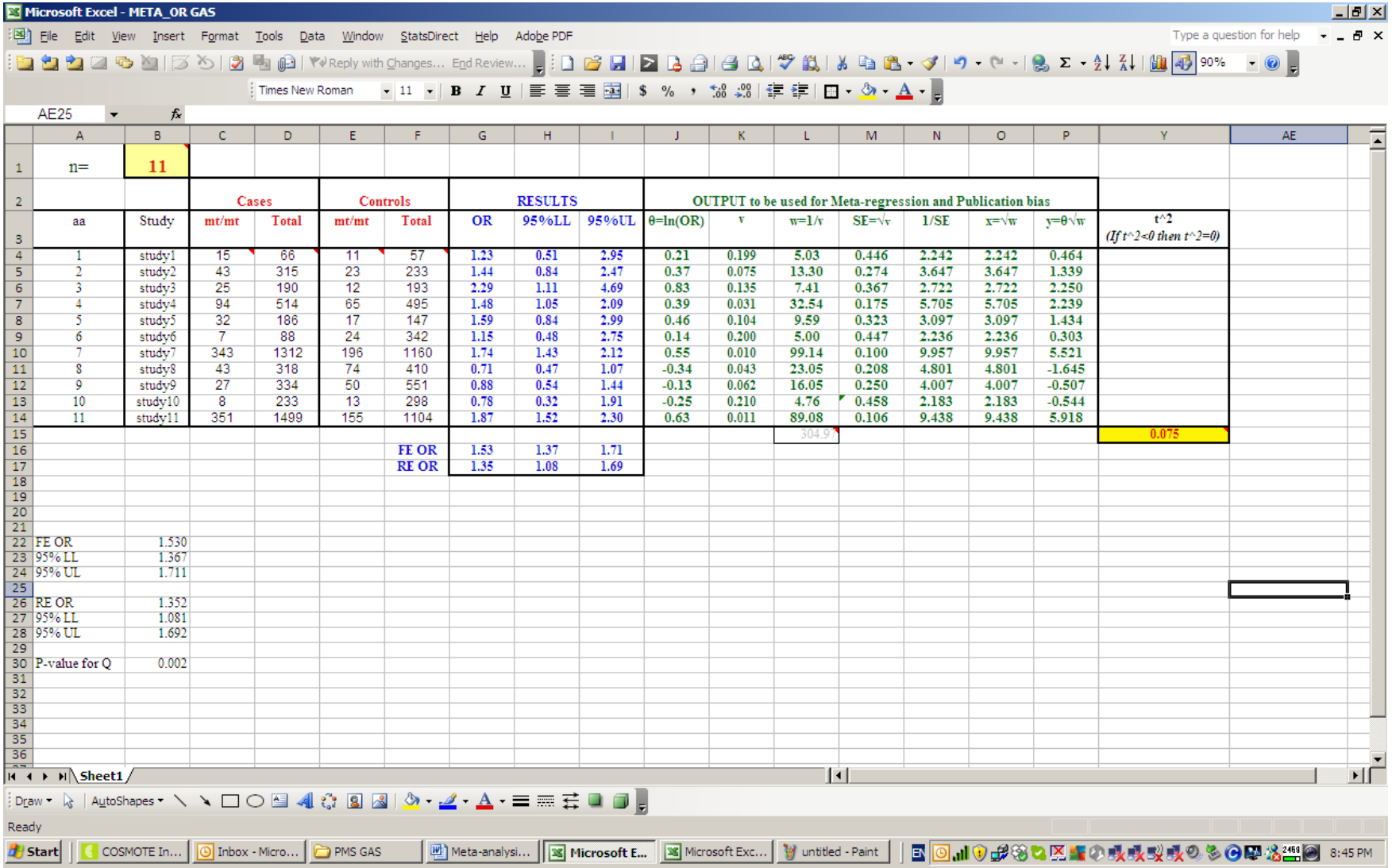
$$\left(e^{\theta_p - 1.96*SE(\theta_p)}, e^{\theta_p + 1.96*SE(\theta_p)}\right) = \left(e^{\ln(OR_p) - 1.96*SE}, e^{\ln(OR_p) + 1.96*SE}\right) =$$

$$= \left(e^{0.32}, e^{0.54}\right) = (1.37, 1.71)$$

The 1 is not included in the 95% CI, thus the FE pooled ORp is significant.

Practice in XL

In XL, the file META_OR _GAS produces the FE ORp and the respective 95% CI.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Y	AE
1	n=	11																
2			Cases		Controls		RESULTS			OUTPUT to be used for Meta-regression and Publication bias								
3	aa	Study	mt/mt	Total	mt/mt	Total	OR	95%LL	95%UL	θ=ln(OR)	v	w=1/v	SE=√v	1/SE	x=√w	y=θ√w	t^2 (If t^2<0 then t^2=0)	
4	1	study1	15	66	11	57	1.23	0.51	2.95	0.21	0.199	5.03	0.446	2.242	2.242	0.464		
5	2	study2	43	315	23	233	1.44	0.84	2.47	0.37	0.075	13.30	0.274	3.647	3.647	1.339		
6	3	study3	25	190	12	193	2.29	1.11	4.69	0.83	0.135	7.41	0.367	2.722	2.722	2.250		
7	4	study4	94	514	65	495	1.48	1.05	2.09	0.39	0.031	32.54	0.175	5.705	5.705	2.239		
8	5	study5	32	186	17	147	1.59	0.84	2.99	0.46	0.104	9.59	0.323	3.097	3.097	1.434		
9	6	study6	7	88	24	342	1.15	0.48	2.75	0.14	0.200	5.00	0.447	2.236	2.236	0.303		
10	7	study7	343	1312	196	1160	1.74	1.43	2.12	0.55	0.010	99.14	0.100	9.957	9.957	5.521		
11	8	study8	43	318	74	410	0.71	0.47	1.07	-0.34	0.043	23.05	0.208	4.801	4.801	-1.645		
12	9	study9	27	334	50	551	0.88	0.54	1.44	-0.13	0.062	16.05	0.250	4.007	4.007	-0.507		
13	10	study10	8	233	13	298	0.78	0.32	1.91	-0.25	0.210	4.76	0.458	2.183	2.183	-0.544		
14	11	study11	351	1499	155	1104	1.87	1.52	2.30	0.63	0.011	89.08	0.106	9.438	9.438	5.918		
15												304.9					0.075	
16						FE OR	1.53	1.37	1.71									
17						RE OR	1.35	1.08	1.69									
18																		
19																		
20																		
21																		
22	FE OR	1.530																
23	95% LL	1.367																
24	95% UL	1.711																
25																		
26	RE OR	1.352																
27	95% LL	1.081																
28	95% UL	1.692																
29																		
30	P-value for Q	0.002																
31																		
32																		
33																		
34																		
35																		
36																		

Testing for heterogeneity

The FE does not consider the variability across studies and assumes that the studies are homogeneous in terms of θi.

Thus, in order to use the FE model, we need first to test whether a significant heterogeneity across studies exists.

If heterogeneity does not exist, then we are eligible to use the FE model; otherwise the RE model should be used.

The test for heterogeneity is based on the following formula (Q-statistic):

$$\mathbf{Q} = \sum_{\mathbf{i}=1}^{\mathbf{n}} \mathbf{w_i} \left(\theta_{\mathbf{i}} - \theta_{\mathbf{p}} \right)^2$$

The Q-statistic is a weighted sum of squares of the deviations of individual θi‘s, θi=ln(ORi), from the pooled estimate θp.

When the θi’s are homogeneous, Q follows a χ^2-distribution with n−1 df.

If Q is less than the 10% point of the χ^2-distribution with n-1 df, there is no significant heterogeneity across studies.

Example - MTHFR C677T and Breast Cancer

MTHFR C677T GAS	Cases		Controls			RESULTS		
	mt/mt	Total	mt/mt	Total	θ =ln(OR)	OR	95%LL	95%UL
1	15	66	11	57	0.21	1.23	0.51	2.95
2	43	315	23	233	0.36	1.44	0.84	2.47
3	25	190	12	193	0.83	2.29	1.11	4.69
4	94	514	65	495	0.39	1.48	1.05	2.09
5	32	186	17	147	0.46	1.59	0.84	2.99
6	7	88	24	342	0.14	1.15	0.48	2.75
7	343	1312	196	1160	0.55	1.74	1.43	2.12
8	43	318	74	410	-0.34	0.71	0.47	1.07
9	27	334	50	551	-0.13	0.88	0.54	1.44
10	8	233	13	298	-0.25	0.78	0.32	1.91
11	351	1499	155	1104	0.63	1.87	1.52	2.30

The pooled estimate of treatment difference is θp = ln(ORp) = 0.43

The heterogeneity Q-statistic is:

$$Q = \sum_{i=1}^{11} w_i \left(\theta_i - \theta_p\right)^2 = w_1 \left(\theta_1 - \theta_p\right)^2 + \ldots + w_{11} \left(\theta_{11} - \theta_p\right)^2 =$$

$$5.03\left(0.21 - 0.43\right)^2 + \ldots + 89.08\left(0.63 - 0.43\right)^2 = 27.87$$

The value Q=27.87 is greater than the 10% point of the χ^2-distribution with n-1=11-1=10 df which is 15.99 (see Table below).

Thus, there is significant heterogeneity across studies (P<0.10).

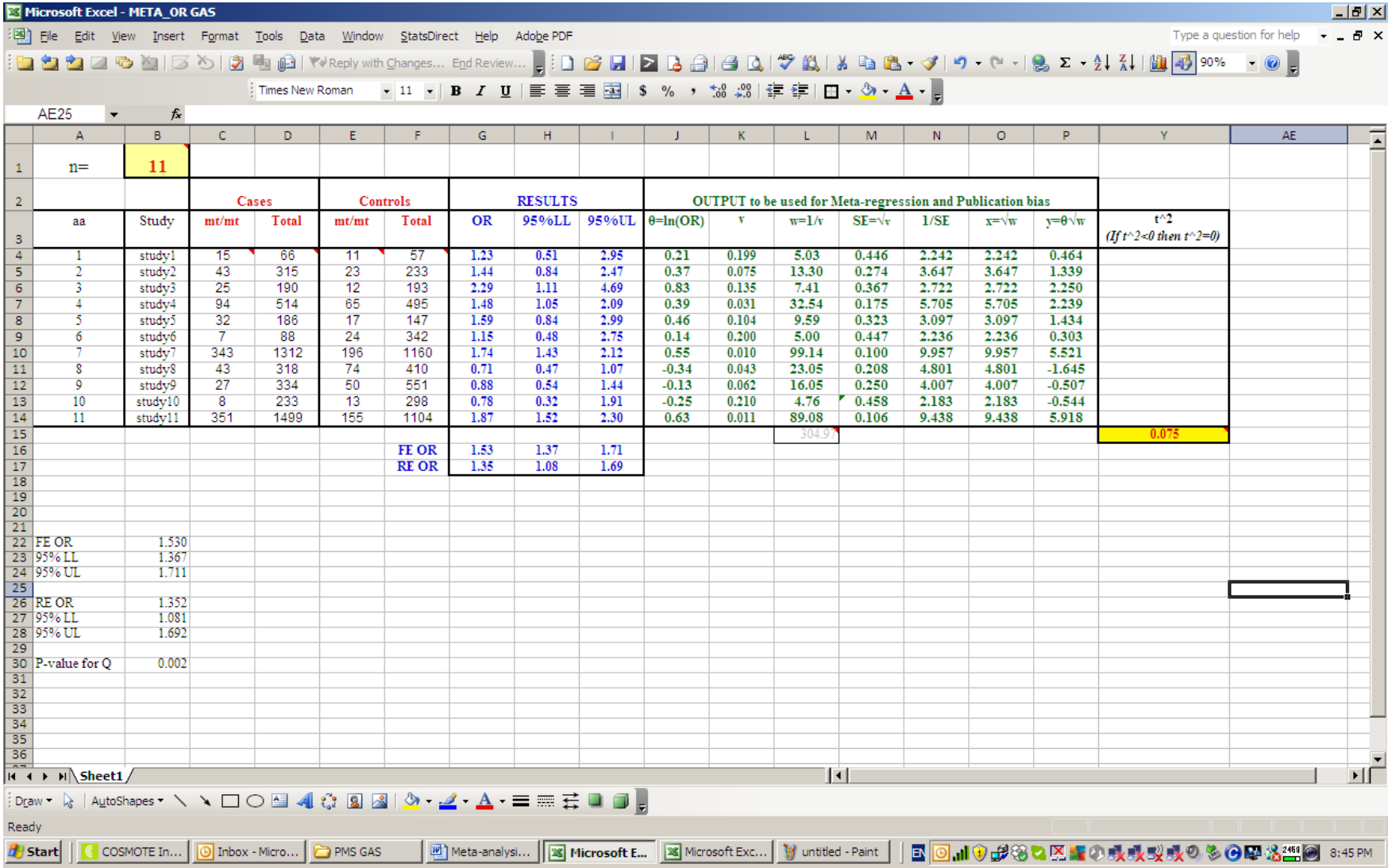
Table. Percentage points of the χ^2 distribution.

	P value			*P* value			*P* value	
d.f.	0.1	0.05	d.f.	0.1	0.05	d.f.	0.1	0.05
1	2.71	3.84	11	17.28	19.68	21	29.62	32.67
2	4.61	5.99	12	18.55	21.03	22	30.81	33.92
3	6.25	7.81	13	19.81	22.36	23	32.01	35.17
4	7.78	9.49	14	21.06	23.68	24	33.20	36.42
5	9.24	11.07	15	22.31	25.00	25	34.38	37.65
6	10.64	12.59	16	23.54	26.30	26	35.56	38.89
7	12.02	14.07	17	24.77	27.59	27	36.74	40.11
8	13.36	15.51	18	25.99	28.87	28	37.92	41.34
9	14.68	16.92	19	27.20	30.14	29	39.09	42.56
10	15.99	18.31	20	28.41	31.41	30	40.26	43.77

Since there is significant heterogeneity across studies the RE model for estimating the pooled OR should be used to draw inferences.

Practice in XL

In XL, the file META_OR_GAS tests for heterogeneity. The exact P-value for Q is P=0.002.



Microsoft Excel - META_OR GAS

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Y	AE
1	n=	11																
2			Cases		Controls		RESULTS			OUTPUT to be used for Meta-regression and Publication bias								
3	aa	Study	mt/mt	Total	mt/mt	Total	OR	95%LL	95%UL	θ=ln(OR)	v	w=1/v	SE=√v	1/SE	x=√w	y=θ√w	t^2 (If t^2<0 then t^2=0)	
4	1	study1	15	66	11	57	1.23	0.51	2.95	0.21	0.199	5.03	0.446	2.242	2.242	0.464		
5	2	study2	43	315	23	233	1.44	0.84	2.47	0.37	0.075	13.30	0.274	3.647	3.647	1.339		
6	3	study3	25	190	12	193	2.29	1.11	4.69	0.83	0.135	7.41	0.367	2.722	2.722	2.250		
7	4	study4	94	514	65	495	1.48	1.05	2.09	0.39	0.031	32.54	0.175	5.705	5.705	2.239		
8	5	study5	32	186	17	147	1.59	0.84	2.99	0.46	0.104	9.59	0.323	3.097	3.097	1.434		
9	6	study6	7	88	24	342	1.15	0.48	2.75	0.14	0.200	5.00	0.447	2.236	2.236	0.303		
10	7	study7	343	1312	196	1160	1.74	1.43	2.12	0.55	0.010	99.14	0.100	9.957	9.957	5.521		
11	8	study8	43	318	74	410	0.71	0.47	1.07	-0.34	0.043	23.05	0.208	4.801	4.801	-1.645		
12	9	study9	27	334	50	551	0.88	0.54	1.44	-0.13	0.062	16.05	0.250	4.007	4.007	-0.507		
13	10	study10	8	233	13	298	0.78	0.32	1.91	-0.25	0.210	4.76	0.458	2.183	2.183	-0.544		
14	11	study11	351	1499	155	1104	1.87	1.52	2.30	0.63	0.011	89.08	0.106	9.438	9.438	5.918		
15												304.9					0.075	
16						FE OR	1.53	1.37	1.71									
17						RE OR	1.35	1.08	1.69									
18																		
19																		
20																		
21																		
22	FE OR	1.530																
23	95% LL	1.367																
24	95% UL	1.711																
25																		
26	RE OR	1.352																
27	95% LL	1.081																
28	95% UL	1.692																
29																		
30	P-value for Q	0.002																
31																		
32																		
33																		
34																		
35																		
36																		

Sheet1

Ready

Start | COSMOTE In... | Inbox - Micro... | PMS GAS | Meta-analysi... | Microsoft E... | Microsoft Exc... | untitled - Paint | 8:45 PM

Random effects model

When there is significant heterogeneity between studies, the pooled estimate of the treatment differences is calculating using the RE model since it incorporates to the calculations a between study variance.

The RE pooled estimate of the treatment difference (ie the meta-analysis' global outcome or pooled effect) is again a weighted mean of θi's:

$$\theta_p^* = \frac{\sum_{i=1}^{n} w_i^* \theta_i}{\sum_{i=1}^{n} w_i^*} = \frac{w_1^* \theta_1 + \ldots + w_n^* \theta_n}{w_1^* + \ldots + w_n^*}$$

where is the weight of study i which is equal to the inverse of the variance (precision) of θi,

$$w_i^* = \frac{1}{\mathrm{var}(\theta_i)}$$

and

$$\mathrm{var}(\theta_i) \cong \left(\frac{1}{w_i} + \tau^2 \right)$$

where τ^2 is the between studies variance (ie in the RE model, the variance of θi incorporates the variance of treatment differences across studies).

The τ^2 is given by the (DerSimonian and Laird) formula:

$$\tau^2 = \frac{Q - (n-1)}{\sum_{i=1}^{n} w_i - \left(\sum_{i=1}^{n} w_i^2 \Big/ \sum_{i=1}^{n} w_i \right)}$$

If $\tau^2<0$ then it is set $\tau 2=0$.

The standard error of $\theta^*_p = \ln(OR^*_p)$ is given by

$$SE(\theta^*_p) = \sqrt{\frac{1}{\sum_{i=1}^{n} w_i^*}}$$

The 95% CI for $\theta^*_p = \ln(OR^*_p)$ is given by

$$\left(\theta^*_p - 1.96 * SE\left(\theta^*_p\right),\ \theta^*_p + 1.96 * SE\left(\theta^*_p\right)\right)$$

In order to estimate the RE pooled , we calculate the anti-log of , i.e. we calculate the exponential of (note: $e^{\ln(x)}=x$)

$$OR_p^* = e^{\theta_p^*} = e^{\ln(OR_p^*)}$$

The 95% CI of the RE OR^*_p, is calculated by taking the anti-log of the limits of the CI for $\theta^*_p=\ln(OR^*_p)$, i.e. we calculate the exponentials of the upper and lower limits:

$$\left(e^{\theta_p^* - 1.96 * SE\left(\theta_p^*\right)}, e^{\theta_p^* + 1.96 * SE\left(\theta_p^*\right)} \right)$$

Example - MTHFR C677T and Breast Cancer

MTHFR C677T GAS	Cases		Controls			RESULTS		
	mt/mt	Total	mt/mt	Total	θ =ln(OR)	v	w=1/v	w*
1	15	66	11	57	0.21	0.199	5.03	3.66
2	43	315	23	233	0.36	0.075	13.30	6.67
3	25	190	12	193	0.83	0.135	7.41	4.77
4	94	514	65	495	0.39	0.031	32.54	9.49
5	32	186	17	147	0.46	0.104	9.59	5.59
6	7	88	24	342	0.14	0.200	5.00	3.64
7	343	1312	196	1160	0.55	0.010	99.14	11.80
8	43	318	74	410	-0.34	0.043	23.05	8.47
9	27	334	50	551	-0.13	0.062	16.05	7.30
10	8	233	13	298	-0.25	0.210	4.76	3.51
11	351	1499	155	1104	0.63	0.011	89.08	11.64

$Q=27.87$

$$\tau^2 = \frac{Q-(n-1)}{\sum_{i=1}^{11} w_i - \sum_{i=1}^{11} w_i^2 \Big/ \sum_{i=1}^{11} w_i} = \frac{Q-(n-1)}{w_1+\ldots+w_{11} - \frac{w_1^2+\ldots+w_{11}^2}{w_1+\ldots+w_{11}}} =$$

$$\frac{27.87-(11-1)}{5.03+\ldots+89.08 - \frac{5.03^2+\ldots+89.08^2}{5.03+\ldots+89.08}} = 0.075$$

Example - MTHFR C677T and Breast Cancer

MTHFR C677T GAS	Cases		Controls			RESULTS		
	mt/mt	Total	mt/mt	Total	θ =ln(OR)	v	w=1/v	w*
1	15	66	11	57	0.21	0.199	5.03	3.66
2	43	315	23	233	0.36	0.075	13.30	6.67
3	25	190	12	193	0.83	0.135	7.41	4.77
4	94	514	65	495	0.39	0.031	32.54	9.49
5	32	186	17	147	0.46	0.104	9.59	5.59
6	7	88	24	342	0.14	0.200	5.00	3.64
7	343	1312	196	1160	0.55	0.010	99.14	11.80
8	43	318	74	410	-0.34	0.043	23.05	8.47
9	27	334	50	551	-0.13	0.062	16.05	7.30
10	8	233	13	298	-0.25	0.210	4.76	3.51
11	351	1499	155	1104	0.63	0.011	89.08	11.64

The weight of study 1 incorporating the across study variability is:

$$w_1^* = \frac{1}{\left(\frac{1}{w_1} + \tau^2\right)} = \frac{1}{\left(\frac{1}{5.03} + 0.075\right)} = 3.66$$

For study 11 the weight is:

$$w_{11}^* = \frac{1}{\left(\frac{1}{w_{11}} + \tau^2\right)} = \frac{1}{\left(\frac{1}{89.08} + 0.075\right)} = 11.64$$

Example - MTHFR C677T and Breast Cancer

MTHFR C677T GAS	Cases		Controls			RESULTS		
	mt/mt	Total	mt/mt	Total	θ =ln(OR)	v	w=1/v	w*
1	15	66	11	57	0.21	0.199	5.03	3.66
2	43	315	23	233	0.36	0.075	13.30	6.67
3	25	190	12	193	0.83	0.135	7.41	4.77
4	94	514	65	495	0.39	0.031	32.54	9.49
5	32	186	17	147	0.46	0.104	9.59	5.59
6	7	88	24	342	0.14	0.200	5.00	3.64
7	343	1312	196	1160	0.55	0.010	99.14	11.80
8	43	318	74	410	-0.34	0.043	23.05	8.47
9	27	334	50	551	-0.13	0.062	16.05	7.30
10	8	233	13	298	-0.25	0.210	4.76	3.51
11	351	1499	155	1104	0.63	0.011	89.08	11.64

