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.

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 <u>http://www.quantpsy.org/chisq/chisq.htm</u>):

	Br	east Cancer		H	Helathy controls					
MTHFR C677T GAS	TT	ТС	СС	TT	ТС	CC	P-value			
1	15	23	28	11	11 21		0.70			
2	43 162 110 23 92				92	118	<0.01			
3	25 85 80				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		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			

	Bre	east Cance	r	H	elathy contr	ols	
MTHFR C677T GAS	TT	ТС	CC	TT	TC	СС	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	94 242 17		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)

	Bre	east Cance	r	Н	elathy contr	ols		
MTHFR C677T GAS	TT	ТС	CC	TT	TC	СС	P-value	
1	15	5 23 28 11		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	7 38		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	$n_T = 66$	<i>n_C</i> = 57	<i>n</i> = 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 In(OR) is:
$$\theta = \ln OR = \ln \left(\frac{s_T f_C}{s_C f_T}\right)$$

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

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

Then, the 95% CI of $\theta = \ln(OR)$ is

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: $e^{\ln(x)}=x$)

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

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	<i>n_T</i> = 66	<i>n_C</i> = 57	<i>n</i> = 123

The OR of BC when homozygous mt/mt relative to wtcarrier (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	<i>s</i> = 26
wt-carrier (mt/wt+wt/wt)	$f_{T}^{} = 51$	f _C = 46	f = 97
Total	<i>n_T</i> = 66	<i>n_C</i> = 57	<i>n</i> = <i>123</i>

$$\theta = \ln(OR) = \ln(1.23) = 0.21$$
$$var(\theta) = \frac{1}{s_{T}} + \frac{1}{s_{C}} + \frac{1}{f_{T}} + \frac{1}{f_{C}} = \frac{1}{15} + \frac{1}{11} + \frac{1}{51} + \frac{1}{46} = 0.20$$
$$SE(\theta) = \sqrt{var(\theta)} = \sqrt{0.20} = 0.45$$

Then, the 95% CI of θ =In(OR) is

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

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

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	Ca	ISES	Con	trols	RESULTS			
GAS	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	24 342		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	Ca	ISES	Con	trols	RESULTS			
GAS	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 2.99	
5	32	186	17	147	1.59	0.84		
6	7	88	24	342	1.15	0.48	2.75	
7	343	1312	196	1160	1.74	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	Ca	ises	Con	trols	RESULTS			
GAS	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	Ca	ises	Con	trols	RESULTS			
GAS	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	13 298		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 metaanalysis. Also, data to be used in SPSS for analysis later are shown (blue and green color).