The RE pooled estimate of the treatment difference θp* is:

$$\theta_p^* = \frac{\sum_{i=1}^{11} w_i^* \theta_i}{\sum_{i=1}^{11} w_i^*} = \frac{w_1^*\theta_1 + \ldots + w_{11}^*\theta_{11}}{w_1^* + \ldots + w_{11}^*} = \frac{3.66(0.21) \ldots + 11.64(0.63)}{3.66 + \ldots + 11.64} = 0.302$$

Example - MTHFR C677T and Breast Cancer

MTHFR C677T GAS	Cases		Controls			RESULTS		
	mt/mt	Total	mt/mt	Total	θ =ln(OR)	v	w=1/v	w*
1	15	66	11	57	0.21	0.199	5.03	3.66
2	43	315	23	233	0.36	0.075	13.30	6.67
3	25	190	12	193	0.83	0.135	7.41	4.77
4	94	514	65	495	0.39	0.031	32.54	9.49
5	32	186	17	147	0.46	0.104	9.59	5.59
6	7	88	24	342	0.14	0.200	5.00	3.64
7	343	1312	196	1160	0.55	0.010	99.14	11.80
8	43	318	74	410	-0.34	0.043	23.05	8.47
9	27	334	50	551	-0.13	0.062	16.05	7.30
10	8	233	13	298	-0.25	0.210	4.76	3.51
11	351	1499	155	1104	0.63	0.011	89.08	11.64

$$se(\theta_p^* \cong \sqrt{\frac{1}{\sum_{i=1}^{11} w_i^*}} = \sqrt{\frac{1}{w_1^* + \ldots + w_{11}^*}} = \sqrt{\frac{1}{3.66 + \ldots + 11.64}} = 0.114$$

The 95% CI of $\theta^*_p = \ln(OR^*_p)$ is:

$$\left(\theta^*_p - 1.96 se\widehat{\theta^*_p},\ \theta^*_p + 1.96 se\widehat{\theta^*_p}\right) =$$

$$= \left(0.302 - 1.96 \times 0.114,\ 0.302 + 1.96 \times 0.114\right) =$$

$$= \left(0.079,\ 0.525\right)$$

The **RE** pooled is calculated by taking the anti-log of , i.e. we calculate the exponential of (note: $\mathbf{e^{ln(x)}=x}$):

$$\mathbf{OR_p^* = e^{\theta_p^*} = e^{\ln(OR_p^*)} = e^{0.302} = 1.352}$$

In order to estimate the 95% CI of we calculate the anti-log, i.e. we calculate the exponentials of the upper and lower limits:

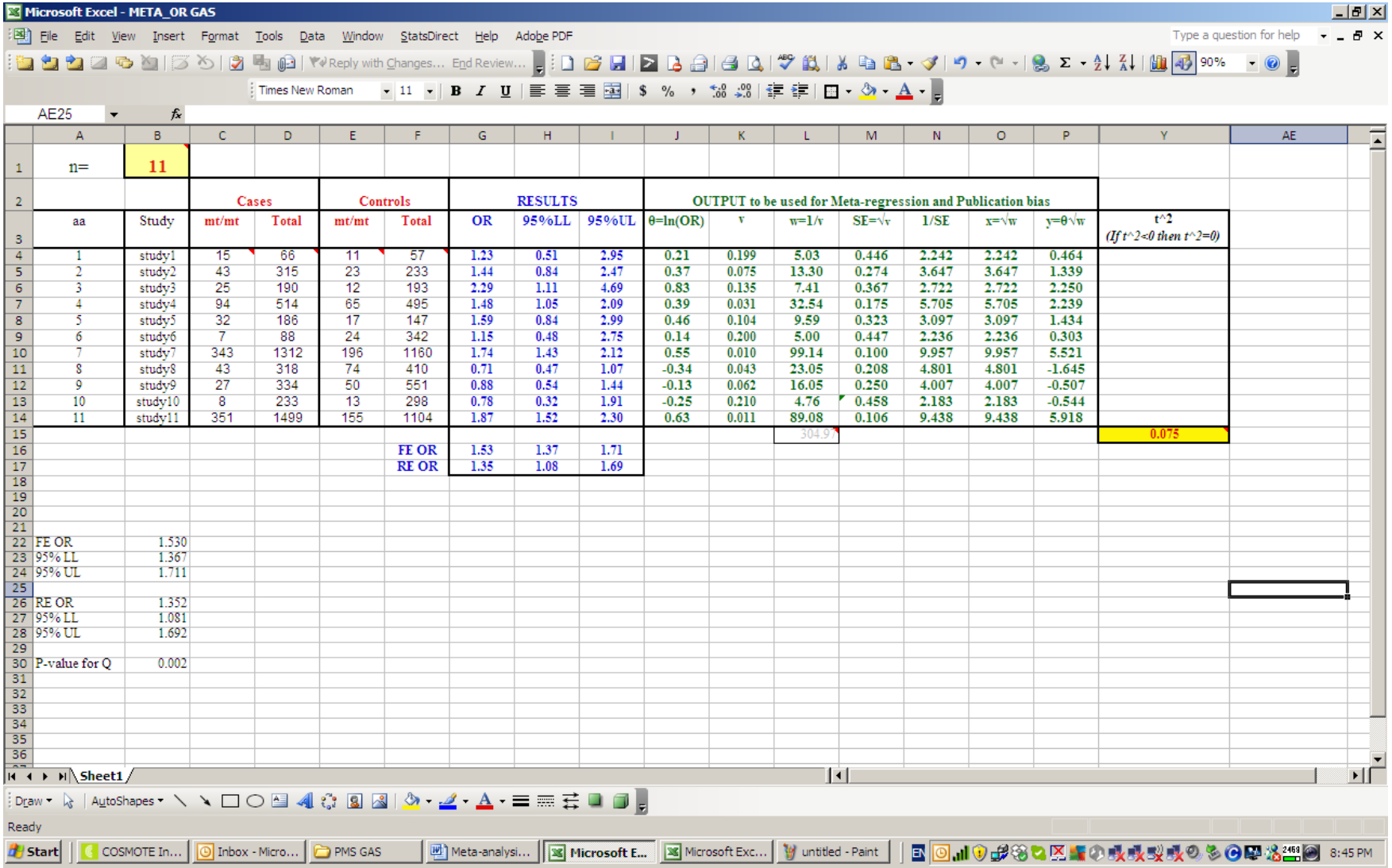
$$\left(\mathbf{e^{\theta_p^* - 1.96*SE}, e^{\theta_p^* + 1.96*SE}}\right) = \left(\mathbf{e^{0.079}, e^{0.525}}\right) = \mathbf{(1.082, 1.691)}$$

The 1 is included in the 95% CI, thus the RE pooled ORp is not significant!

The 95% CI for RE ORp is always wider than the 95% CI for FE OR.

Practice in XL

In XL, the file META_OR_GAS produces the RE ORp and the respective 95% CI.



Microsoft Excel - META_OR GAS

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Y	AE
1	n=	11																
2			Cases		Controls		RESULTS			OUTPUT to be used for Meta-regression and Publication bias								
3	aa	Study	mt/mt	Total	mt/mt	Total	OR	95%LL	95%UL	θ=ln(OR)	v	w=1/v	SE=√v	1/SE	x=√w	y=θ√w	t^2 (If t^2<0 then t^2=0)	
4	1	study1	15	66	11	57	1.23	0.51	2.95	0.21	0.199	5.03	0.446	2.242	2.242	0.464		
5	2	study2	43	315	23	233	1.44	0.84	2.47	0.37	0.075	13.30	0.274	3.647	3.647	1.339		
6	3	study3	25	190	12	193	2.29	1.11	4.69	0.83	0.135	7.41	0.367	2.722	2.722	2.250		
7	4	study4	94	514	65	495	1.48	1.05	2.09	0.39	0.031	32.54	0.175	5.705	5.705	2.239		
8	5	study5	32	186	17	147	1.59	0.84	2.99	0.46	0.104	9.59	0.323	3.097	3.097	1.434		
9	6	study6	7	88	24	342	1.15	0.48	2.75	0.14	0.200	5.00	0.447	2.236	2.236	0.303		
10	7	study7	343	1312	196	1160	1.74	1.43	2.12	0.55	0.010	99.14	0.100	9.957	9.957	5.521		
11	8	study8	43	318	74	410	0.71	0.47	1.07	-0.34	0.043	23.05	0.208	4.801	4.801	-1.645		
12	9	study9	27	334	50	551	0.88	0.54	1.44	-0.13	0.062	16.05	0.250	4.007	4.007	-0.507		
13	10	study10	8	233	13	298	0.78	0.32	1.91	-0.25	0.210	4.76	0.458	2.183	2.183	-0.544		
14	11	study11	351	1499	155	1104	1.87	1.52	2.30	0.63	0.011	89.08	0.106	9.438	9.438	5.918		
15												304.9					0.075	
16						FE OR	1.53	1.37	1.71									
17						RE OR	1.35	1.08	1.69									
18																		
19																		
20																		
21																		
22	FE OR	1.530																
23	95% LL	1.367																
24	95% UL	1.711																
25																		
26	RE OR	1.352																
27	95% LL	1.081																
28	95% UL	1.692																
29																		
30	P-value for Q	0.002																
31																		
32																		
33																		
34																		
35																		
36																		

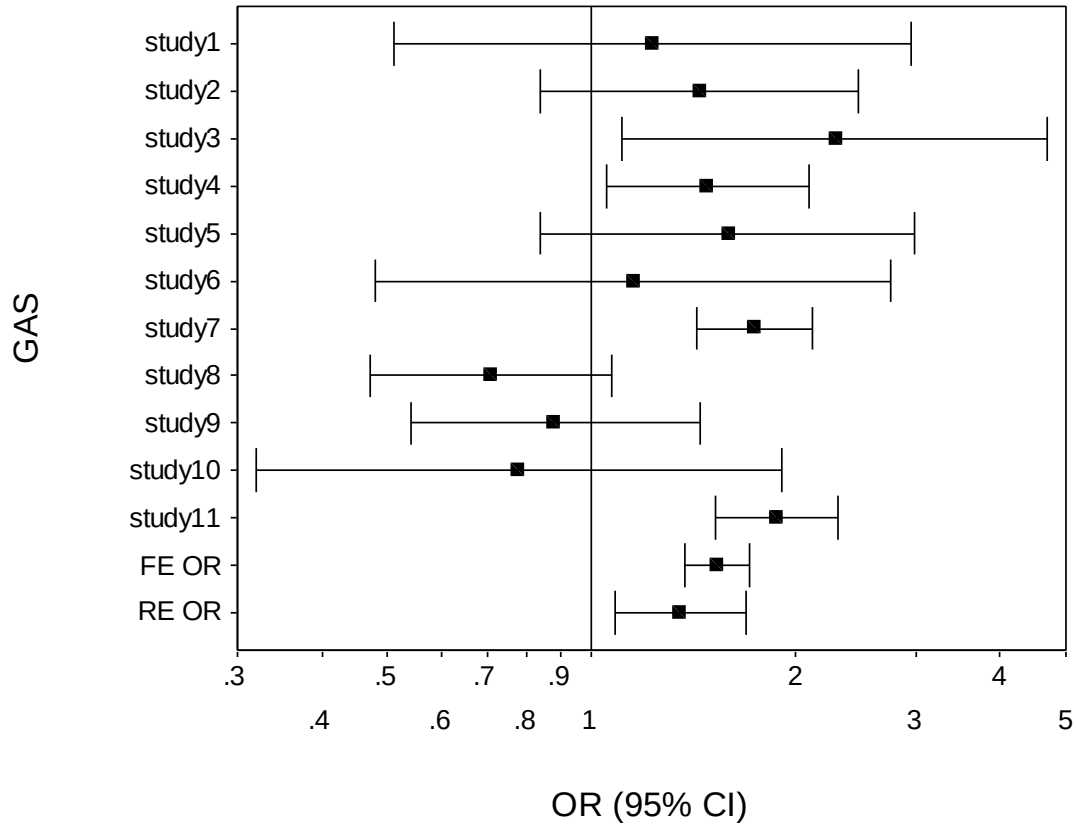
Sheet1

Ready

Start | COSMOTE In... | Inbox - Micro... | PMS GAS | Meta-analysi... | Microsoft E... | Microsoft Exc... | untitled - Paint | 8:45 PM

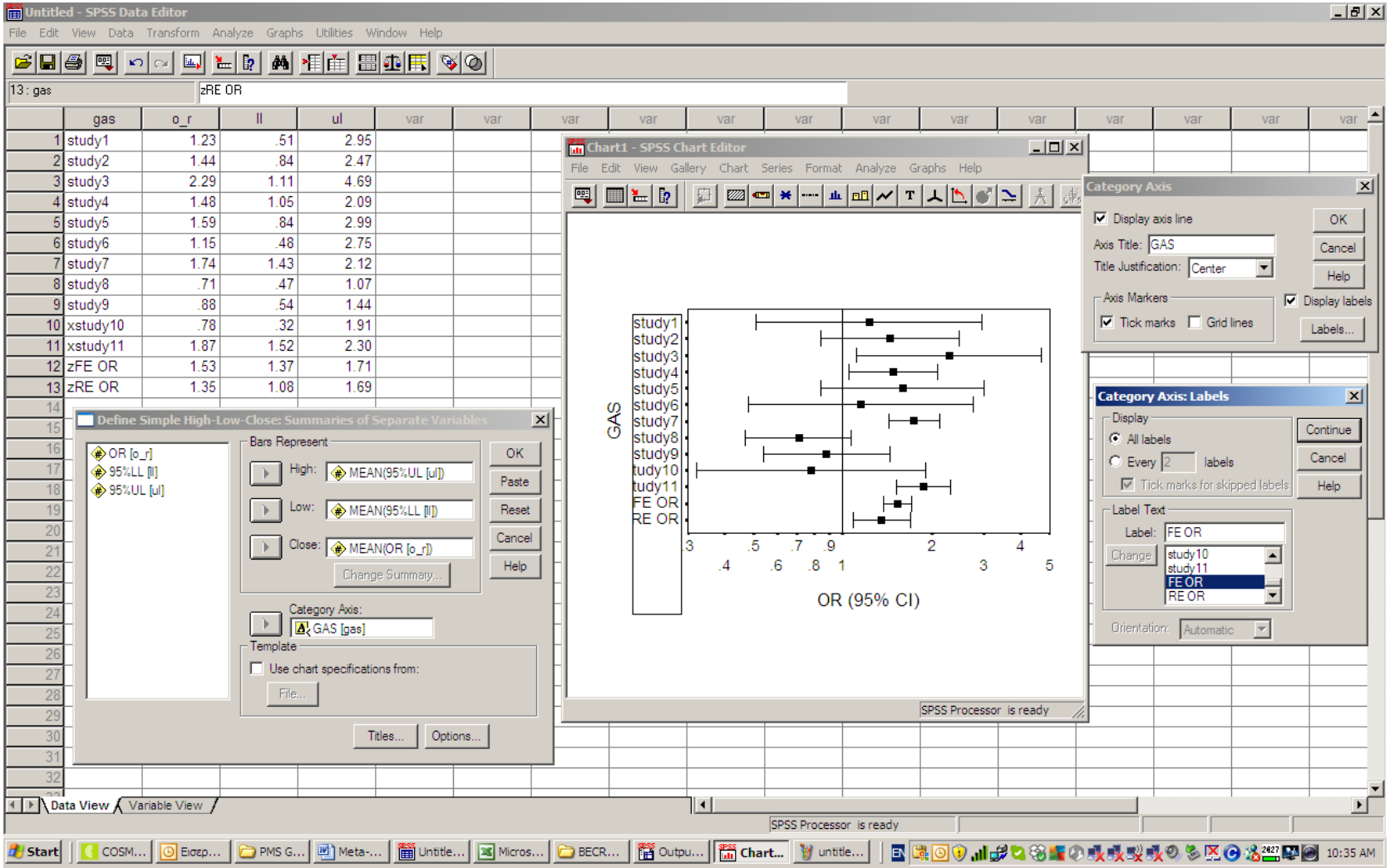
Example - MTHFR C677T and Breast Cancer

In plotting the results, the y-axis depicts the studies and the x-axis the ORs with the respective 95% CIs, the x-axis is always shown in logarithmic scale for making the visual presentation easy. The FE and RE pooled estimates are also shown.



Practice in SPSS

Copy-paste the data from XL file META_OR to SPSS. Then, in SPSS, the graph is contracted as follows:



Practice

A GAS investigating the association between the alleles ADH2*1 (mt) and ADH2*2 (wt) with alcoholism produced the following genotype distributions:

			ADH2					
			*1/*1		*1/*2		*2/*2	
Year	Author	Racial	alcoholic	healthy	alcoholic	healthy	alcoholic	healthy
1993	Sherman	Caucasian	7	18	19	3	19	2
1994	Muramatsu	Chinese	13	12	8	43	11	50
1994	Thomasson	Taiwan	3	1	28	10	63	54
1995	Maezawa	Japanese	30	2	28	22	38	36
1996	Chen	Taiwan	14	0	15	19	17	44
1996	Higuchi	Japanese	204	33	224	160	227	268
1997	Espinós	Caucasian	62	58	9	12	0	1
1999	Chen CC	Chinese	130	43	106	205	104	297
2000	Chao	Taiwan	51	17	129	102	101	122
2001	Lee	Korean	10	6	32	18	64	40
2001	Ogurtsov	Caucasian	56	15	51	29	3	6

Perform a full meta-analysis of the GAS.

Dealing with heterogeneity – Subgroup analysis

Heterogeneity can be attributed to various characteristics of the individual GAS included in the meta-analysis (such race, clinical settings, study quality, etc).

One way to deal with heterogeneity is to perform a subgroup analysis by each characteristic and to explore which characteristic contributes to study heterogeneity.

Example - MTHFR C677T and Breast Cancer

			Cases		Controls					
i		Population	mt/mt	Total	mt/mt	Total	OR_i	$\theta_i=\ln(OR_i)$	v_i	$w_i=1/v_i$
1	study1	Whites	15	66	11	57	1.23	0.21	0.199	5.03
2	study2	Whites	43	315	23	233	1.44	0.37	0.075	13.30
3	study3	E. Asians	25	190	12	193	2.29	0.83	0.135	7.41
4	study4	E. Asians	94	514	65	495	1.48	0.39	0.031	32.54
5	study5	E. Asians	32	186	17	147	1.59	0.46	0.104	9.59
6	study6	Whites	7	88	24	342	1.15	0.14	0.200	5.00
7	study7	E. Asians	343	1312	196	1160	1.74	0.55	0.010	99.14
8	study8	Whites	43	318	74	410	0.71	-0.34	0.043	23.05
9	study9	Whites	27	334	50	551	0.88	-0.13	0.062	16.05
10	study10	Whites	8	233	13	298	0.78	-0.25	0.210	4.76
11	study11	E. Asians	351	1499	155	1104	1.87	0.63	0.011	89.08

However, in studies 3, 4, 5, 7 and 11 the origin of the population was East Asians and in the rest studies the population was Whites.

We would like to explore whether "race" contributes to study heterogeneity.

1) Subgroup analysis for East Asians

For East Asians, the pooled OR is

FE ORp=1.76 with 95% CI (1.55, 2.00)

The P-value of Q (heterogeneity test) is PQ=0.75.

Thus, there is no significant heterogeneity across studies (since PQ≥0.10) and only the FE OR will be considered (note that RE OR coincides with the FE OR).

2) Subgroup analysis for Whites

For Whites, the pooled OR is

FE ORp=0.94 with 95% CI (0.74, 1.19)

The P-value of Q (heterogeneity test) is PQ=0.41.

Thus, there is no significant heterogeneity across studies (since PQ≥0.10) and only the FE OR will be considered.

The subgroup analysis for East Asians produced absolutely different results from the Whites

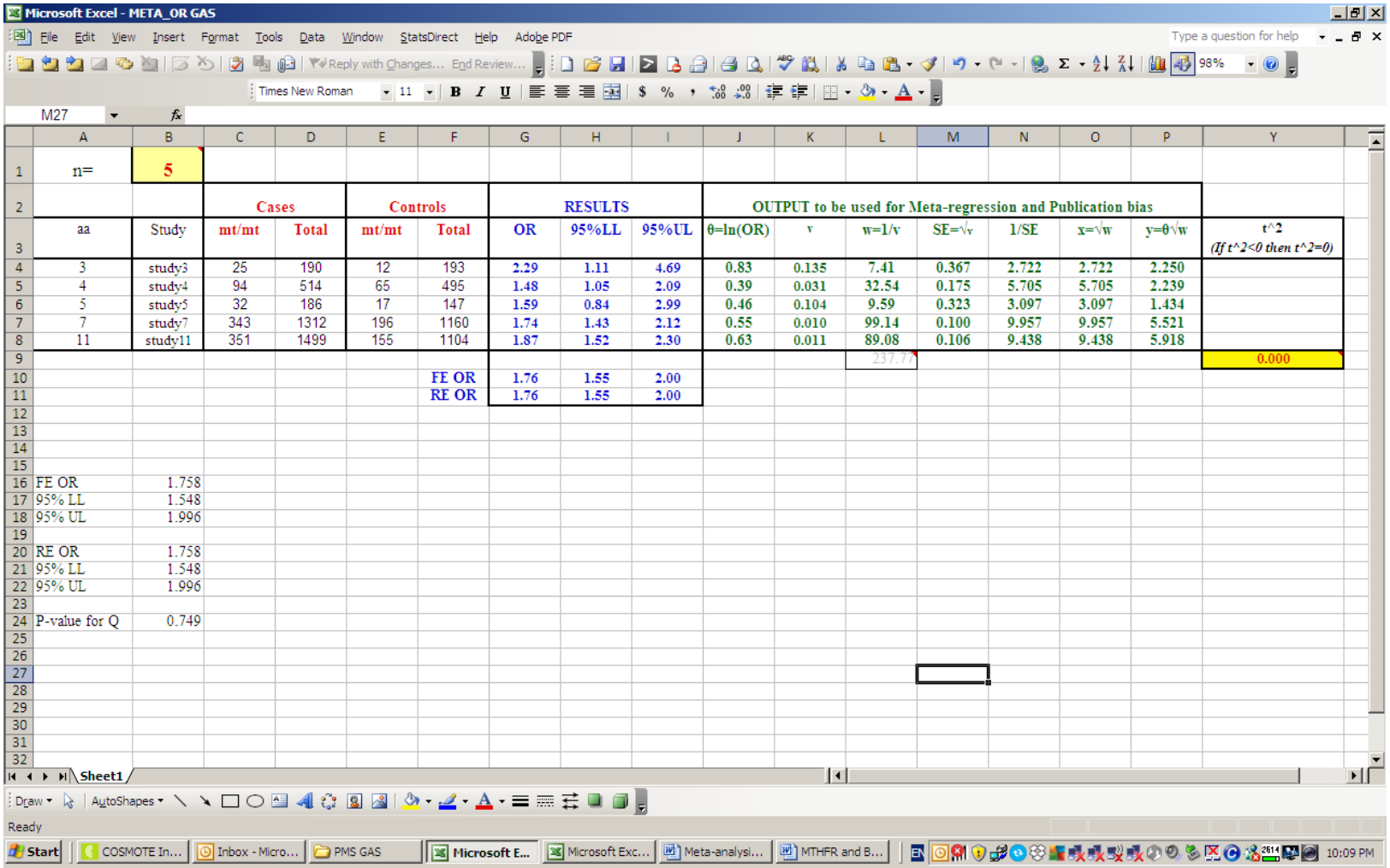
E. Asians ORp=1.76 (1.55, 2.00) and PQ=0.75

Whites vs. ORp=0.94 (0.74, 1.19) and PQ=0.41

This diversity in ORp and the lack of heterogeiety in subgroup analyses implies that “race” has a significant effect in the overall meta-analysis and contributes in the heterogeneity across studies.

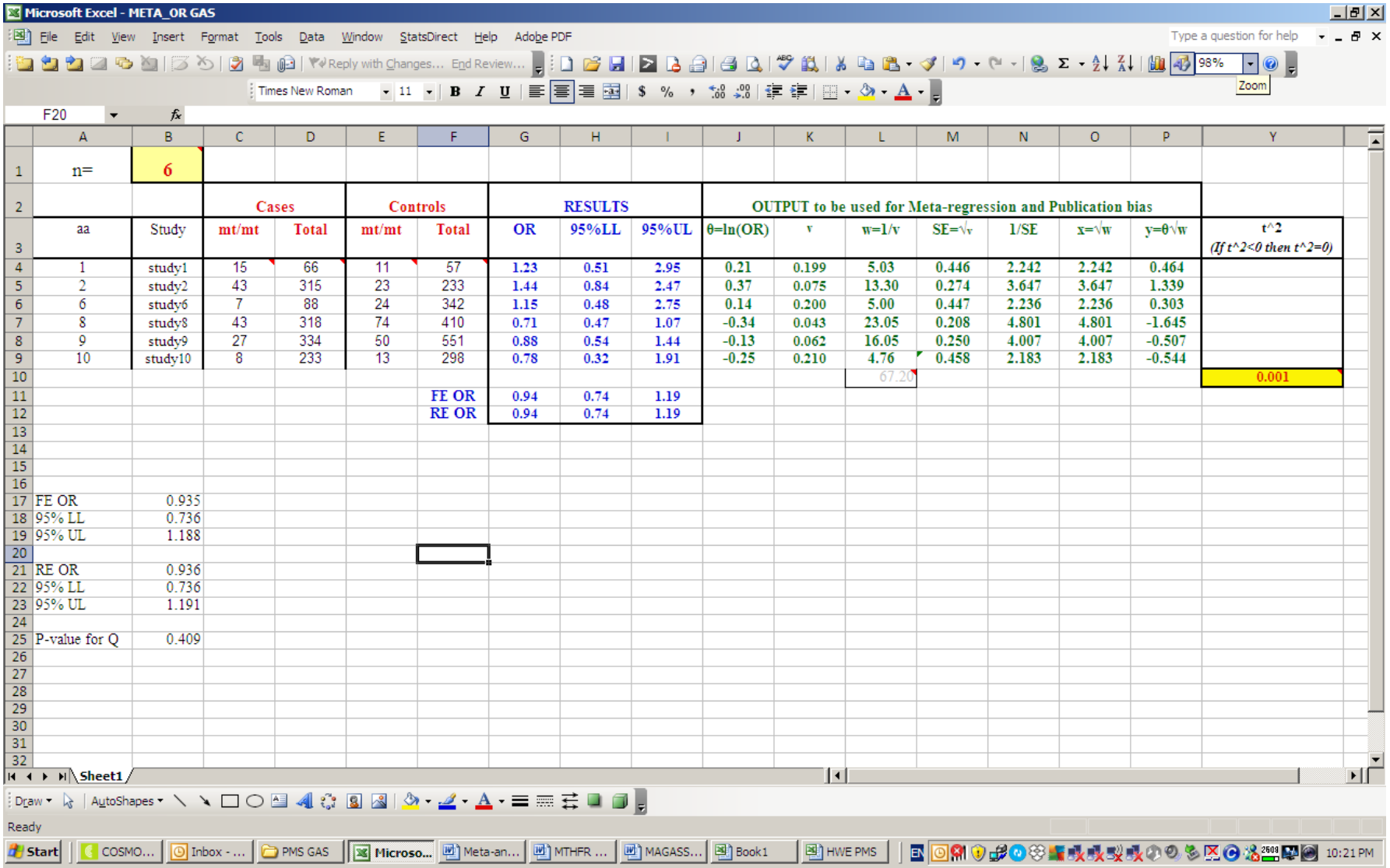
Practice in XL

In XL, the file META_OR_GAS produces the FE ORp and RE ORp with the respective 95% CIs for each subgroup (for analyzing the East Asians, just delete the respective rows with the Whites and change the n from 11 to 5). Close the file without saving it.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Y
1	n=	5															
2			Cases		Controls		RESULTS			OUTPUT to be used for Meta-regression and Publication bias							
3	aa	Study	mt/mt	Total	mt/mt	Total	OR	95%LL	95%UL	θ=ln(OR)	v	w=1/v	SE=√v	1/SE	x=√w	y=θ√w	t^2 (If t^2<0 then t^2=0)
4	3	study3	25	190	12	193	2.29	1.11	4.69	0.83	0.135	7.41	0.367	2.722	2.722	2.250	
5	4	study4	94	514	65	495	1.48	1.05	2.09	0.39	0.031	32.54	0.175	5.705	5.705	2.239	
6	5	study5	32	186	17	147	1.59	0.84	2.99	0.46	0.104	9.59	0.323	3.097	3.097	1.434	
7	7	study7	343	1312	196	1160	1.74	1.43	2.12	0.55	0.010	99.14	0.100	9.957	9.957	5.521	
8	11	study11	351	1499	155	1104	1.87	1.52	2.30	0.63	0.011	89.08	0.106	9.438	9.438	5.918	
9												237.77					0.000
10						FE OR	1.76	1.55	2.00								
11						RE OR	1.76	1.55	2.00								
12																	
13																	
14																	
15																	
16	FE OR	1.758															
17	95% LL	1.548															
18	95% UL	1.996															
19																	
20	RE OR	1.758															
21	95% LL	1.548															
22	95% UL	1.996															
23																	
24	P-value for Q	0.749															
25																	
26																	
27																	
28																	
29																	
30																	
31																	
32																	

For analyzing the Whites, open the file META_OR_GAS and just delete the respective rows with the East Asians and change the n from 11 to 6. Close the file without saving it.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Y
1	n=	6															
2			Cases		Controls		RESULTS			OUTPUT to be used for Meta-regression and Publication bias							
3	aa	Study	mt/mt	Total	mt/mt	Total	OR	95%LL	95%UL	θ=ln(OR)	v	w=1/v	SE=√v	1/SE	x=√w	y=θ√w	t^2 (If t^2<0 then t^2=0)
4	1	study1	15	66	11	57	1.23	0.51	2.95	0.21	0.199	5.03	0.446	2.242	2.242	0.464	
5	2	study2	43	315	23	233	1.44	0.84	2.47	0.37	0.075	13.30	0.274	3.647	3.647	1.339	
6	6	study6	7	88	24	342	1.15	0.48	2.75	0.14	0.200	5.00	0.447	2.236	2.236	0.303	
7	8	study8	43	318	74	410	0.71	0.47	1.07	-0.34	0.043	23.05	0.208	4.801	4.801	-1.645	
8	9	study9	27	334	50	551	0.88	0.54	1.44	-0.13	0.062	16.05	0.250	4.007	4.007	-0.507	
9	10	study10	8	233	13	298	0.78	0.32	1.91	-0.25	0.210	4.76	0.458	2.183	2.183	-0.544	
10												67.20					0.001
11						FE OR	0.94	0.74	1.19								
12						RE OR	0.94	0.74	1.19								
13																	
14																	
15																	
16																	
17	FE OR	0.935															
18	95% LL	0.736															
19	95% UL	1.188															
20																	
21	RE OR	0.936															
22	95% LL	0.736															
23	95% UL	1.191															
24																	
25	P-value for Q	0.409															
26																	
27																	
28																	
29																	
30																	
31																	
32																	

Practice

A GAS investigating the association between the alleles ADH2*1 (mt) and ADH2*2 (wt) with alcoholism produced the following genotype distributions:

			ADH2					
			*1/*1		*1/*2		*2/*2	
Year	Author	Racial	alcoholic	healthy	alcoholic	healthy	alcoholic	healthy
1993	Sherman	Caucasian	7	18	19	3	19	2
1994	Muramatsu	Chinese	13	12	8	43	11	50
1994	Thomasson	Taiwan	3	1	28	10	63	54
1995	Maezawa	Japanese	30	2	28	22	38	36
1996	Chen	Taiwan	14	0	15	19	17	44
1996	Higuchi	Japanese	204	33	224	160	227	268
1997	Espinós	Caucasian	62	58	9	12	0	1
1999	Chen CC	Chinese	130	43	106	205	104	297
2000	Chao	Taiwan	51	17	129	102	101	122
2001	Lee	Korean	10	6	32	18	64	40
2001	Ogurtsov	Caucasian	56	15	51	29	3	6

Perform subgroup analysis by “race”.

Sensitivity analysis for studies of poor quality

The sensitivity analysis examines the effect of excluding specific studies from the meta-analysis, ie examines the impact of excluding these studies in the pooled estimate of OR and in heterogeneity.

Sensitivity analysis for studies of poor quality

Inspection of whether genotype frequencies of controls (disease-free subjects) conform to Hardy-Weinberg equilibrium (HWE) provides an indication of the quality in the design and conduct of GAS.

Departures from HWE can be due to:

- **genotyping errors,**
- **population stratification (population stratification includes differences between groups of ethnic origin or differences between groups of similar ethnic origin but with a limited admixture) and**
- **selection bias in the recruitment of controls**

Example - MTHFR C677T and Breast Cancer

		Cases		Controls					
i		mt/mt	Total	mt/mt	Total	OR_i	$\theta_i=\ln(OR_i)$	v_i	$w_i=1/v_i$
1	study1	15	66	11	57	1.23	0.21	0.199	5.03
2	study2	43	315	23	233	1.44	0.37	0.075	13.30
3	study3	25	190	12	193	2.29	0.83	0.135	7.41
4	study4	94	514	65	495	1.48	0.39	0.031	32.54
5	study5	32	186	17	147	1.59	0.46	0.104	9.59
6	study6	7	88	24	342	1.15	0.14	0.200	5.00
7	study7	343	1312	196	1160	1.74	0.55	0.010	99.14
8	study8	43	318	74	410	0.71	-0.34	0.043	23.05
9	study9	27	334	50	551	0.88	-0.13	0.062	16.05
10	study10	8	233	13	298	0.78	-0.25	0.210	4.76
11	study11	351	1499	155	1104	1.87	0.63	0.011	89.08

Only in study 5, the HWE was marginally significant (P=0.07) and thus a sensitivity analysis was performed for this study.

In Sensitivity analysis,

PQ=0.001.

RE ORp=1.33 with 95% CI (1.05, 1.69).

In the full analysis,

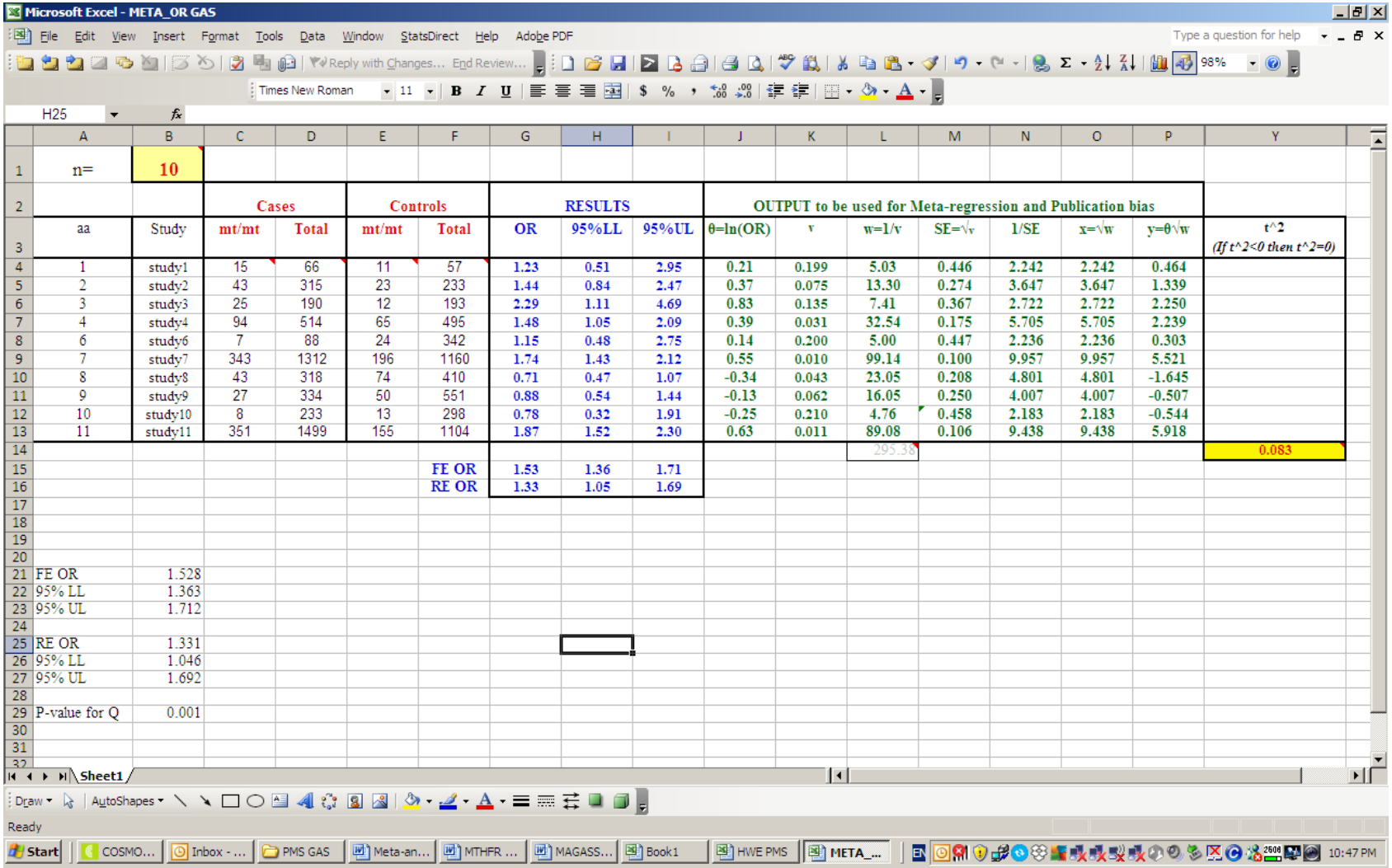
PQ=0.002 which was significant (PQ<0.10).

RE ORp=1.35 with 95% CI (1.08, 1.69)

Thus, exclusion of the study not in HWE does not change the pattern of results of the full analysis.

Practice in XL

In XL, the file META_OR_GAS produces the FE ORp and RE ORp with the respective 95% CIs for sensitivity analysis (just delete the row corresponding to Study 5 and change the n from 11 to 10).



Microsoft Excel - META_OR GAS

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Y
1	n=	10															
2			Cases		Controls		RESULTS			OUTPUT to be used for Meta-regression and Publication bias							
3	aa	Study	mt/mt	Total	mt/mt	Total	OR	95%LL	95%UL	θ=ln(OR)	v	w=1/v	SE=√v	1/SE	x=√w	y=θ√w	t^2 (If t^2<0 then t^2=0)
4	1	study1	15	66	11	57	1.23	0.51	2.95	0.21	0.199	5.03	0.446	2.242	2.242	0.464	
5	2	study2	43	315	23	233	1.44	0.84	2.47	0.37	0.075	13.30	0.274	3.647	3.647	1.339	
6	3	study3	25	190	12	193	2.29	1.11	4.69	0.83	0.135	7.41	0.367	2.722	2.722	2.250	
7	4	study4	94	514	65	495	1.48	1.05	2.09	0.39	0.031	32.54	0.175	5.705	5.705	2.239	
8	6	study6	7	88	24	342	1.15	0.48	2.75	0.14	0.200	5.00	0.447	2.236	2.236	0.303	
9	7	study7	343	1312	196	1160	1.74	1.43	2.12	0.55	0.010	99.14	0.100	9.957	9.957	5.521	
10	8	study8	43	318	74	410	0.71	0.47	1.07	-0.34	0.043	23.05	0.208	4.801	4.801	-1.645	
11	9	study9	27	334	50	551	0.88	0.54	1.44	-0.13	0.062	16.05	0.250	4.007	4.007	-0.507	
12	10	study10	8	233	13	298	0.78	0.32	1.91	-0.25	0.210	4.76	0.458	2.183	2.183	-0.544	
13	11	study11	351	1499	155	1104	1.87	1.52	2.30	0.63	0.011	89.08	0.106	9.438	9.438	5.918	
14												295.38					0.083
15						FE OR	1.53	1.36	1.71								
16						RE OR	1.33	1.05	1.69								
17																	
18																	
19																	
20																	
21	FE OR	1.528															
22	95% LL	1.363															
23	95% UL	1.712															
24																	
25	RE OR	1.331															
26	95% LL	1.046															
27	95% UL	1.692															
28																	
29	P-value for Q	0.001															
30																	
31																	
32																	

Sheet1

Ready

Start | COSMO... | Inbox - ... | PMS GAS | Meta-an... | MTHFR ... | MAGASS... | Book1 | HWE PMS | META_... | 10:47 PM

Practice

A GAS investigating the association between the alleles ADH2*1 (mt) and ADH2*2 (wt) with alcoholism produced the following genotype distributions:

			ADH2					
			*1/*1		*1/*2		*2/*2	
Year	Author	Racial	alcoholic	healthy	alcoholic	healthy	alcoholic	healthy
1993	Sherman	Caucasian	7	18	19	3	19	2
1994	Muramatsu	Chinese	13	12	8	43	11	50
1994	Thomasson	Taiwan	3	1	28	10	63	54
1995	Maezawa	Japanese	30	2	28	22	38	36
1996	Chen	Taiwan	14	0	15	19	17	44
1996	Higuchi	Japanese	204	33	224	160	227	268
1997	Espinós	Caucasian	62	58	9	12	0	1
1999	Chen CC	Chinese	130	43	106	205	104	297
2000	Chao	Taiwan	51	17	129	102	101	122
2001	Lee	Korean	10	6	32	18	64	40
2001	Ogurtsov	Caucasian	56	15	51	29	3	6

Perform a sensitivity analysis for GAS with the controls not in HWE.

Sensitivity analysis for big studies

If we are interested in examining the impact of the biggest studies in the meta-analysis results, then we perform a sensitivity analysis, i.e. we exclude these studies from the meta-analysis and we re-calculated the pooled estimate of the treatment difference and the heterogeneity test.

Example - MTHFR C677T and Breast Cancer

		Cases		Controls					
i		mt/mt	Total	mt/mt	Total	OR_i	$\theta_i=\ln(OR_i)$	v_i	$w_i=1/v_i$
1	study1	15	66	11	57	1.23	0.21	0.199	5.03
2	study2	43	315	23	233	1.44	0.37	0.075	13.30
3	study3	25	190	12	193	2.29	0.83	0.135	7.41
4	study4	94	514	65	495	1.48	0.39	0.031	32.54
5	study5	32	186	17	147	1.59	0.46	0.104	9.59
6	study6	7	88	24	342	1.15	0.14	0.200	5.00
7	study7	343	1312	196	1160	1.74	0.55	0.010	99.14
8	study8	43	318	74	410	0.71	-0.34	0.043	23.05
9	study9	27	334	50	551	0.88	-0.13	0.062	16.05
10	study10	8	233	13	298	0.78	-0.25	0.210	4.76
11	study11	351	1499	155	1104	1.87	0.63	0.011	89.08

A sensitivity analysis was performed for biggest studies (e.g. Studies 4, 7 and 11) and the meta-analysis results are as follows:

In the Sensitivity analysis,

FE ORp=1.07 with 95% CI (0.87, 1.33)

RE ORp=1.13 with 95% CI (0.85, 1.52)

PQ=0.10 which is marginally significant.

In the full analysis,

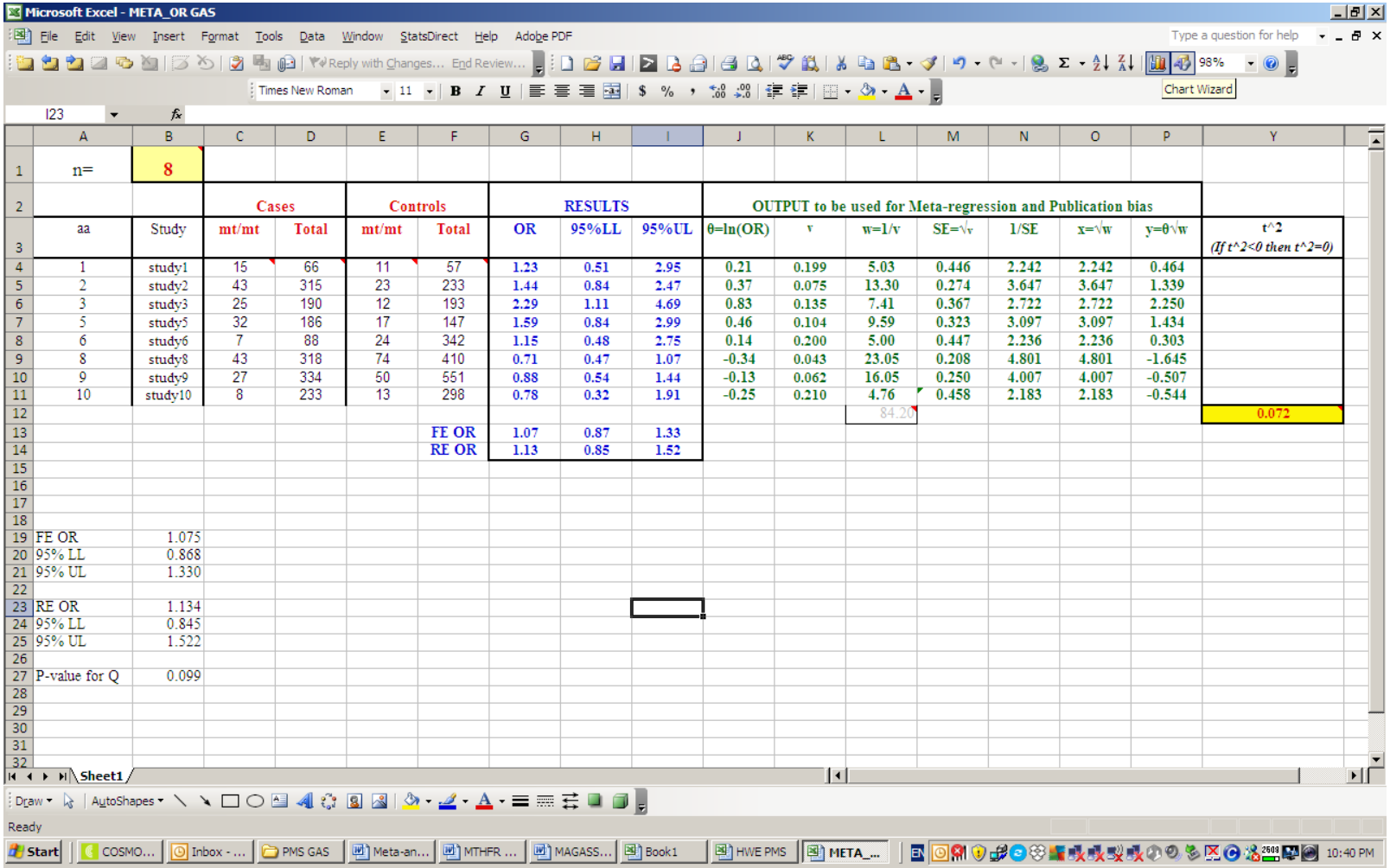
PQ=0.002 which was significant (PQ<0.10).