X	Microsoft Excel	- META_OR	GAS																_ 8 ×
: 🗷	<u> </u>	ew <u>I</u> nsert	F <u>o</u> rmat	<u>T</u> ools <u>D</u> a	ta <u>W</u> indov	v <u>S</u> tatsDire	ect <u>H</u> elp	Ado <u>b</u> e PDF									Type a que	stion for help	- 8 ×
	ı 🖄 🖄 🖾 🦻	b b 13	3		Reply with	n Changes	End Revie	N 📕 🗄 🗋	💕 📕 I	> 🖪 🔒	1 🛋 🖪	ABC 🛍	ሯ 🗈 🖺	- 🎸 🔊	- (1 - 1	😣 Σ -	<u>2</u> ↓ <u>2</u> ↓ <u>111</u> <u>3</u> 90%	- 🕜 📃	
				Times New	Roman	• 11 • I	B 7 I		=	\$ %	e.0 .00		- 1 - 8 - 1	A					
	AE25	fv		,			D 1 <u>1</u>			φ 70 ×	.00 ->.0 :		· · · · ·	- · F					
	A	B	с	D	E	F	G	н		J	К	L	м	N	0	Р	Y	AE	
			_				_												-
1	n=	11																	
2			Ca	ases	Cor	itrols		RESULTS		OU	TPUT to b	e used for M	leta-regres	sion and P	ublication	bias			
	aa	Study	mt/mt	Total	mt/mt	Total	OR	95%LL	95%UL	$\theta = \ln(OR)$	v	w=1/v	SE=√v	1/SE	x=√w	y=0√w	t^2		
3	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	(1) t^2<0 then t^2=0)		
5	2	study1 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	/	study/	343	312	190	410	1.74	0.47	2.12	0.55	0.010	99.14	0.100	9.957	9.957	5.521			
12	9	studya study9	27	334	50	551	0.71	0.47	1.07	-0.13	0.045	16.05	0.208	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						REOR	1.35	1.08	1.69										
19																			
20																			
21																			
22	FE OK	1.350																	
24	95% UL	1.711																	
25																			
26	RE OR	1.352																	
27	95% LL	1.081																	
29	3376 62	1.052																	
30	P-value for Q	0.002																	
31																			
32																			
34																			
35																			
36																			
H.	(→ → \ Sheet1	/								_			•						
) D <u>r</u>	aw 👻 🔓 A <u>u</u> toS	hapes 🔹 🔨	$\mathbf{X} \square ($) 🔄 🐗	्रि 🚨 🖉	🛯 🌺 🗕 🚪	🖉 🗕 🗸 🗸	≡≣≣		Ŧ									
Rea	dv																		
					-						0.5	1 10	1			-			
	Start COS	SMOTE In	Inbox	- Micro	PMS GAS	<u> </u>	Meta-anal	/si 🛛 🗶 🖡	licrosoft E.	- Micro	soft Exc	😗 untitle	d - Paint	🗳 © "I	1 🕖 🚅 🧐		/ X X X X 9 % (💆 💕 🐞 2408 🧭	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.

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

MTHFR	HFR Cases		Con	trols	RESULTS			
C677T GAS	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.



OR (95% CI)

Practice in SPSS

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

File Edit	ed - SPSS Data View Data	a Editor Transform Ar	nalyze Graph	ns Utilities W	indow Help				
	4 🖭 🗠	a 💷 🖁	- 12 44	·[] [] []	 ata ⊞ ®	୬ାଭା			
14:1									
	study	n_o		ul	var	var	Var	Var 📶 Chart1 - SP55 Chart Editor	
1	study1	1.23	.51	2.95				File Edit View Gallery Chart Series Format Analyze Graphs Help	
2	study2	1.44	.84	2.47				늘 》 ☆ 囫 ☜ ★ ▲ 凾 〃 ▾ ᆾ Ւ ♂ ↘ 옷 感	
	study3	2.29	1.11	2.09					
	study5	1.40	8/	2.03					
6	study6	1.55	48	2.33				+	
7	study7	1.13	1.43	2.13					
8	study?	.71	.47	1.07					
9	study9	.88	.54	1.44					
10	zstudy10	.78	.32	1.91				Study I	
11	zstudy11	1.87	1.52	2.30				study2	
12	1							study3	
13		- f: f:l- l	utale tana Ch	C		- h - M - vi- h l			
14		enne Simple i	nigh-Low-Cia	ose: Summari	es of Separ				
15	(*	OR [o_r]	Da	ars Represent			ок		
16		95%LL [I]		🕞 High: 🧃	MEAN(95%	UL [ul])	Parte	study6	
17		95%UL [ul]					Taste	study7	
18				Low:	MEAN(95%		Reset	ctudu 8	
19				Close:		D	Cancel		
20				<u> </u>	MEAN(OR]	0_1)	Help	study9	
21					Change Sum	mary		study10	
22				Category	Avis:				
23				Call GAS	[study]	_	-		
24			Te	emplate					
20				Use chart spe	cifications from	n:			
20	<u></u>			File				OR (95% CI)	
28									
29					Titles	Options			
30							-		
31									
32	!							SPSS Processor is ready	
	ata View 🖉 Va	riahle View 1							
ISPSS Processor is ready									
者 Start									
Juli			··· •						

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:

$$\theta_{\mathbf{p}} = \frac{\sum_{i=1}^{n} \mathbf{w}_{i} \theta_{i}}{\sum_{i=1}^{n} \mathbf{w}_{i}} = \frac{\mathbf{w}_{1} \theta_{1} + \dots + \mathbf{w}_{n} \theta_{n}}{\mathbf{w}_{1} + \dots + \mathbf{w}_{n}}$$