RE ORp=1.35 with 95% CI (1.08, 1.69)

Thus, the exclusion of Studies 4, 7 and 11 changes the pattern of results in the estimation of the pooled estimate of the treatment difference and in heterogeneity.

Practice in XL

In XL, the file META_OR_GAS produces the FE ORp and RE ORp with the respective 95% CIs for sensitivity analysis (just delete the row corresponding to Studies 4, 7 and 11 and change the n from 11 to 8).



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Y
1	n=	8															
2			Cases		Controls		RESULTS			OUTPUT to be used for Meta-regression and Publication bias							
3	aa	Study	mt/mt	Total	mt/mt	Total	OR	95%LL	95%UL	θ=ln(OR)	v	w=1/v	SE=√v	1/SE	x=√w	y=θ√w	t^2 (If t^2<0 then t^2=0)
4	1	study1	15	66	11	57	1.23	0.51	2.95	0.21	0.199	5.03	0.446	2.242	2.242	0.464	
5	2	study2	43	315	23	233	1.44	0.84	2.47	0.37	0.075	13.30	0.274	3.647	3.647	1.339	
6	3	study3	25	190	12	193	2.29	1.11	4.69	0.83	0.135	7.41	0.367	2.722	2.722	2.250	
7	5	study5	32	186	17	147	1.59	0.84	2.99	0.46	0.104	9.59	0.323	3.097	3.097	1.434	
8	6	study6	7	88	24	342	1.15	0.48	2.75	0.14	0.200	5.00	0.447	2.236	2.236	0.303	
9	8	study8	43	318	74	410	0.71	0.47	1.07	-0.34	0.043	23.05	0.208	4.801	4.801	-1.645	
10	9	study9	27	334	50	551	0.88	0.54	1.44	-0.13	0.062	16.05	0.250	4.007	4.007	-0.507	
11	10	study10	8	233	13	298	0.78	0.32	1.91	-0.25	0.210	4.76	0.458	2.183	2.183	-0.544	
12												84.20					0.072
13						FE OR	1.07	0.87	1.33								
14						RE OR	1.13	0.85	1.52								
15																	
16																	
17																	
18																	
19	FE OR	1.075															
20	95% LL	0.868															
21	95% UL	1.330															
22																	
23	RE OR	1.134															
24	95% LL	0.845															
25	95% UL	1.522															
26																	
27	P-value for Q	0.099															
28																	
29																	
30																	
31																	
32																	

Selection bias

The pooled estimate of OR can be systematically influenced by the selection of studies for inclusion in the meta-analysis.

Then, in the meta-analysis, bias may be introduced in two different ways:

A) by including studies which have themselves produced biased estimates of the OR, and

B) by excluding eligible studies because the relevant data are not available.

Bias in individual GAS

A) The individual studies with biased estimates of the OR may introduce bias in meta-analysis due to:

-The inclusion of studies with methodological flaws (such as inappropriate patient selection)

-The chronological order in which studies are conducted. Small early studies may produce larger estimates of OR than from later larger studies.

Publication bias

B) The data of eligible studies can be missing due to Publication bias.

Publication bias is introduced when the meta-analysis is restricted to the synthesis of results obtained only from studies which have been published.

-Often, the decision to submit or accept a paper with the results of a GAS for publication in a journal is influenced by the significance of results:

> **a large GAS with significant OR is more likely to be published than a small GAS with non-significant OR.**

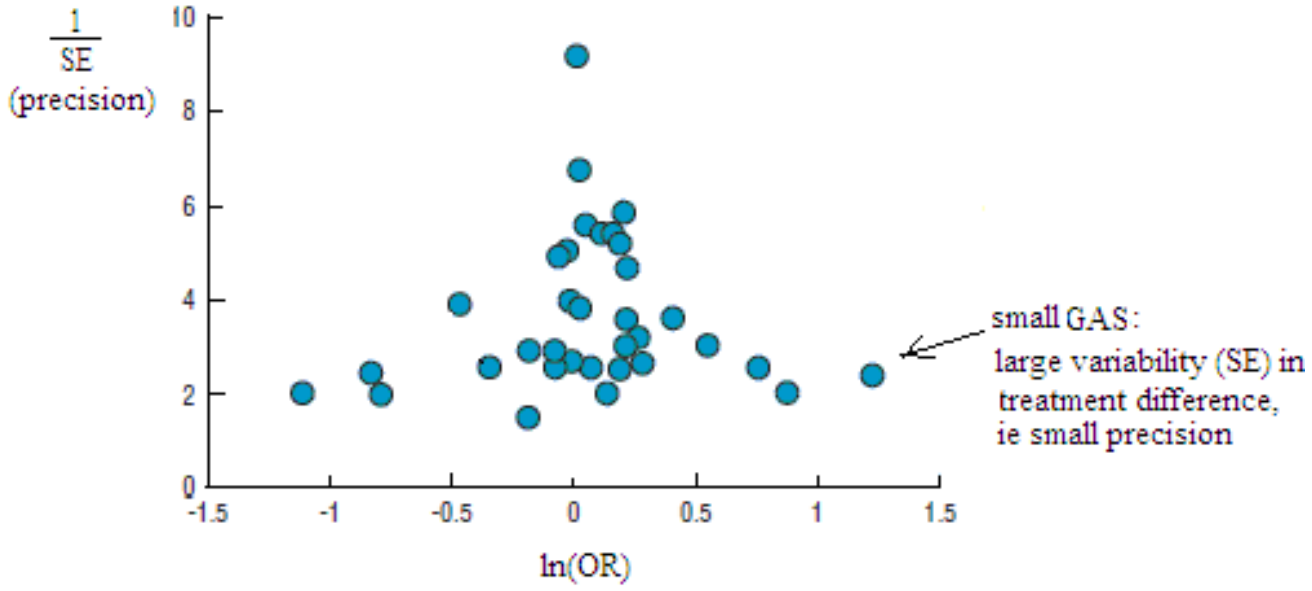
-Also, GAS indicating that an wt/wt plays a role in disease development are less likely to be published than those indicating that mt-carriers a role in disease development.

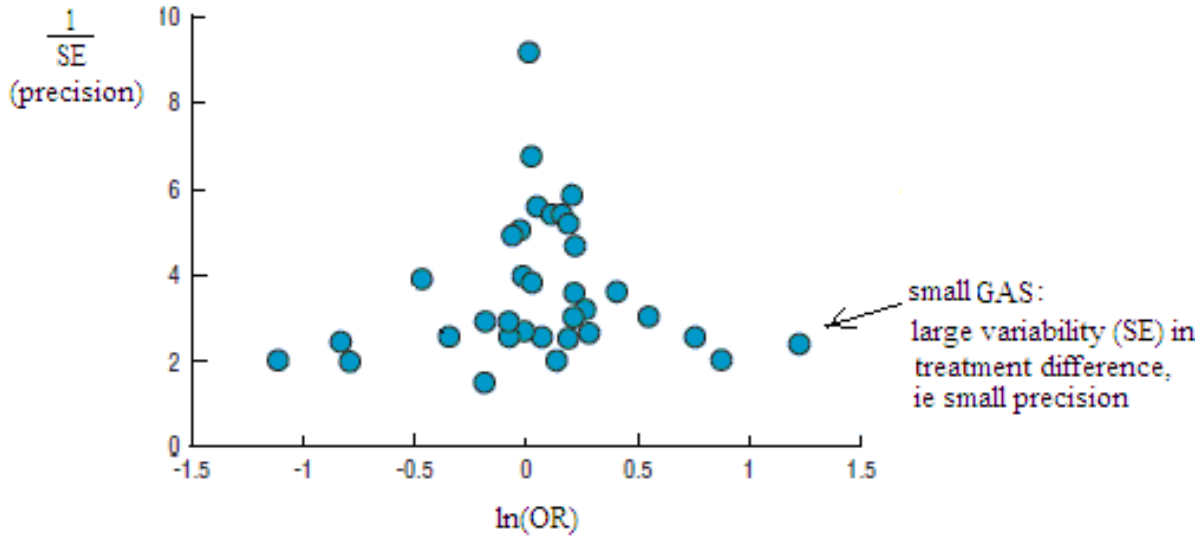
A remedy to deal with selection bias is to perform a Sensitivity analysis.

Funnel plot

The simplest and most commonly used method to detect publication bias is an informal examination of a funnel plot.

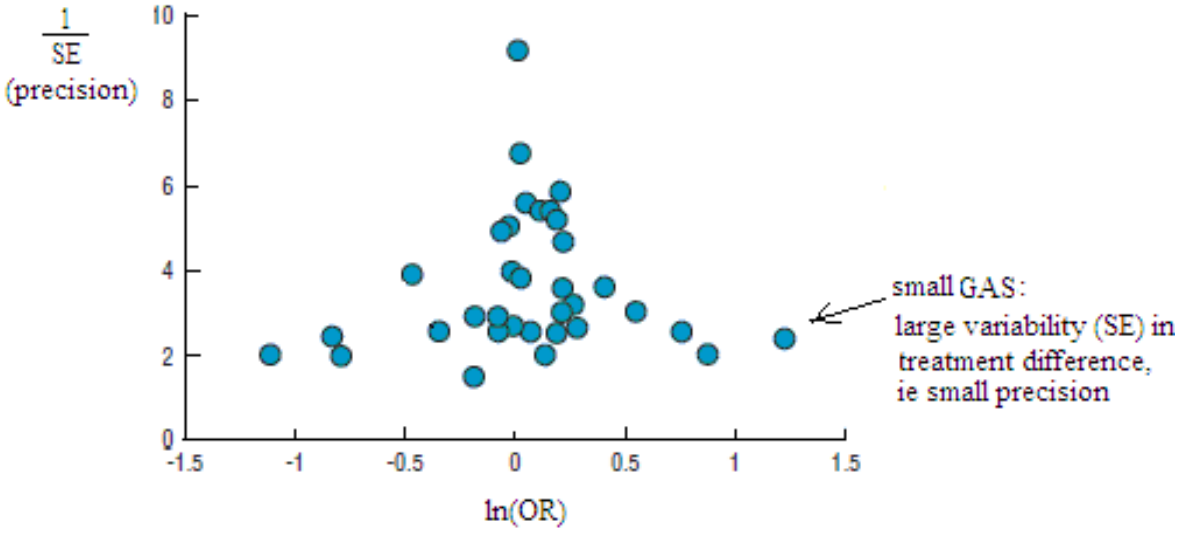
A funnel plot is a plot of each study's θi=ln(ORi) against the precision (expressed as 1/SEi).





In the absence of publication bias, this plot will resemble a symmetrical inverted shaped like a funnel.

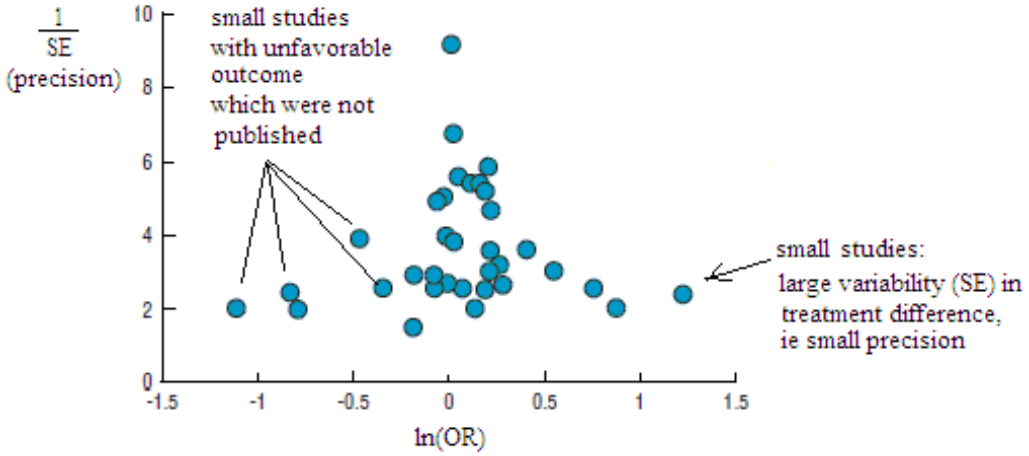
Then, the spread of results will be wide at the bottom of the graph where small studies are placed, and will become narrower as the studies become larger.

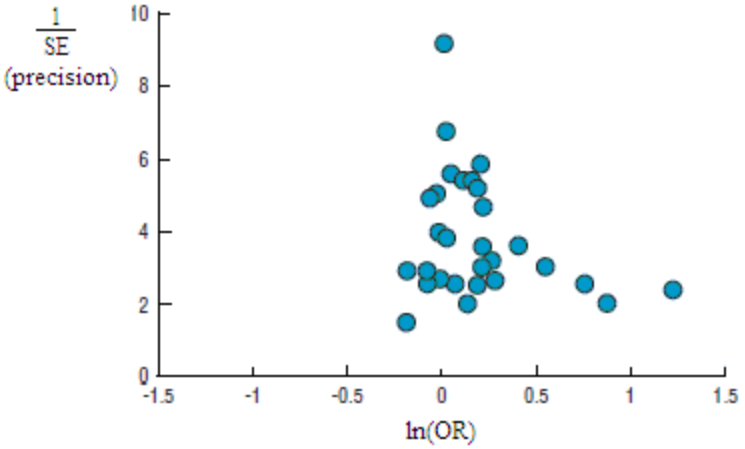


This funnel shape is expected because GAS of smaller size (which are more numerous) have increasingly large variation (ie small precision) in the estimates of their ln(OR).

Note that a small GAS has a large SE and small precision since precision= 1/SE.

However, since smaller or nonsignificant GAS are less likely to be published, GAS in the bottom left hand corner (when a undesirable outcome is being considered, e.g. wt/wt is related to the disease) of the plot are often omitted, creating a degree of asymmetry in the funnel.



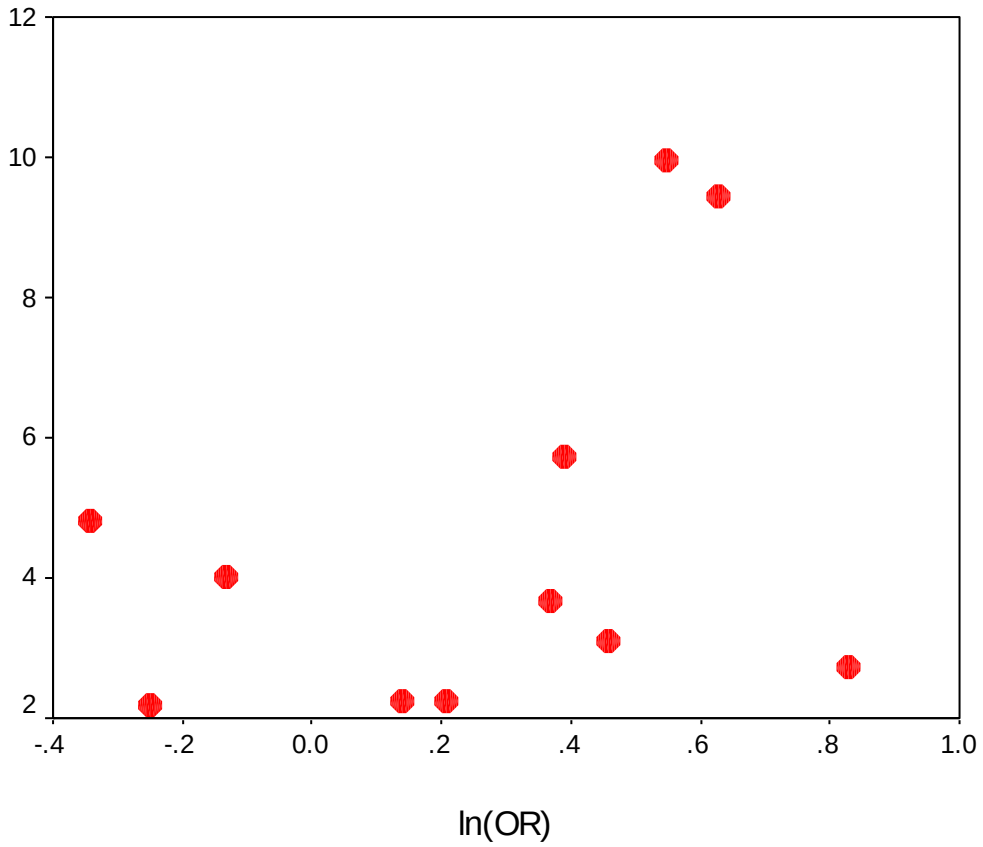


Thus, an asymmetry in funnel plot indicates the existence of publication bias.

Example - MTHFR C677T and Breast Cancer

		Cases		Controls					
i		mt/mt	Total	mt/mt	Total	OR_i	$\theta_i=\ln(OR_i)$	v_i	$w_i=1/v_i$
1	study1	15	66	11	57	1.23	0.21	0.199	5.03
2	study2	43	315	23	233	1.44	0.37	0.075	13.30
3	study3	25	190	12	193	2.29	0.83	0.135	7.41
4	study4	94	514	65	495	1.48	0.39	0.031	32.54
5	study5	32	186	17	147	1.59	0.46	0.104	9.59
6	study6	7	88	24	342	1.15	0.14	0.200	5.00
7	study7	343	1312	196	1160	1.74	0.55	0.010	99.14
8	study8	43	318	74	410	0.71	-0.34	0.043	23.05
9	study9	27	334	50	551	0.88	-0.13	0.062	16.05
10	study10	8	233	13	298	0.78	-0.25	0.210	4.76
11	study11	351	1499	155	1104	1.87	0.63	0.011	89.08

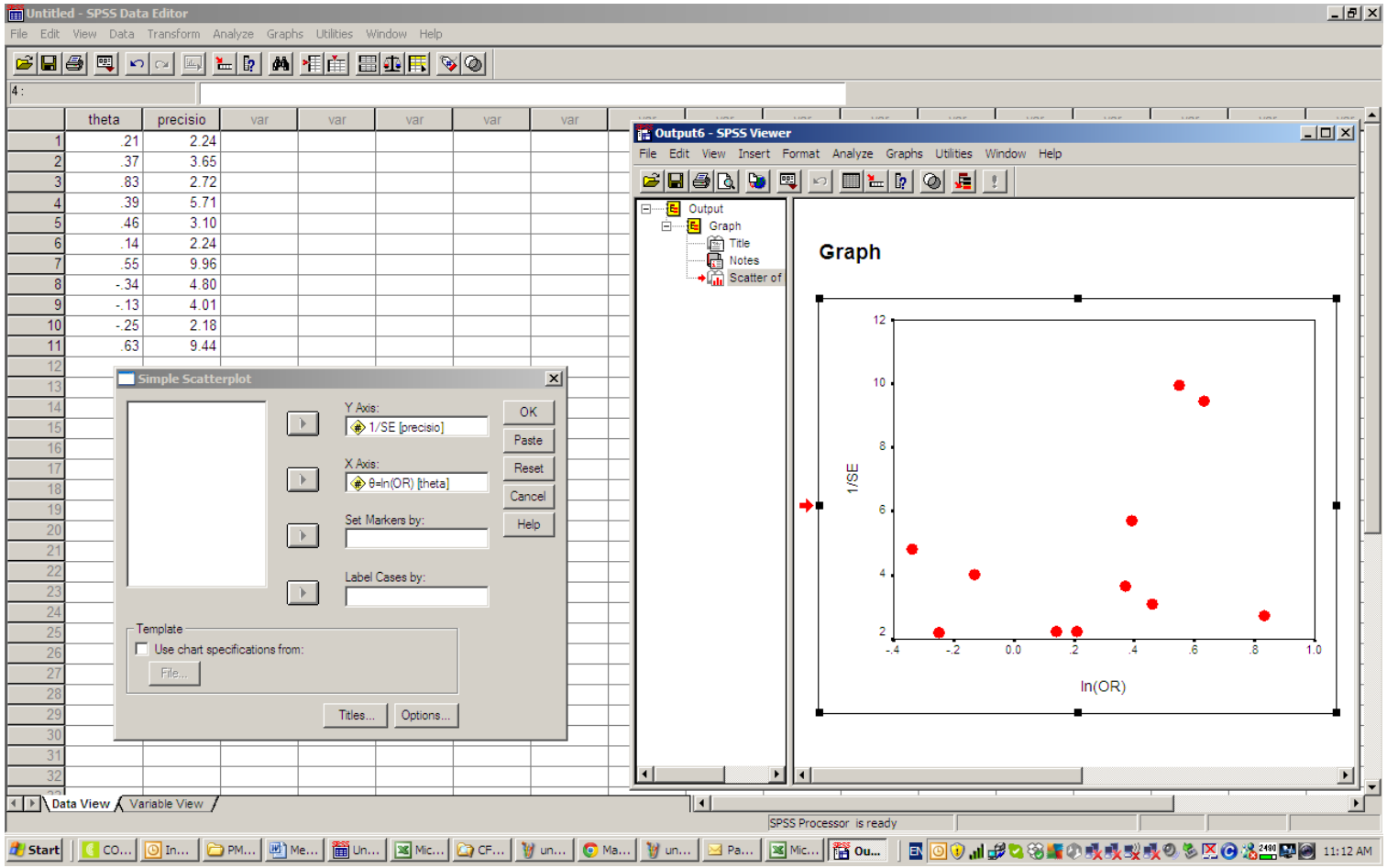
We could test for publication bias graphically using XL



The funnel plot shows some symmetry and therefore, there might be an indication of lack of publication bias.

Practice in SPSS

Copy-paste the data from XL file META_OR_GAS to SPSS. Then, the funnel plot for exploring the existence of publication bias can be drawn in SPSS:



Egger's test

A formal test for publication bias can be based on linear regression analysis (Egger's test). In particular, the following regression line is fitted:

$$y_i = a + bx_i$$ **for i=1 to n, where n is the number of studies**

yi is the standardized estimate of θi: $y_i = \theta_i \sqrt{w_i}$

xi is the precision: $x_i = \sqrt{w_i}$

,

$$\theta_i \sqrt{w_i} = a + b\sqrt{w_i}$$

$$\theta_i \sqrt{w_i} = a + b\sqrt{w_i}$$

A test of publication bias would be a test whether the intercept a is equal to zero.

The intercept, a, provides a measure of funnel plot asymmetry: the larger its deviation from zero, the more pronounced is the asymmetry of the funnel plot (ie existence of publication bias).

$$\theta_i \sqrt{w_i} = a + b\sqrt{w_i}$$

The intercept a and slope b can be obtained by performing a typical least-squares regression of yi on xi (using SPSS).

Then, in testing whether the intercept a is 0, the statistic is t=a/SE(a) compared against the 5% point of the t-distribution with n-2 df.

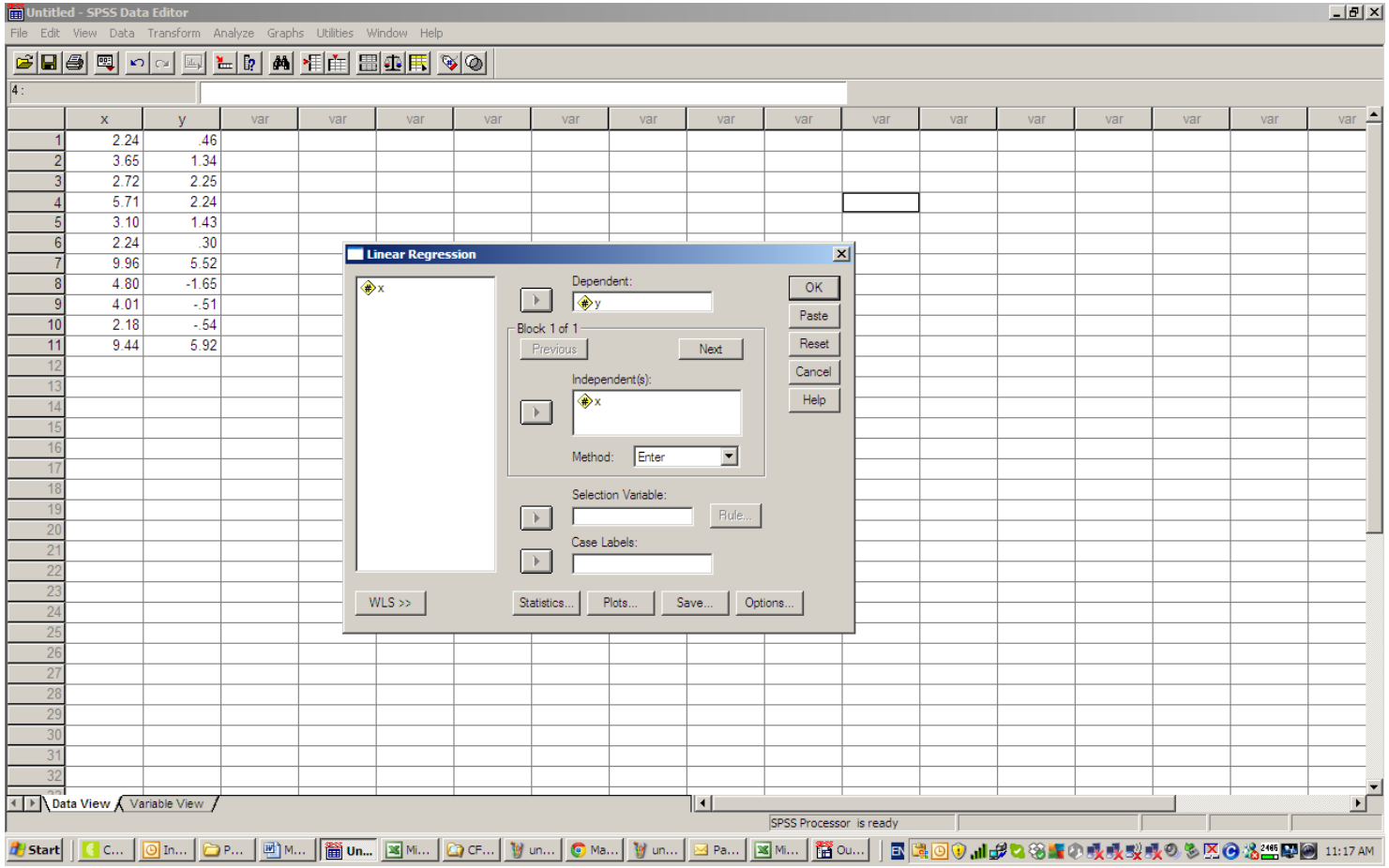
Thus, if t is less than the 5% point of the t-distribution with n-2 df, there is no indication of publication bias.

Example - MTHFR C677T and Breast Cancer

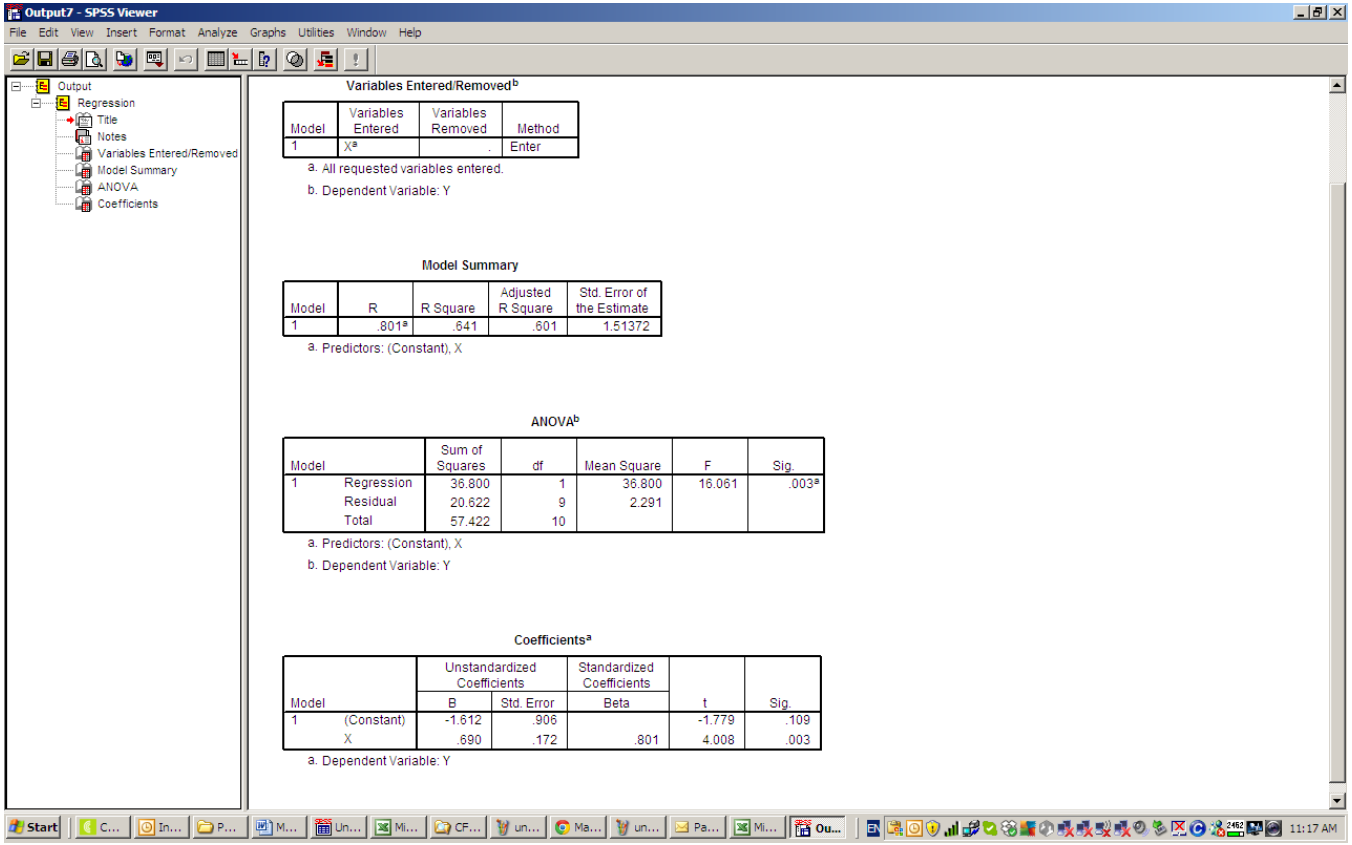
		Cases		Controls					
i		mt/mt	Total	mt/mt	Total	OR_i	$\theta_i=\ln(OR_i)$	v_i	$w_i=1/v_i$
1	study1	15	66	11	57	1.23	0.21	0.199	5.03
2	study2	43	315	23	233	1.44	0.37	0.075	13.30
3	study3	25	190	12	193	2.29	0.83	0.135	7.41
4	study4	94	514	65	495	1.48	0.39	0.031	32.54
5	study5	32	186	17	147	1.59	0.46	0.104	9.59
6	study6	7	88	24	342	1.15	0.14	0.200	5.00
7	study7	343	1312	196	1160	1.74	0.55	0.010	99.14
8	study8	43	318	74	410	0.71	-0.34	0.043	23.05
9	study9	27	334	50	551	0.88	-0.13	0.062	16.05
10	study10	8	233	13	298	0.78	-0.25	0.210	4.76
11	study11	351	1499	155	1104	1.87	0.63	0.011	89.08

Practice in SPSS

Copy-paste the data from XL file META_OR to SPSS. Then, a typical least-squares regression of on is fitted (i.e. we fit a regression line of xi on yi) to obtain the intercept a and slope b using SPSS as follows:



The produced output with the results of the regression analysis is shown below:



The P-value for testing whether the intercept a is 0 is P=0.109, i.e. the intercept a is not significant since P≥0.05.

Therefore, there is no significant publication bias in the meta-analysis.

Practice

A GAS investigating the association between the alleles ADH2*1 (mt) and ADH2*2 (wt) with alcoholism produced the following genotype distributions:

			ADH2					
			*1/*1		*1/*2		*2/*2	
Year	Author	Racial	alcoholic	healthy	alcoholic	healthy	alcoholic	healthy
1993	Sherman	Caucasian	7	18	19	3	19	2
1994	Muramatsu	Chinese	13	12	8	43	11	50
1994	Thomasson	Taiwan	3	1	28	10	63	54
1995	Maezawa	Japanese	30	2	28	22	38	36
1996	Chen	Taiwan	14	0	15	19	17	44
1996	Higuchi	Japanese	204	33	224	160	227	268
1997	Espinós	Caucasian	62	58	9	12	0	1
1999	Chen CC	Chinese	130	43	106	205	104	297
2000	Chao	Taiwan	51	17	129	102	101	122
2001	Lee	Korean	10	6	32	18	64	40
2001	Ogurtsov	Caucasian	56	15	51	29	3	6

Test for publication bias the GAS included in the meta-analysis.

Cumulative and recursive cumulative meta-analysis

Cumulative and recursive cumulative meta-analyses is another way to explore heterogeneity in risk effect for a genetic model in time.

They provide a framework for updating a genetic effect from all studies and a measure of how much the genetic effect changes as evidence accumulates.

Cumulative meta-analysis

In cumulative meta-analysis, studies are ordered by publication year, and then, the pooled OR is obtained when a new study is published.

Thus, cumulative meta-analysis indicates the trend in estimated risk effect.

Recursive cumulative meta-analysis

In recursive cumulative meta-analysis, the relative change in cumulative pooled OR in each publication year is calculated (cumulative OR in next year/ cumulative OR in current year).

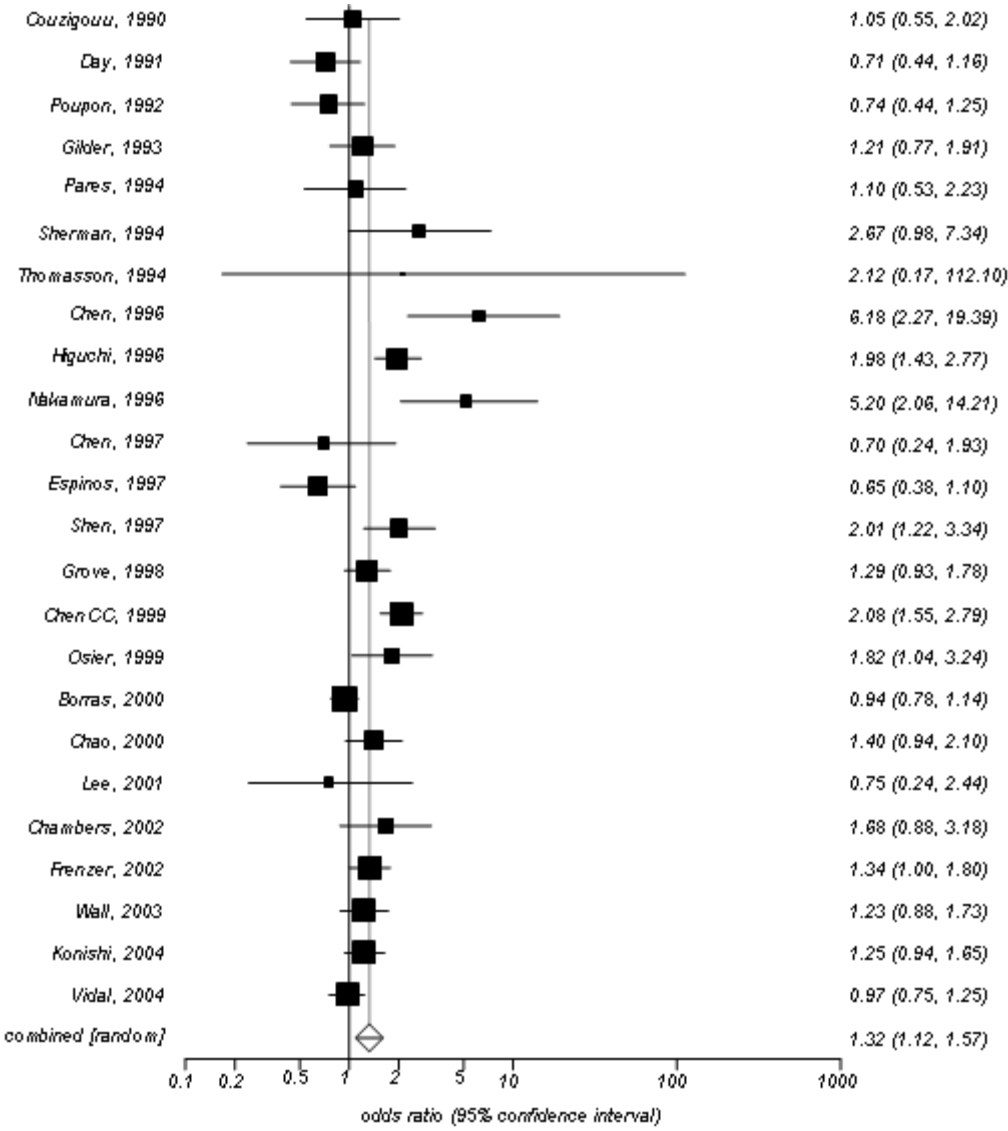
Thus, recursive cumulative meta-analysis indicates the stability in risk effect.

Wide oscillations in risk effect early in the course of accumulating evidence are usually associated with major changes in risk effect in the future.

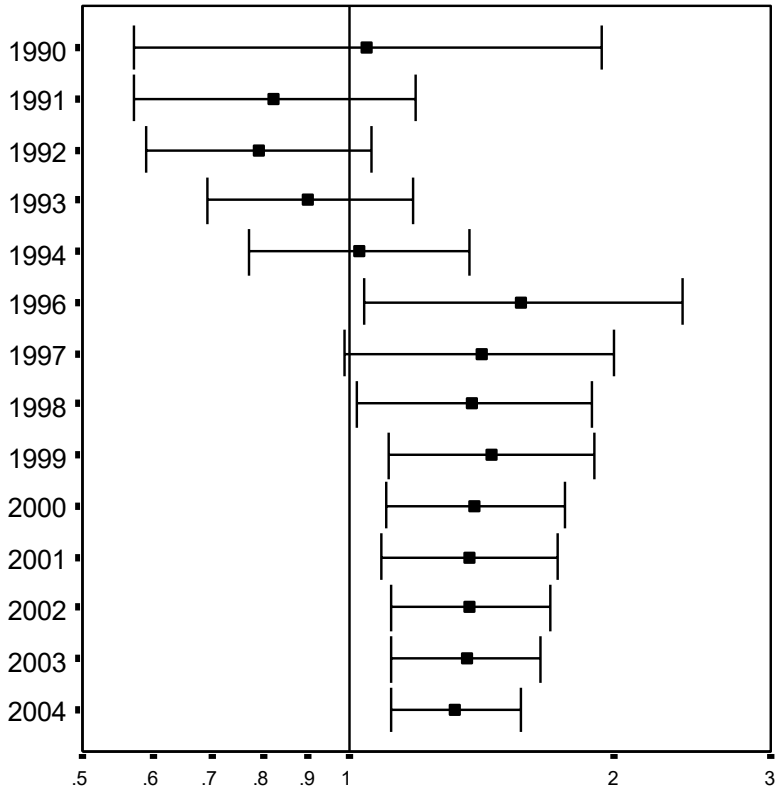
If the oscillations remain in time then more information is required to draw safe conclusion on the magnitude of the risk effect.

Example

A meta-analysis for investigating the association between alcoholism and the ADH3 (wt=*1, mt=*2) gene polymorphism for the recessive model (*2/*2 vs. rest) produced significant heterogeneity (p<0.01) and OR=1.32 (1.12-1.57)



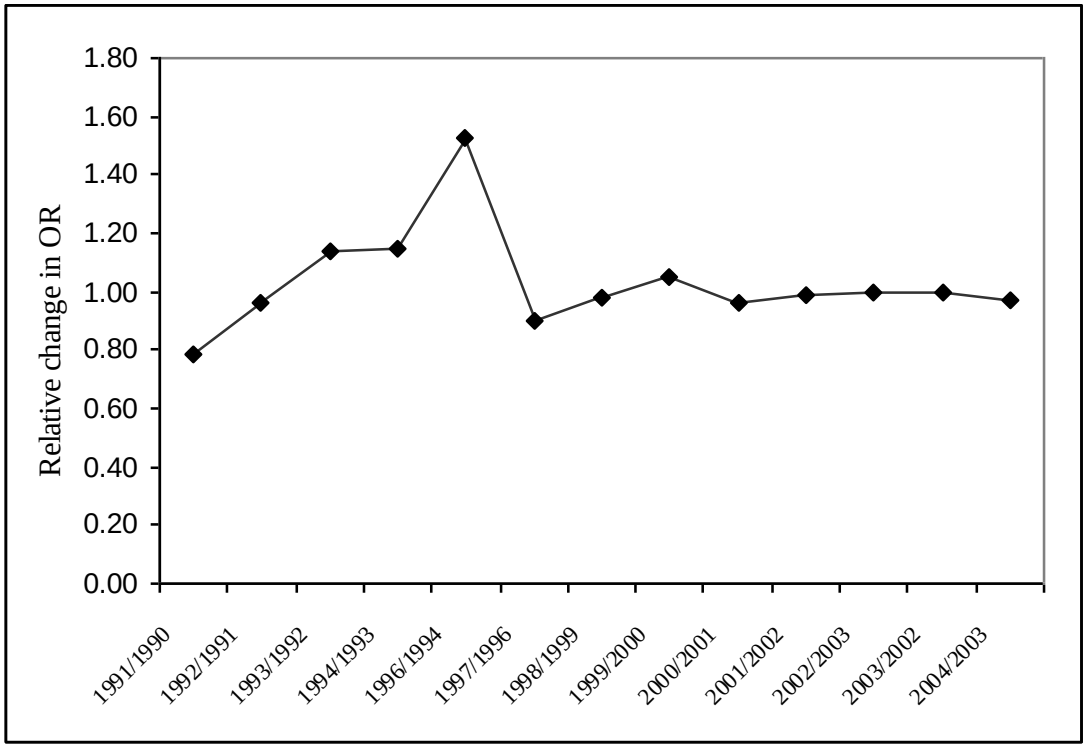
The cumulative meta-analysis for the recessive model showed a trend of association as information accumulates:



In recursive cumulative meta-analysis, the relative change in OR fluctuated around the value of OR=1.0, and it stabilizes after information step 2000/2001.

The scatter of the relative changes in OR is wider at the beginning, and then, it shrinks as evidence accumulates.

This stability indicates that there is enough evidence to draw safe conclusion about the risk effect of ADH3 gene polymorphism in alcoholism.



The other genetic models

The meta-analysis of the other genetic models (dominant, additive, co-dominant) is similar to the recessive model.

When more than on model is significant, the identification of the mode of inheritance can be based on the heuristic algorithm presented in the "Analysis of GAS" course.

Genetic model-free approach

A genetic model-free approach for testing the association between disease status (disease vs. healthy) and genotype is the generalized odds ratio (ORG).

The ORG is a single statistic that utilizes the complete genotype distribution (not merging genotypes like in the co-dominant model) and provides an estimate of the overall risk effect.

Definition of ORG

ORG is the probability of a subject being diseased relative to probability of being free of disease, given that the diseased subject has a higher mutational load than the non-diseased.

$$OR_G = \frac{\text{Probability being diseased, diseased has high mutational load}}{\text{Probability of being non-diseased, non-diseased has low mutational load}}$$

When ORG>1 then an increased genetic exposure (mutational load) implies disease.

ORGGASMA

“ORGGASMA”: a software for implementing the generalized odds ratio methodology for the analysis and meta-analysis of GAS.

The software “ORGGASMA” (together with instructions how to operate it) is freely available and it can be downloaded form the web site http://biomath.med.uth.gr

Download the “ORGGASMA” software and operated it only for the “cmd” command of windows (do not double click the icon).