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

$$\boldsymbol{\theta}_{\mathbf{p}} = \frac{\sum_{i=1}^{n} \mathbf{w}_{i} \boldsymbol{\theta}_{i}}{\sum_{i=1}^{n} \mathbf{w}_{i}} = \frac{\mathbf{w}_{1} \boldsymbol{\theta}_{1} + \dots + \mathbf{w}_{n} \boldsymbol{\theta}_{n}}{\mathbf{w}_{1} + \dots + \mathbf{w}_{n}}$$

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

$$\mathbf{w}_{i} = \frac{1}{\mathbf{v}_{i}}$$

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

The standard error of $\theta p = \ln(ORp)$ is given by

$$\mathbf{SE}(\boldsymbol{\theta}_{\mathbf{p}}) = \sqrt{\frac{1}{\sum_{i=1}^{n} \mathbf{w}_{i}}}$$

The 95% CI for $\theta p = \ln(ORp)$ 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 $\theta p = ln(ORp)$,

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

$$\mathbf{OR}_{\mathbf{p}} = \mathbf{e}^{\mathbf{\theta}_{\mathbf{p}}} = \mathbf{e}^{\ln(\mathbf{OR}_{\mathbf{p}})}$$

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: $e^{\ln(x)}=x$)

$$\left(\boldsymbol{e}^{\boldsymbol{\theta}_{p}-1.96*SE\left(\boldsymbol{\theta}_{p}\right)}, \, \boldsymbol{e}^{\boldsymbol{\theta}_{p}+1.96*SE\left(\boldsymbol{\theta}_{p}\right)}\right)$$

MTHFR C677T	Cas	es	Controls		RESULTS			
GAS	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 $\theta p = ln(ORp)$ is

$$\Theta_{p} = \frac{\sum_{i=1}^{8} w_{i} \Theta_{i}}{\sum_{i=1}^{8} w_{i}} = \frac{w_{1} \Theta_{1} + \dots + w_{11} \Theta_{11}}{w_{1} + \dots + w_{11}} = \frac{5.03 \times 0.21 + \dots + 89.08 \times 0.63}{5.03 + \dots + 89.08} = 0.43$$

MTHFR C677T	Cas	es	Controls		RESULTS			
GAS	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 $\theta p = ln(ORp)$ is

$$\Theta_{p} = \frac{\sum_{i=1}^{8} w_{i} \Theta_{i}}{\sum_{i=1}^{8} w_{i}} = \frac{w_{1} \Theta_{1} + \dots + w_{11} \Theta_{11}}{w_{1} + \dots + w_{11}} = \frac{5.03 \times 0.21 + \dots + 89.08 \times 0.63}{5.03 + \dots + 89.08} = 0.43$$

MTHFR C677T	Cases		Controls		RESULTS			
GAS	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 $\theta p = \ln(ORp)$ is

se(
$$\theta_{p} = \sqrt{\frac{1}{\sum_{i=1}^{11} w_{i}}} = \sqrt{\frac{1}{w_{1} + ... + w_{11}}} = \sqrt{\frac{1}{5.03 + ... + 89.08}} = 0.057$$

The 95% CI of $\theta p = \ln(ORp)$ is

$$(\theta_{p} - 1.96 \text{se}_{\theta_{p}} , \theta_{p} + 1.96 \text{se}_{\theta_{p}}) =$$

=(0.429-1.96×0.057, 0.429+1.96×0.057)=

=(0.32, 0.54)

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

$$\left(\mathbf{e}^{\theta_{p} - 1.96^{*} SE(\theta_{p})}, \, \mathbf{e}^{\theta_{p} + 1.96^{*} SE(\theta_{p})} \right) = \left(\mathbf{e}^{\ln(OR_{p}) - 1.96^{*} SE}, \, \mathbf{e}^{\ln(OR_{p}) + 1.96^{*} SE} \right) = \\ = \left(\mathbf{e}^{0.32}, \, \mathbf{e}^{0.54} \right) = \left(1.37, \, 1.71 \right)$$

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.