Example – MTHFR C677T and Breast Cancer

Eleven GAS were conducted to investigate the association between MTHFR C677T (wt=C and mt=T) variant and susceptibility to breast cancer (BC). The data are as follows:

		Breast Cancer			Helathy controls		
MTHFR C677T GAS	Race	***TT***	***TC***	***CC***	***TT***	***TC***	***CC***
1	Whites	**15**	**23**	**28**	**11**	**21**	**25**
2	Whites	**43**	**162**	**110**	**23**	**92**	**118**
3	E. Asians	**25**	**85**	**80**	**12**	**87**	**94**
4	E. Asians	**94**	**242**	**178**	**65**	**215**	**215**
5	E. Asians	**32**	**96**	**58**	**17**	**80**	**50**
6	Whites	**7**	**38**	**43**	**24**	**145**	**173**
7	E. Asians	**343**	**695**	**274**	**196**	**577**	**387**
8	Whites	**43**	**140**	**135**	**74**	**196**	**140**
9	Whites	**27**	**141**	**166**	**50**	**259**	**242**
10	Whites	**8**	**91**	**134**	**13**	**104**	**181**
11	E. Asians	**351**	**786**	**362**	**155**	**509**	**440**

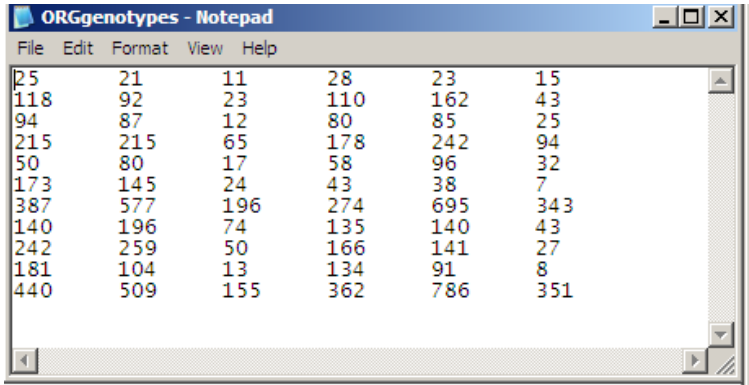
Is there evidence that the variant MTHFR C677T is associated with the development of BC?

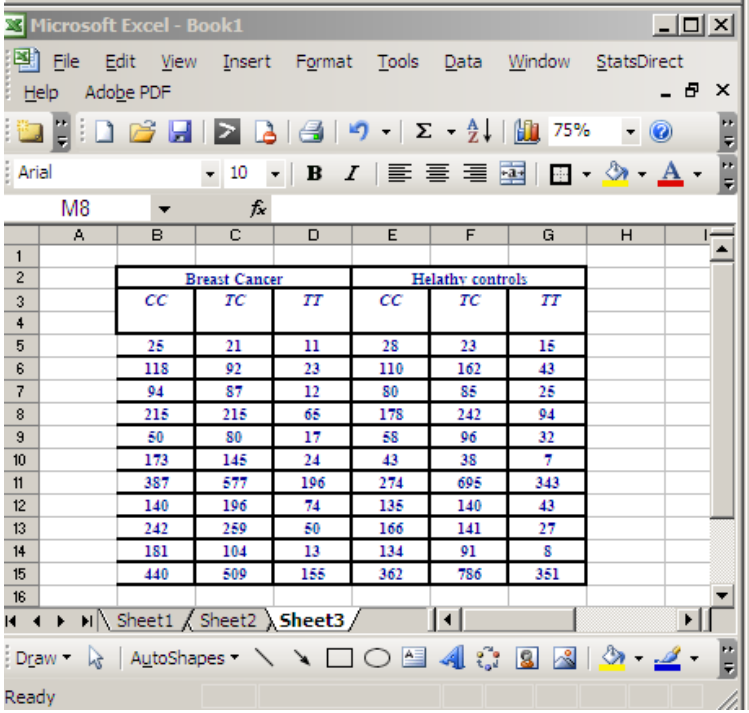
MTHFR C677T and Breast Cancer

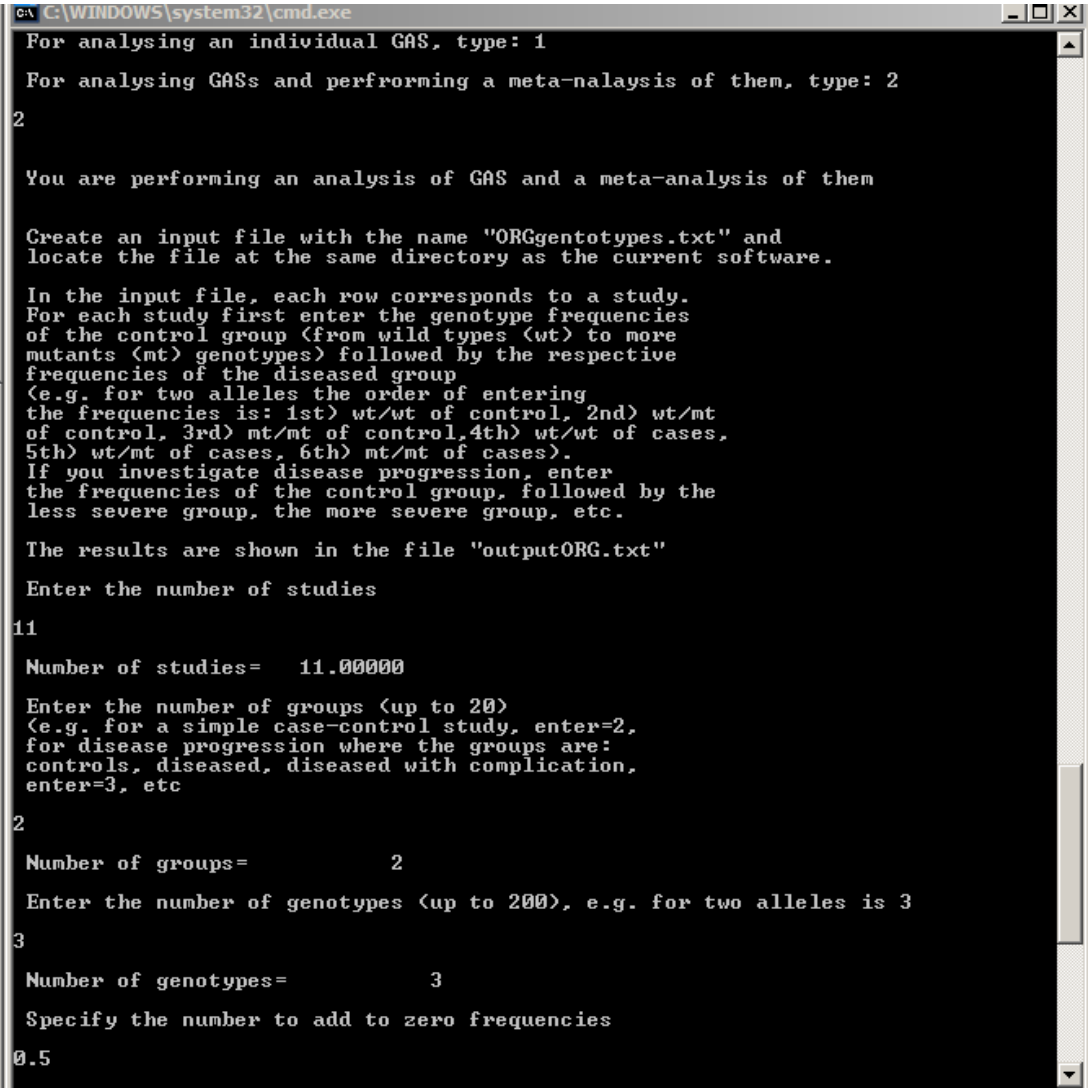
Prior to meta-analysis, we will make an assumption:

Subjects who are homozygous for TT allele have the highest mutational load, those homozygous for CC allele have the lowest, and heterozygous CT have an intermediate level.

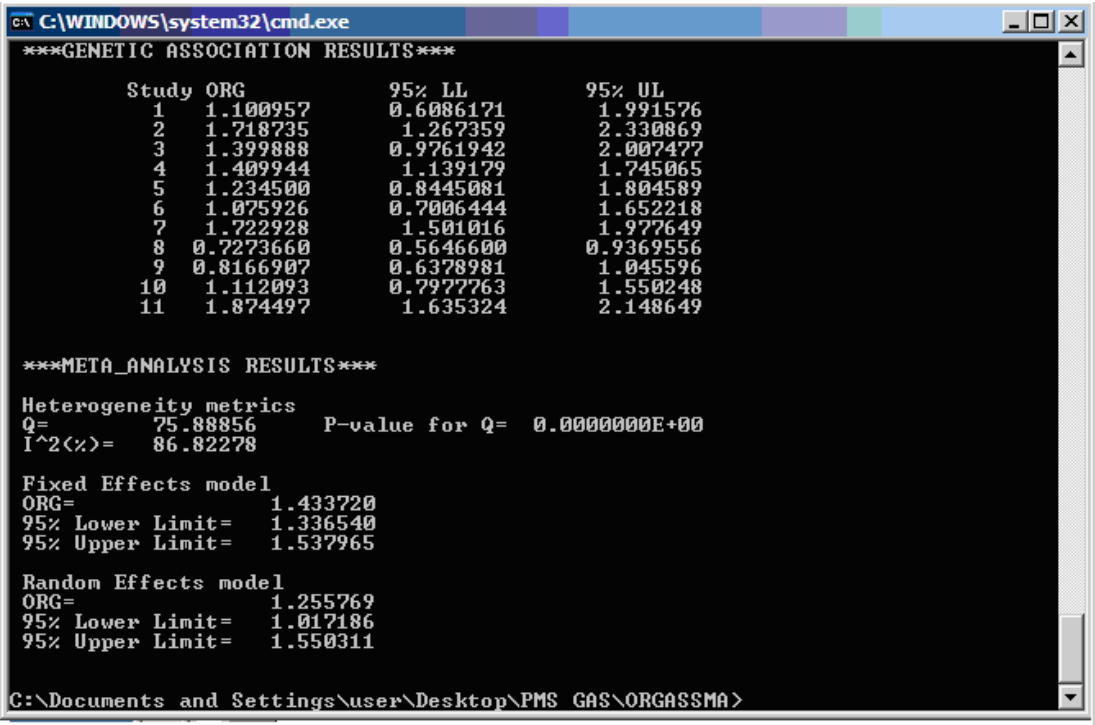
In ORGGASMA, the data are stored in the file “ORGgenotypes.txt” at the same folder with the ORGGASMA.exe file.

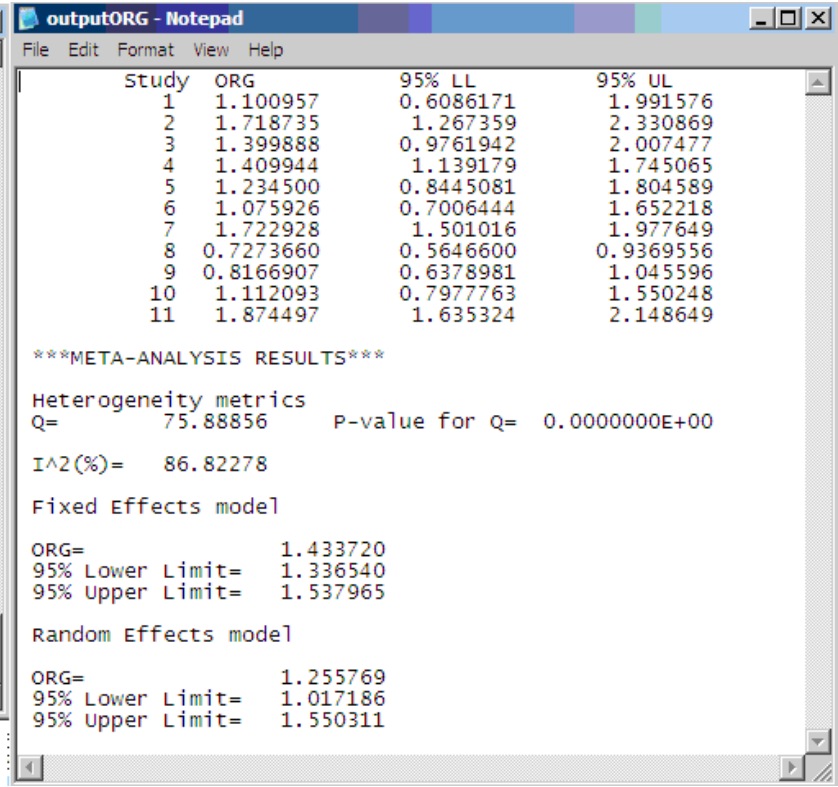






The meta-analysis results were the following (the results were saved in the file “outputORG.txt”):





The heterogeneity between studies was significant P<0.10.

Then, the random effects (RE) model generalized OR was ORG=1.26 (1.02, 1.55).

Since 1 is not included in the 95% CI the ORG is significant.

ORG=1.26

The interpretation of the finding is as follows: For any two subjects, diseased with BC and healthy, the probability of being diseased is 26% higher (relative to the probability of being non-diseased) given that the diseased subject has higher mutational load for the variant MTHFR C677T than the healthy one.

Thus, an increased genetic exposure (mutational load) implies disease.

Subgroup analysis by “race”

A subgroup analysis for each “race” can be performed in a similar way to the full analysis.

Sensitivity analysis

A sensitivity analysis can be performed for the study 5 where the controls are in HWE marginally.

Practice

A GAS investigating the association between the alleles ADH2*1 (mt) and ADH2*2 (wt) with alcoholism produced the following genotype distributions:

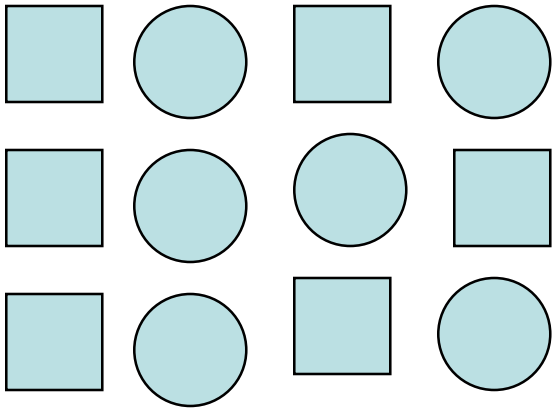
			ADH2					
			*1/*1		*1/*2		*2/*2	
Year	Author	Racial	alcoholic	healthy	alcoholic	healthy	alcoholic	healthy
1993	Sherman	Caucasian	7	18	19	3	19	2
1994	Muramatsu	Chinese	13	12	8	43	11	50
1994	Thomasson	Taiwan	3	1	28	10	63	54
1995	Maezawa	Japanese	30	2	28	22	38	36
1996	Chen	Taiwan	14	0	15	19	17	44
1996	Higuchi	Japanese	204	33	224	160	227	268
1997	Espinós	Caucasian	62	58	9	12	0	1
1999	Chen CC	Chinese	130	43	106	205	104	297
2000	Chao	Taiwan	51	17	129	102	101	122
2001	Lee	Korean	10	6	32	18	64	40
2001	Ogurtsov	Caucasian	56	15	51	29	3	6

Perform a full meta-analysis of the GAS, a subgroup analysis by "race" and sensitivity analysis for GAS with the controls not in HWE.

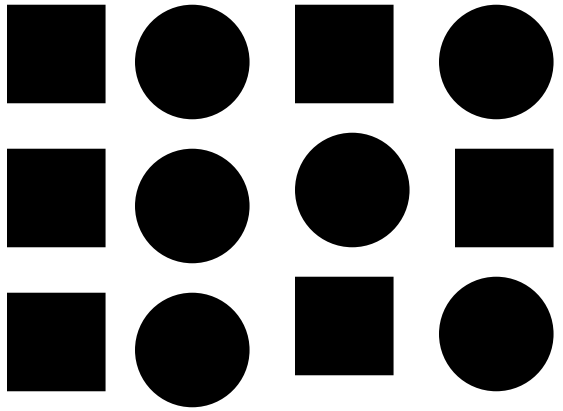
HEGESMA
METRADISC

Genome Wide Association Studies (GWAS)

GWAS is scan of genomic sequence variants which enable to examine hundreds of thousands of SNPs in cases and controls



500-1000 cases

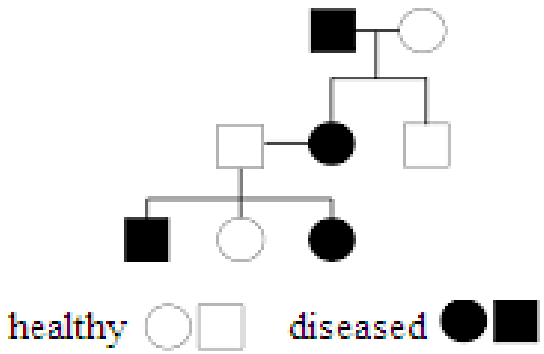


500-1000 controls

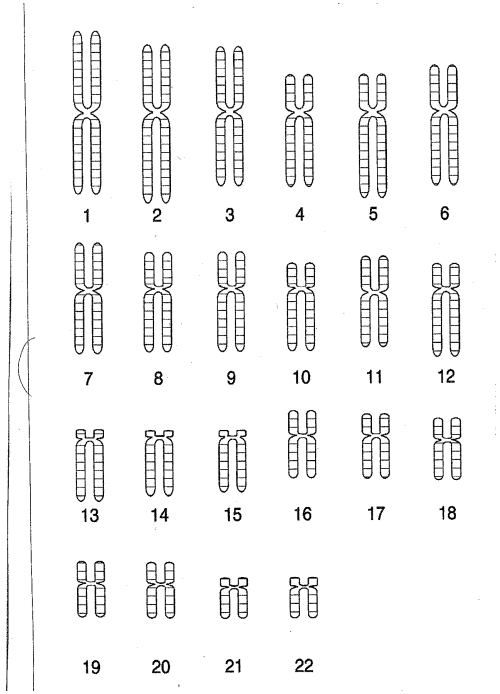
- **Calculate which of the alleles of 300-500,000 SNPs) are more frequent in cases than controls**

Whole Genome Scans or Genome Linkage Scans (WGS)

WGS is a technique to determine linkage of complex disease using families of sibling pairs with the disease



~400 microsatellite (tandem repeat sequence) markers across the genome evenly spaced every 10 cM= 1million base pairs (bps)



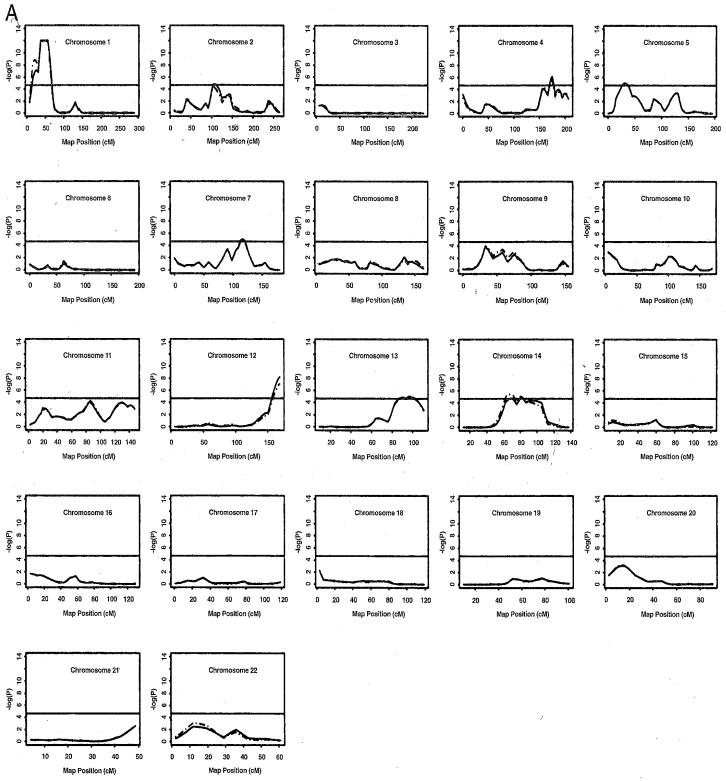
WGS Identifies "linkage peak" with extensive allele sharing in a cohort of sibling pairs with the disease.

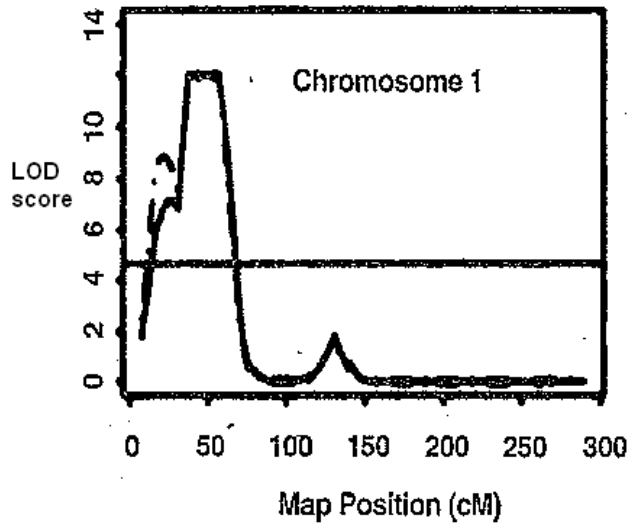
Steps:

1) **>300 families with >1000 affected individuals**

2) **Logarithm of odds ratio (LOD) score >3 for each marker**
i.e. large probability of allele sharing in sibling pairs (of linkage for the marker)

Significant linkage peak => gene(s) in LD with a marker





Genome scans on complex diseases have produced inconclusive inferences:

- **linkage signals tend to be rather weak**
- **number of families and affected sibpairs are relatively small**
- **individual genome scans identify linkage in different chromosomal regions**

HEGESMA

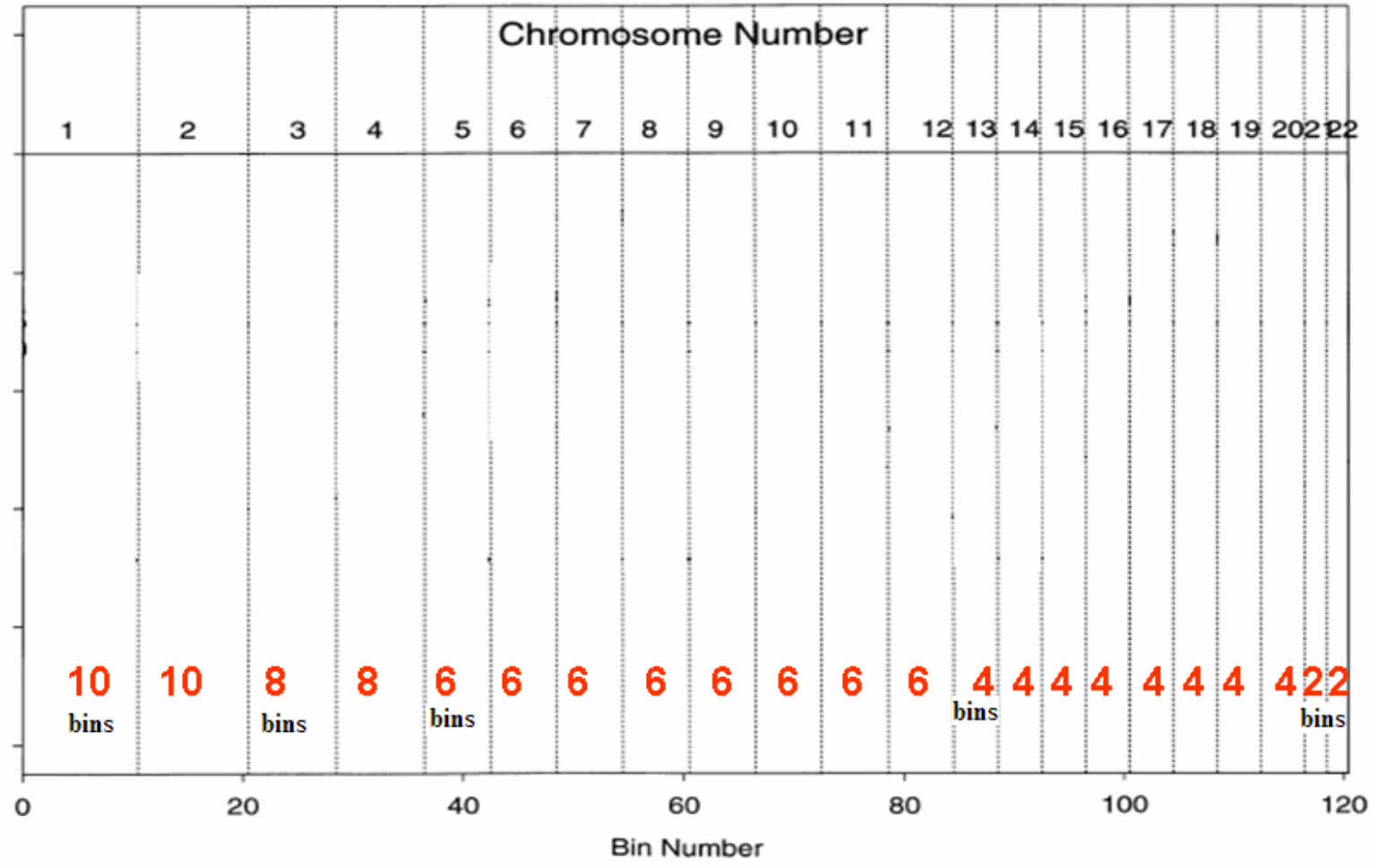
(***Zintzaras and Ioannidis 2005 Genetic Epidemiology, Bioinformatics)***

HEGESMA is a method for synthesizing data from diverse genome scans

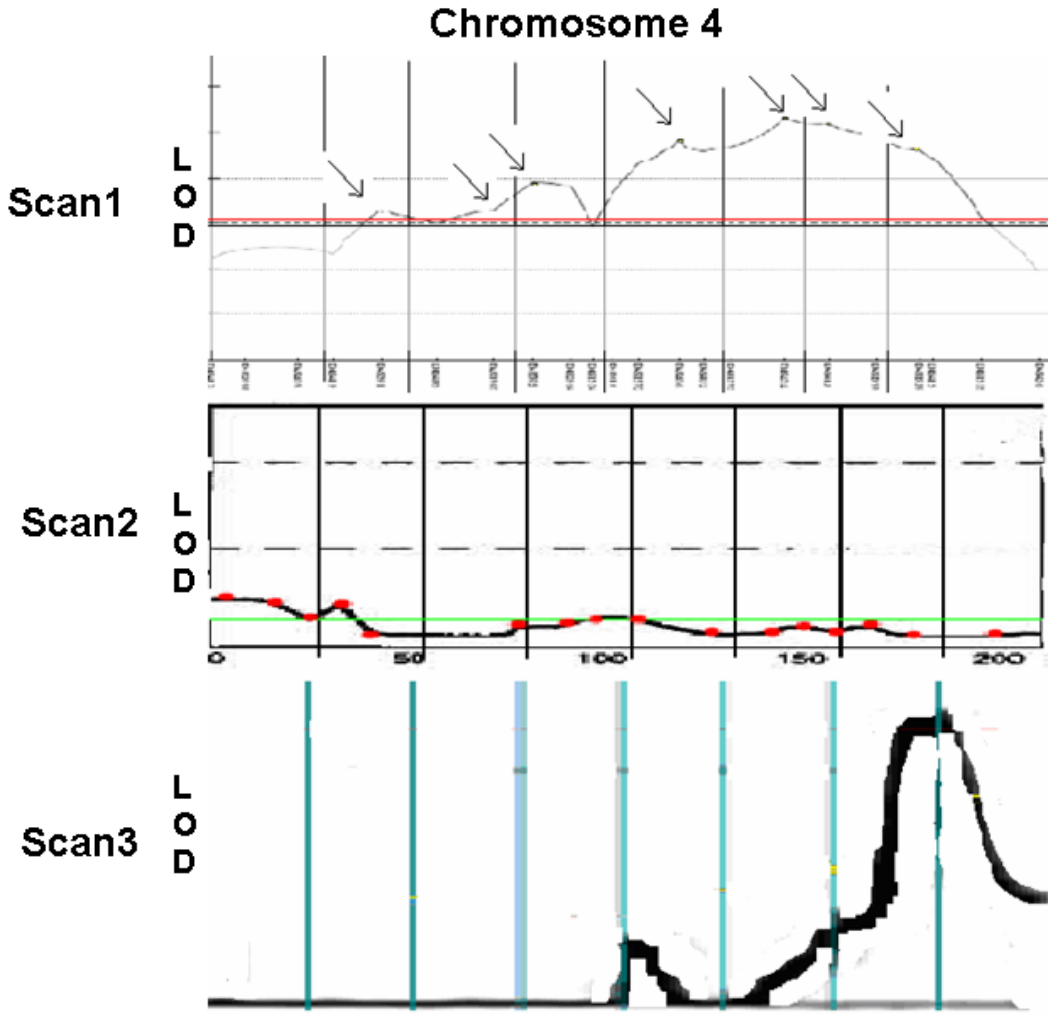
HEGESMA starts by splitting the chromosomes into bins of approximately equal length:

- each bin has a width of 30 cM

- 120 bins in total for the whole genome



For each genome scan, the highest LOD score obtained within the bin is recorded.



The method

- **For each scan the bins are ranked according to their LOD score.**

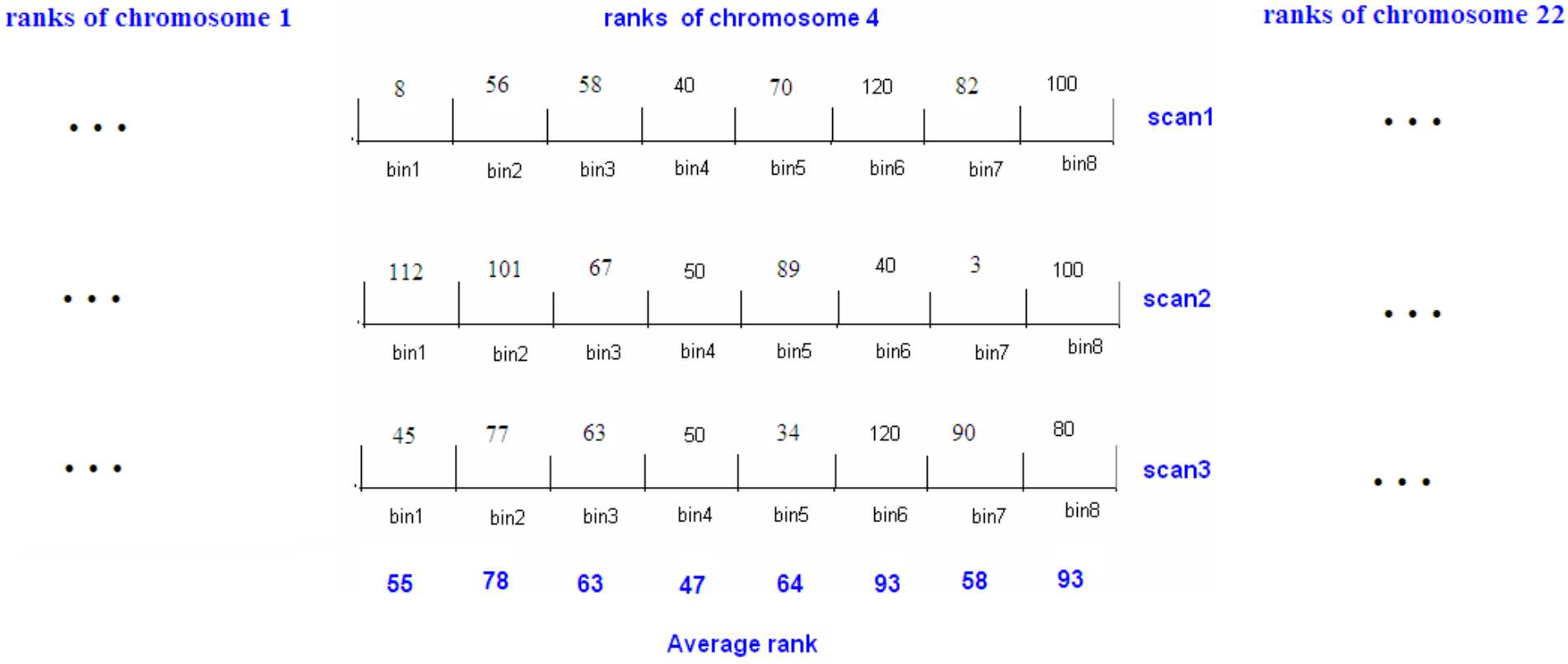
- **Assign with 120 to the highest linkage score and with 1 the lowest.**

- **The ranks for each bin are averaged across scans:**

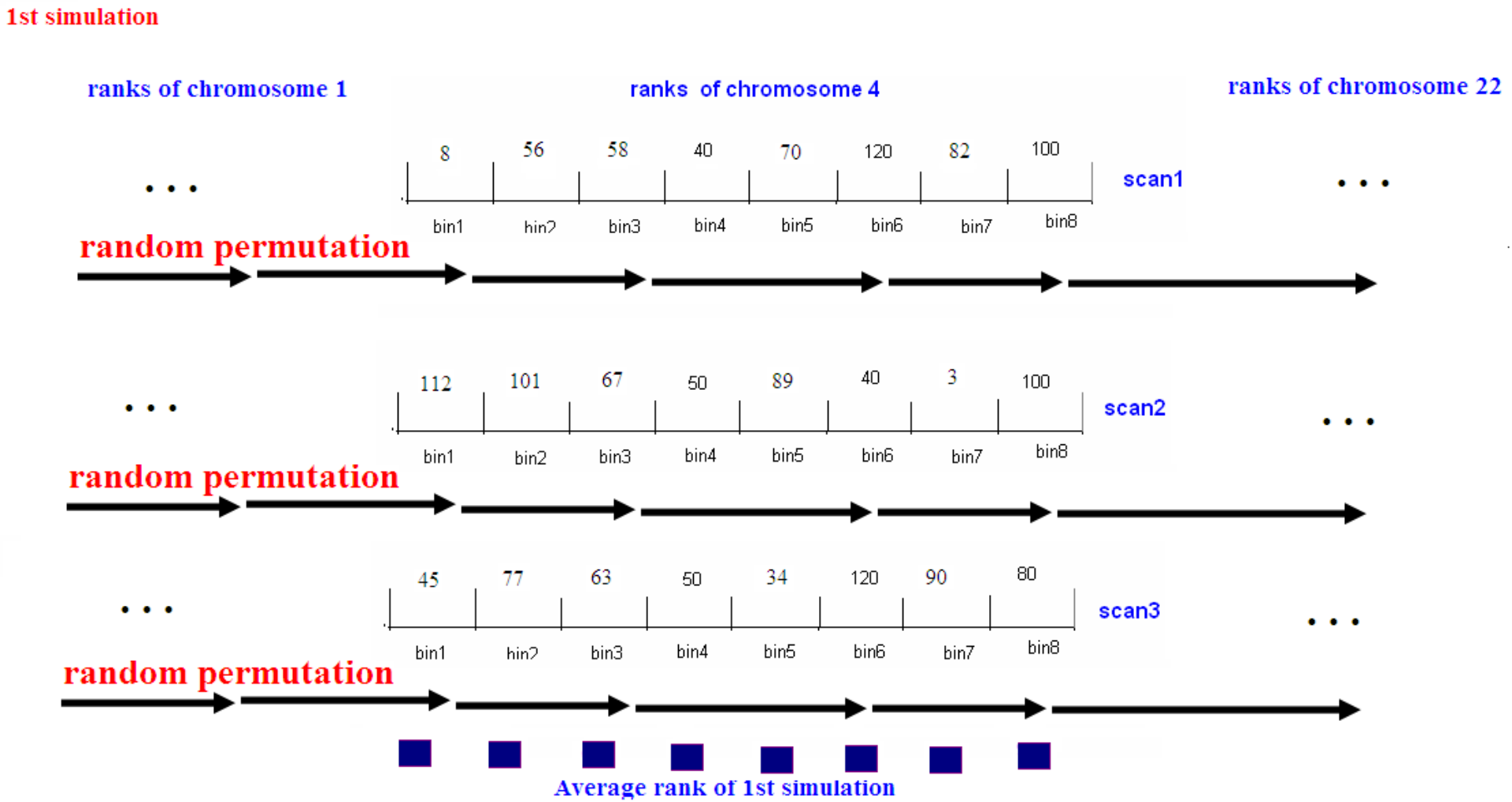
$$R=\Sigma Ri/s$$

Ri is the rank of a bin for study i (i=1 to s studies)

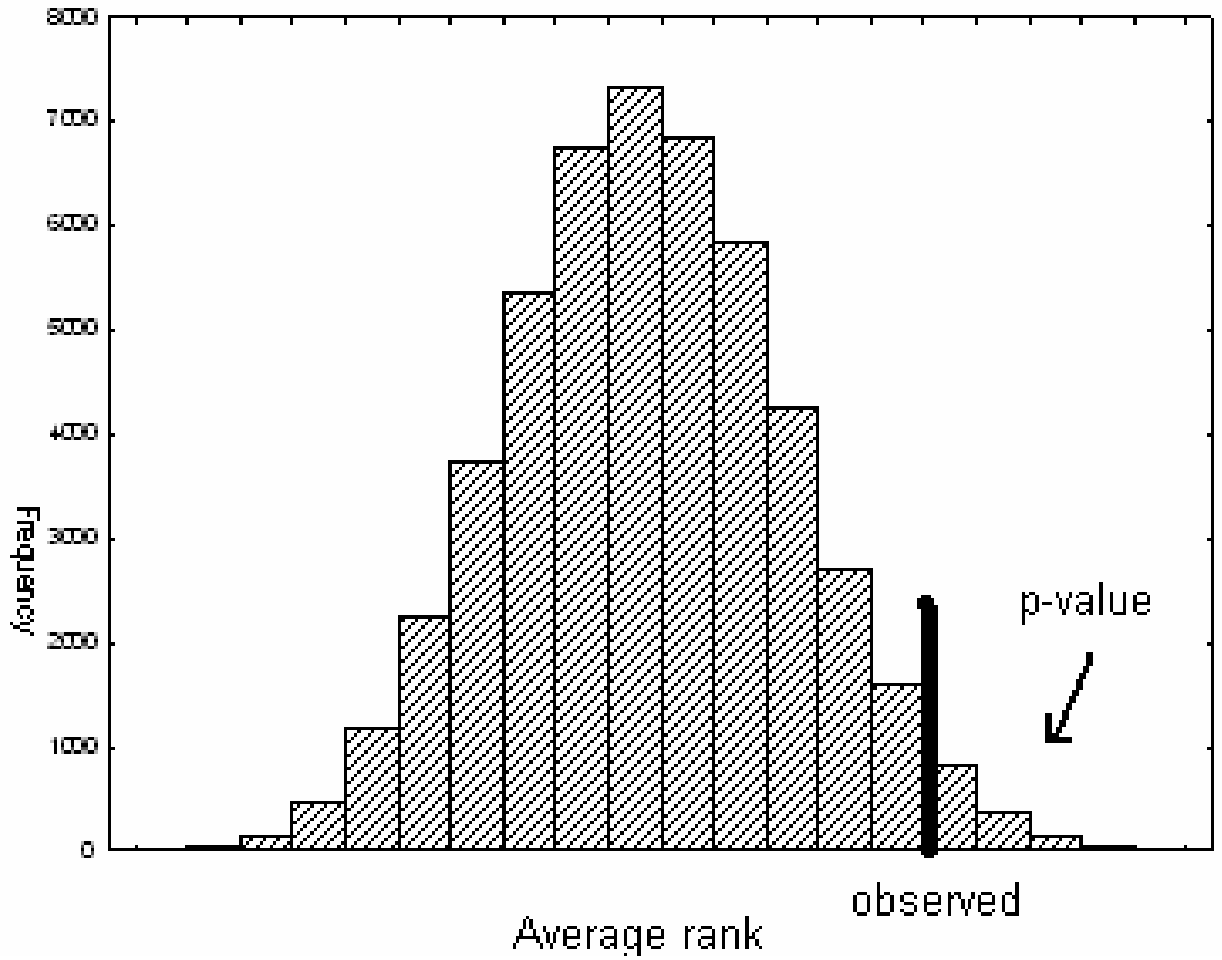
The method


ranks of chromosome 1
ranks of chromosome 4
ranks of chromosome 22
8 56 58 40 70 120 82 100
scan1
bin1 bin2 bin3 bin4 bin5 bin6 bin7 bin8
112 101 67 50 89 40 3 100
scan2
bin1 bin2 bin3 bin4 bin5 bin6 bin7 bin8
45 77 63 50 34 120 90 80
scan3
bin1 bin2 bin3 bin4 bin5 bin6 bin7 bin8
55 78 63 47 64 93 58 93
Average rank

The significance of the average rank R of each bin is assessed against the distribution of average ranks using a Monte-Carlo permutation method.



The significance of the averaged bin ranks were assessed against the null distribution using a large number of random permutations.



When a specific bin has a high average rank then there is evidence for the importance of this bin for linkage with the disease

Identification of chromosomal regions linked to premature MI using HEGESMA

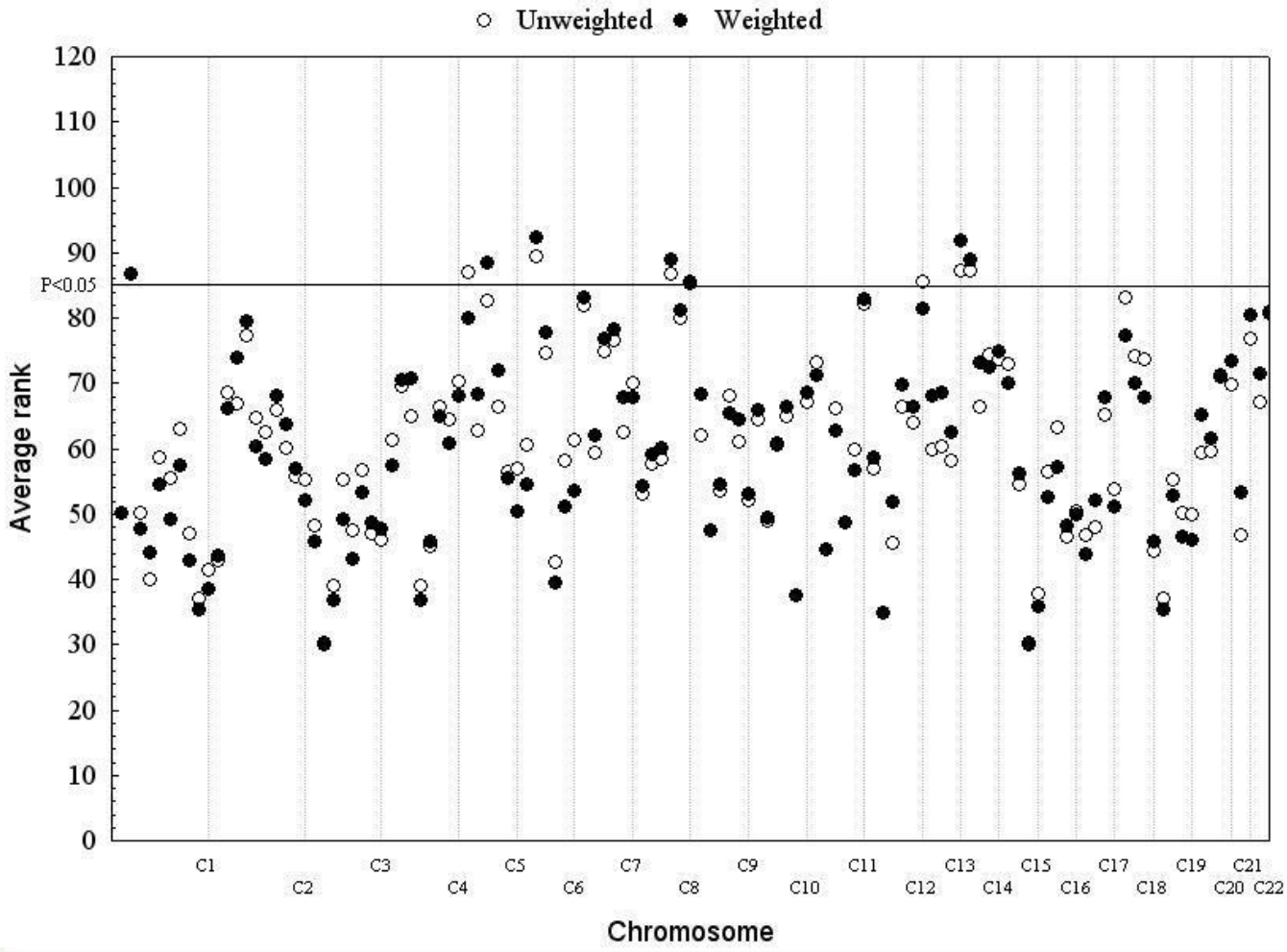
(Zintzaras and Kitsios 2006, J Hum Genet)

- **Characteristics and major results of premature Myocardial Infarction (MI) genome-scans**

First Author, publication year	Ethnicity	Pedigrees	Age (years)	Markers	Weighting	Location with evidence of linkage
Broeckel, 2002	Germany	513	(51.6-56.1)± (0.4-2.4)	394	0.22	14q
Hauser, 2004	Mixed (93.1% Caucasians)	228	<51 for males, <56 for females	395	0.15	1q25, 3q13, 19p13
Wang, 2004	Caucasians	428	44.4±9.7	408	0.2	1p34-36, 2p11, 4q32, 5p14, 7q22, 12q24, 13q32, 14q24
Helgadottir, 2004	Iceland	93	Early onset	1068	0.28	13q12-13
Samani, 2005	UK	847	50.7±7.9 for males 53.4±7.8 for	398	0.15	2p11.2 – 2q21.1

Unweighted (open circle) and weighted for the number of pedigrees (filled circle) average ranks from five premature Myocardial Infarction genome-scans with 120 bins.