Microsoft Excel - META_OR GAS													_ 8 ×						
:2	<u>Eile E</u> dit <u>V</u>	jew <u>I</u> nsert	F <u>o</u> rmat	<u>T</u> ools <u>D</u> a	ta <u>W</u> indov	/ <u>S</u> tatsDire	ct <u>H</u> elp	Ado <u>b</u> e PDF									Type a que	stion for help	- 8 ×
-	i 🖆 🖄 🖾 🤅	o 🛯 🖉	610	🖦 😥 i V	Reply with	<u>Changes</u>	End Review	w 🔲 🗄 🗋	🖻 🔛 I	> 🖪 🔒		ABC 🛍	X 🗈 🖺	- 🍼 🔊	- (21 - 1	😣 Σ -	≜↓ Ž↓ 🏨 📣 90%	- 💿 📘	
				Times New	Roman	- 11 - 1	в 7 т		= =	\$ %	•.0 .00 -		- 29 - 1	A					
	AE25 -	£		, nines iven	Komon	• •	D 1 <u>(</u>	2 = =	-=	φ /0 ^γ	.00 ->.0 =		· <u> </u>	- . E					
_	AEZ5 V	<i>jx</i>	<u> </u>	D	F	E	G	L			K	1	NA	N	0	D	v	٨E	
			<u> </u>		L .		9			,	N	L.	IVI	IN	0	r -	1	AL	-
1	n=	11																	
2			C .	CAC	Cor	trole		RESULTS		01	TPUT to b	a used for \	fets regres	sion and P	ublication l	iac			
-	aa	Study	mt/mt	Total	mt/mt	Total	OR	95%LL	95%UL	$\theta = \ln(OR)$	v	w=1/v	SE=√v	1/SE	x=√w	v=θ√w	t^2		
3																· ·	(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			
	3	study3	20	514	65	193	1.49	1.11	2.00	0.85	0.135	32.54	0.307	5.705	5.705	2.250			
8	5	study-	32	186	17	147	1.59	0.84	2.99	0.35	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	0.055		
15						FT OP	1.52	1.27	1 71			304.97					0.075		
10						PEOR	1.55	1.57	1./1										
18						KL OK	1.55	1.00	1.09										
19																			
20																			
21		1.000																	
22	FE OK	1.530																	
23	95% UL	1.307																	
25																			
26	RE OR	1.352																	-
27	95% LL	1.081																	
28	95% UL	1.092																	
30	P-value for Q	0.002																	
31																			
32																			
33																			
35																			
36																			
14 4	→ → \Sheet	1/			-					<u> </u>					l				
Draw • λ AutoShapes • λ • \Box \bigcirc \Box \checkmark \blacksquare \blacksquare \blacksquare \blacksquare \blacksquare \blacksquare \blacksquare \blacksquare \blacksquare																			
Read	dy																		
<u>a</u>	start 🚺 CO	SMOTE In	🕒 Inbox	- Micro	🚞 PMS GAS		Meta-analy	ysi 💌 M	licrosoft E.	📧 Micro	soft Exc	🦉 untitled	d - Paint	EN 🖸 "1	0 🗳 🛞	V 🔀 🌆 () * * * * * * *	2468 💓	8:45 PM

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

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):

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

The Q-statistic is a weighted sum of squares of the deviations of individual θ i's, θ i=In(ORi), from the pooled estimate θ p.

When the θ i'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

	Ca	ses	Co	ntrols			RESULTS	
MTHFR C677T GAS	mt/mt	Total	mt/mt	Total	$\theta = \ln(OR)$	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 $\theta p = \ln(ORp) = 0.43$ The heterogeneity Q-statistic is:

$$Q = \sum_{i=1}^{11} w_i (\theta_i - \theta_p)^2 = w_1 (\theta_1 - \theta_p)^2 + ... + w_{11} (\theta_{11} - \theta_p)^2 = 5.03 (0.21 - 0.43)^2 + ... + 89.08 (0.63 - 0.43)^2 = 27.87$$

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).