Bins with significant *Prank* in unweighted or weighted analysis are above the line at P<0.05.



The meta-analysis replicated the significance of 4 already reported chromosomal regions:

- **Bin 13.4 (13q33.1-13q34): F7, F10 and IRS2 (CAD)**
- **Bin 5.1 (5p15.33-5p15.1): MTRR**
- **Bin 1.2 (1p36.21-1p35.2): MTHFR and ECE1**
- **Bin 12.6 (12q24.31-12q24.33): SCARB1**

Four new candidate regions were identified:

- **Bin 8.6 (8q24.21-8q24.3): CYP11B2**
- **Bin 8.4 (8q13.2-8q22.2): a novel region**
- **Bin 6.2 (6p22.3-6p21.1): MOG, HSPA1A, LTA, TNF, AGER, HFE, HLA-DR and C4**
- **Bin 14.1(14p13-14q13.1): CAQ14 and PSMA6**

Practice

The ranks of 3 WGS in preeclampsia are as follows:

aa	Chrom	Bin	Scan1	Scan2	Scan3	aa	Chrom	Bin	Scan1	Scan2	Scan3	aa	Chrom	Bin	Scan1	Scan2	Scan3
1	1	1.1	33.5	59	94	41	5	5.5	88	98	42	81	12	12.3	33.5	70	42
2	1	1.2	33.5	60	42	42	5	5.6	33.5	63	105	82	12	12.4	101	25	42
3	1	1.3	83	25	42	43	6	6.1	33.5	50	42	83	12	12.5	94	86	42
4	1	1.4	78	25	42	44	6	6.2	33.5	109	42	84	12	12.6	33.5	25	110
5	1	1.5	33.5	107	42	45	6	6.3	33.5	97	42	85	13	13.1	103	25	92
6	1	1.6	69	84	112	46	6	6.4	33.5	116	42	86	13	13.2	81	25	42
7	1	1.7	33.5	87	98	47	6	6.5	110	108	96	87	13	13.3	33.5	25	42
8	1	1.8	33.5	65	42	48	6	6.6	96	118	106	88	13	13.4	33.5	105	42
9	1	1.9	33.5	88	42	49	7	7.1	104	25	118	89	14	14.1	33.5	75	90
10	1	1.10	33.5	69	42	50	7	7.2	93	53	42	90	14	14.2	33.5	78	42
11	2	2.1	33.5	25	93	51	7	7.3	33.5	25	42	91	14	14.3	33.5	100	42
12	2	2.2	87	25	42	52	7	7.4	84	51	42	92	14	14.4	99	25	42
13	2	2.3	33.5	64	42	53	7	7.5	77	25	42	93	15	15.1	33.5	95	91
14	2	2.4	33.5	114	42	54	7	7.6	86	68	104	94	15	15.2	33.5	81	42
15	2	2.5	33.5	102	42	55	8	8.1	33.5	76	42	95	15	15.3	111	99	42
16	2	2.6	33.5	119	107	56	8	8.2	73	111	42	96	15	15.4	113	25	42
17	2	2.7	33.5	104	42	57	8	8.3	74	94	42	97	16	16.1	75	25	42
18	2	2.8	33.5	115	42	58	8	8.4	33.5	25	42	98	16	16.2	76	74	42
19	2	2.9	33.5	93	42	59	8	8.5	33.5	25	42	99	16	16.3	33.5	25	42
20	2	2.10	109	96	116	60	8	8.6	33.5	25	102	100	16	16.4	33.5	25	101
21	3	3.1	33.5	25	42	61	9	9.1	33.5	25	117	101	17	17.1	33.5	25	42
22	3	3.2	33.5	57	108	62	9	9.2	68	25	42	102	17	17.2	33.5	25	42
23	3	3.3	71	25	42	63	9	9.3	33.5	73	42	103	17	17.3	33.5	25	42
24	3	3.4	70	25	42	64	9	9.4	33.5	113	42	104	17	17.4	33.5	25	109
25	3	3.5	89	77	42	65	9	9.5	33.5	103	42	105	18	18.1	117	91	113
26	3	3.6	98	62	42	66	9	9.6	107	90	42	106	18	18.2	114	25	42
27	3	3.7	33.5	25	42	67	10	10.1	33.5	72	97	107	18	18.3	91	106	42
28	3	3.8	33.5	52	119	68	10	10.2	33.5	92	42	108	18	18.4	33.5	56	42
29	4	4.1	105	25	84	69	10	10.3	116	67	42	109	19	19.1	67	61	100
30	4	4.2	100	25	42	70	10	10.4	119	66	42	110	19	19.2	33.5	25	42
31	4	4.3	80	25	42	71	10	10.5	106	25	42	111	19	19.3	33.5	25	42
32	4	4.4	95	25	42	72	10	10.6	102	25	95	112	19	19.4	33.5	25	99
33	4	4.5	92	25	87	73	11	11.1	33.5	25	111	113	20	20.1	90	25	42
34	4	4.6	82	80	103	74	11	11.2	108	58	42	114	20	20.2	85	54	42
35	4	4.7	79	110	42	75	11	11.3	115	25	42	115	20	20.3	33.5	55	42
36	4	4.8	72	89	114	76	11	11.4	118	117	42	116	20	20.4	33.5	25	120
37	5	5.1	33.5	25	42	77	11	11.5	33.5	120	42	117	21	21.1	33.5	85	89
38	5	5.2	33.5	25	88	78	11	11.6	33.5	101	42	118	21	21.2	33.5	79	42
39	5	5.3	33.5	83	42	79	12	12.1	33.5	25	115	119	22	22.1	112	25	85
40	5	5.4	97	112	86	80	12	12.2	33.5	71	42	120	22	22.2	120	82	42

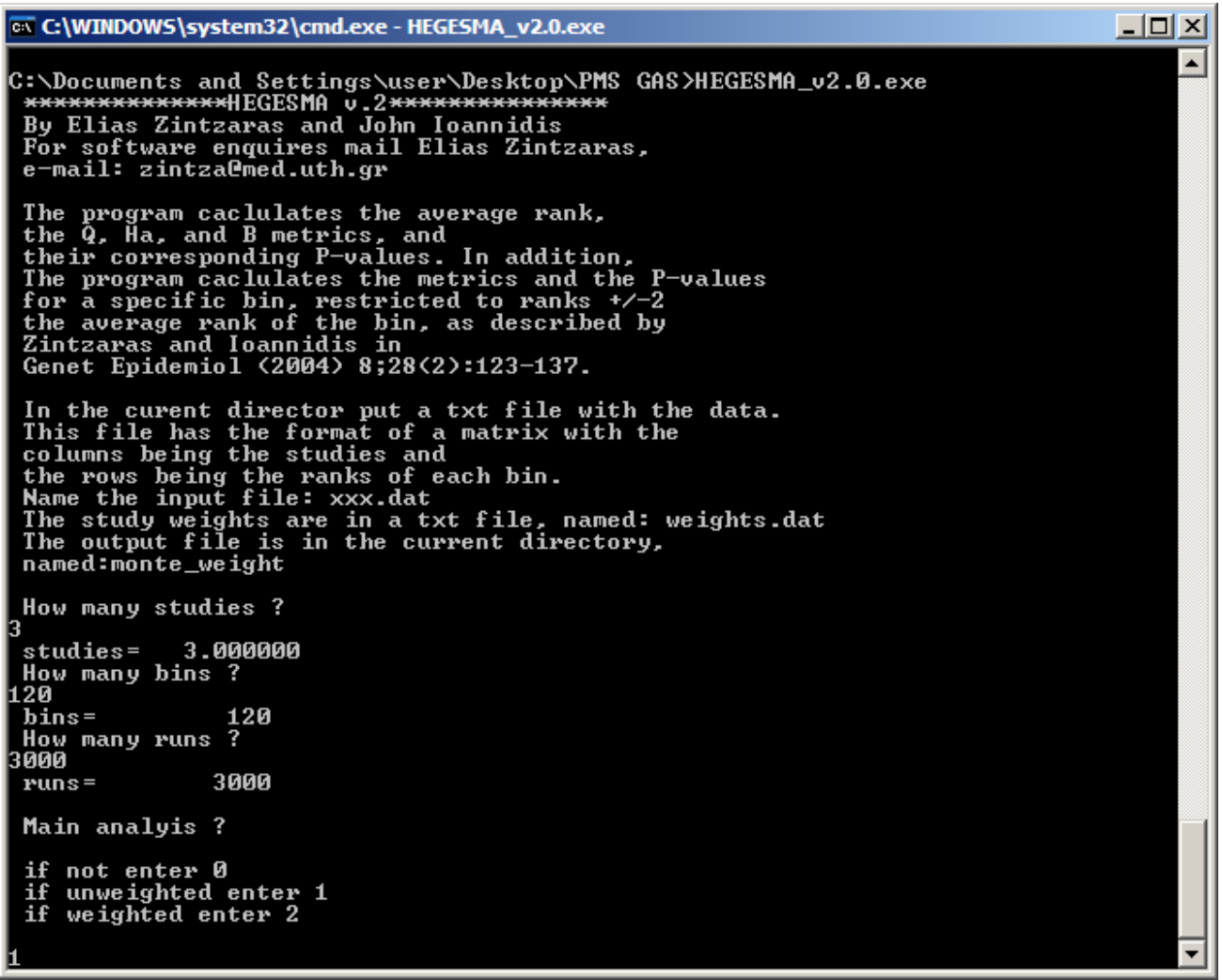
Identify significant bins linked to preeclampsia

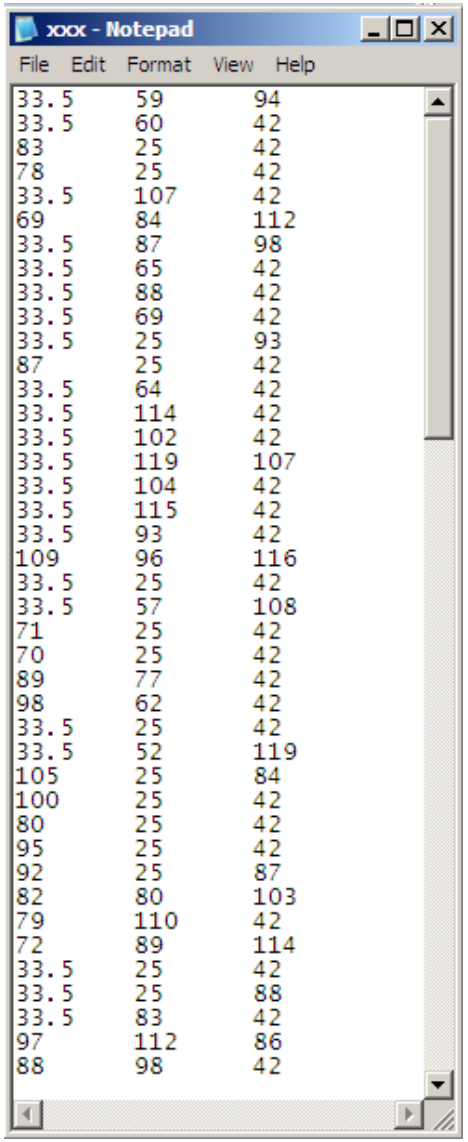
Download the file "HEGESMA_v2.0.exe" from the URL http://biomath.med.uth.gr/

Create the file "xxx.dat" with the input data and put in the same directory with the "HEGESMA_v2.0.exe" file

Then, open the DOS prompt using the "cmd" command

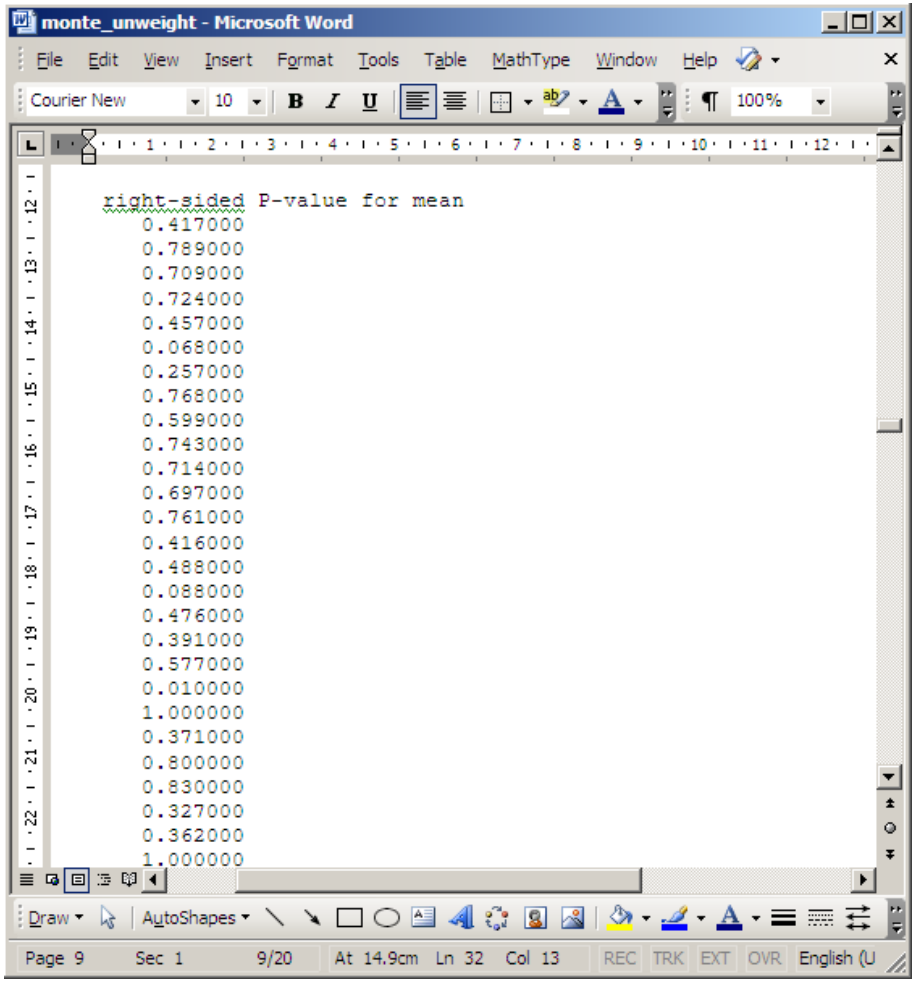
Execute the program "HEGESMA_v2.0.exe"





The output file is named “monte_unweight”

In order to open the file “monte_unweight”, open the Word and then, open the file.



Bins with P<0.05 are linked to preeclampsia

Microarrays

DNA microarrays consist of probes for measuring the expression of thousands of genes in cases and controls.

Over- or under-expressed genes may play a role in disease pathogenesis.

The analysis of microarray data involve:

i) **transformation of the data (e.g. log-transformation and normalization-removing systematic effects and bringing data from different microarrays onto a common scale).**

ii) **assessment of whether there is differential expression between diseased and normal tissue using a generalization of the t-test to adjust for multiple comparisons.**

iii) **Calculation of a P-value (false positive rate) or a Q-value (false discovery rate)**

(Q=(Pxn)/i, n=number of genes, i=sorted rank of P-value)

Inferences are limited by

i) small sample sizes and

ii) inconsistent results across studies.

iii) weakness of associations for each single gene

vi) datasets obtained with different

a) experimental conditions,

b) platforms,

c) analysis techniques,

d) types of samples (e.g. different tissue, different treatment conditions, and even different species).

v) genes under study may also overlap, but not be identical across studies.

METRADISC

(Zintzaras and Ioannidis 2008 Comp Biol Chem)

METRADISC is a generalized meta-analysis method for combining information across microarrays datasets

Average metric

The average rank (R*) for each gene expression across studies is calculated based on the adjusted ranks *(adjusted by the maximum number of tested genes in any of the combined studies).*

$$R^* = (\Sigma R_i)/s$$

where R_i is the rank of the gene under investigation for study i (i=1 to s studies).

Monte Carlo permutation test

The statistical significance of the metric R* is assessed using a Monte Carlo method:

i) **the ranks of each study are randomly permuted**

ii) **the simulated metric R* is calculated**

iii) **the procedure is repeated to generate null distributions for the metric R*.**

iv) **each gene is tested against the null distribution corresponding to the same class of information**

(i.e. gene with information from a number of specific studies are tested against the null distribution derived from these specific studies).

METRADISC is interested in identifying genes that have either very high average ranks (i.e. over-expressed) or very low average ranks (i.e. under-expressed).

Empirical demonstration

Seven prostate cancer microarray studies

Study	Investigator	Controls	Cases	Distinct genes
1	Dhanasekaran	22	59	6414
2	Lapointe	41	62	9756
3	Luo	9	16	5026
4	Ramaswamy	90	190	8958
5	Singh	50	52	6824
6	Welsh	9	25	7107
7	Yu	23	64	7763

- **A total of only 1863 distinct genes were common in all studies.**
- **There were 13580 different genes in the combined datasets.**
- **These genes belonged to 132 information classes**.

METRADISC results

	Average rank (R*)			
	Right-sided	right-sided	left-sided	left-sided
	P<0.0000037*	P<0.001	P<0.0000037*	P<0.001
	N=22	N=192	N=33	N=245

*P<0.05 adjusted for 13580 genes

Genomic Convergence

(Kitsios & Zintzaras 2009 Ann Hum Genet)

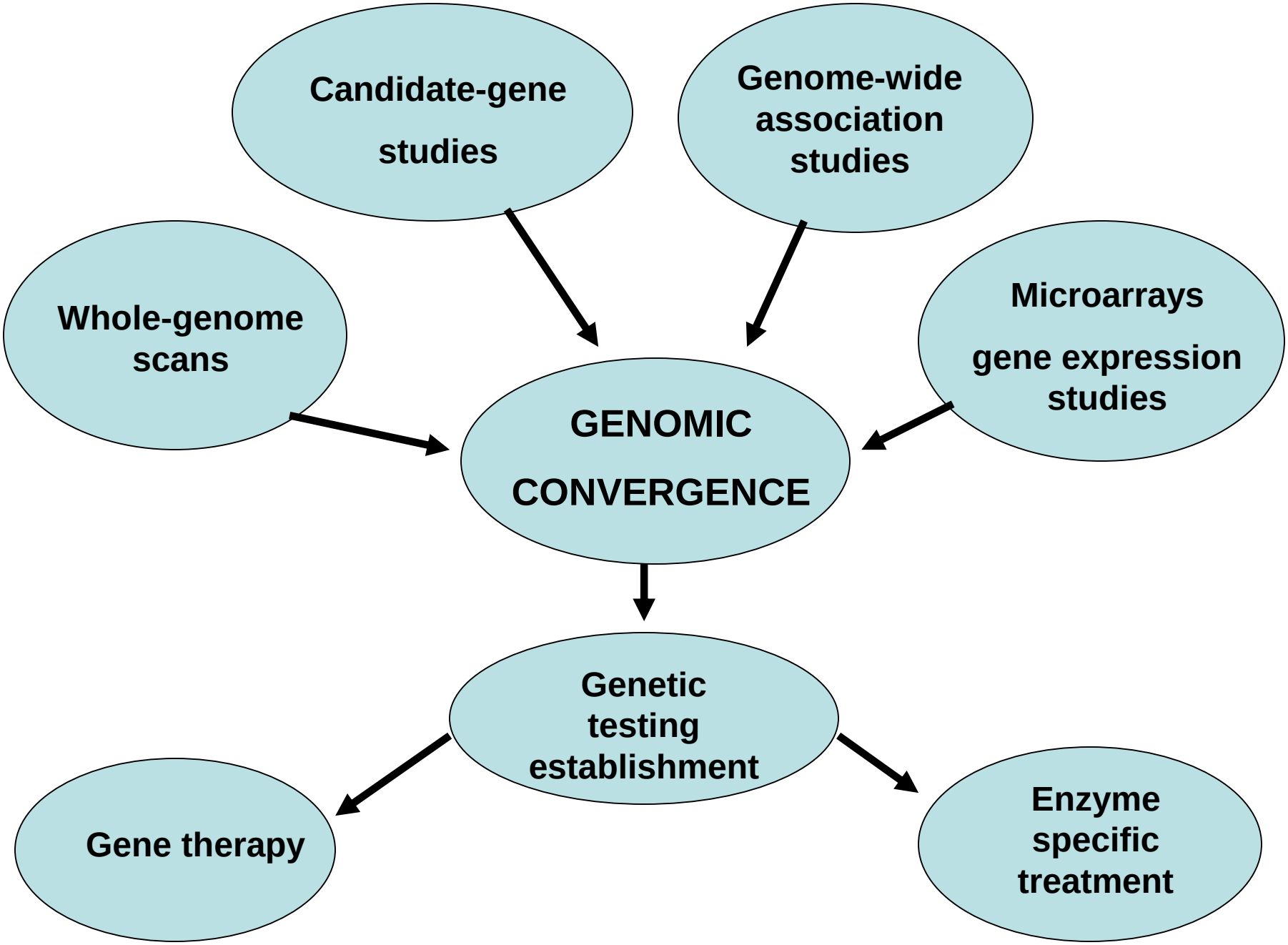
Is there Genomic Convergence?

Do all methods agree in their results?

GAS, GWAS and WGS have produced inconsistent results across them.

CAD and/or MI	Hypothesis-free		Hypothesis-driven
Approaches	**WGS**	**GWAS**	**GAS**
No of studies	**10 (6 MI, 4 CAD and MI) and 2 meta-analyses (MI, MI and CAD)**	**5 (2 MI, 3 CAD and MI)**	**1018 and 18 meta-analyses**
No of markers/genes investigated	**400 markers/scan**	**>20,000 genes (>500,000 SNPs)**	**203 genes**
No of significant findings	**40 regions showing linkage (LOD>2)**	**15 SNPs (p<10E-06)**	**17 genes (based on meta-analyses)**
Replicated findings across approaches	**9p21.3 (Chr9 at 18-48cM) with 4 SNPs:** **rs1333049 C/G (risk allele: C)** **rs10757274 A/G (risk allele: G)** **rs2383206 A/G (risk allele: G)** **rs10757278 A/G (risk allele: G)** **(Wang 2004, Samani 2007, Helgadottir 2007, McPherson 2007, WTCCC 2007)**		

Determination of genetic variants in association to disease susceptibility can only be verified when there is replication validity within and between studies of the same and different design



(Kitsios and Zintzaras 2009 Ann Hum Genet)