	P value	. •	.		~ ^			
	<i>P</i> v	alue		P	value		P va	alue
<u>d.f.</u>	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

Table. Percentage points of the χ^2 distribution.

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.

There are a second and the second an																			
:2	<u>File E</u> dit <u>V</u>	iew <u>I</u> nsert	F <u>o</u> rmat	<u>T</u> ools <u>D</u> at	ta <u>W</u> indow	<u>S</u> tatsDire	ct <u>H</u> elp	Ado <u>b</u> e PDF									Type a que	stion for help 🔸	_ 8 ×
-	🔁 🖄 🖾 🖣	o 🖄 🖾	8	Ba 🔂 (V	Reply with	Changes	End Review	🗋 i 🖻	pa 🖂 🛛	> 🖪 🖨		ABC 1	X 🗈 🕰	- 🍼 🔄	- (21 →]	🔍 Σ 🔻	A ↓ A ↓ M A 90%	- 💿 🗌	
-				Times New	Poman		- • 7 11			• •/ •	•.0 .00 j z		- ^ - /			69			
	AE06 -	2		, nines new	Kullan	• 11 •	b 1 <u>0</u>		-=	¢ 70 7	.00 →.0 ≧		• • • 4	- • -					
	AE20	/x	0	D	F	E	6	ш	1	1	V	1	M	N	0	D	v	٨E	
-	A		U U		E	F	6		1	J	N	L	IVI	IN	0	r	, i i i i i i i i i i i i i i i i i i i	AL	<u> </u>
1	n=	11																	
2 Cases Controls RESULTS OUTPUT to be used for Meta-regression and Publication bias																			
-	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		
3																· ·	(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 study3	43	315	23	233	1.44	0.84	2.47	0.37	0.075	7.41	0.274	3.047	3.047	2.250			
7	4	study5	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	studyб	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 etudy10	21	233	50	208	0.88	0.54	1.44	-0.13	0.002	10.05	0.250	2 183	2 183	-0.507			
14	11	study10	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																			
20																			
21																			
22	FE OR	1.530																	
23	95% LL 95% UL	1.307																	
25																			
26	RE OR	1.352																	
27	95% LL 95% TT	1.081																	
29	3376 CL	1.092																	
30	P-value for Q	0.002																	
31																			
33																			
34																			
35																			
36																			
	► ► Sheet:	1/										Ŀ							
Dra	w 🕶 🔓 A <u>u</u> tos	Shapes 🔹 🔨	$\mathbf{X} \square ($) 🗠 🖪	🕄 🚨 🖉	🛯 🖄 🕶 🚽	🖉 • <u>A</u> •	= = ₹		-									
Read	ły																		
樻 s	itart 🛛 🚺 CO	SMOTE In	🕒 Inbox	- Micro	🚞 PMS GAS	P	Meta-analy	si 💌 M	licrosoft E.	📧 Micro	soft Exc	🦉 untitled	l - Paint	ان 🖸 🖪	1) 🗳 🍪	🔁 🔀 🌆	0 <u>14 14 14 0</u> 8 (3 🔛 🔏 2468 🥘	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 θ i's:

$$\boldsymbol{\theta}_{\mathbf{p}}^{*} = \frac{\sum_{i=1}^{n} \mathbf{w}_{i}^{*} \boldsymbol{\theta}_{i}}{\sum_{i=1}^{n} \mathbf{w}_{i}^{*}} = \frac{\mathbf{w}_{1}^{*} \boldsymbol{\theta}_{1} + \dots + \mathbf{w}_{n}^{*} \boldsymbol{\theta}_{n}}{\mathbf{w}_{1}^{*} + \dots + \mathbf{w}_{n}^{*}}$$

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



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

The τ² is given by the (DerSimonian and Laird) formula:

$$\tau^{2} = \frac{\mathbf{Q} - (\mathbf{n} - 1)}{\sum_{\mathbf{i}=1}^{\mathbf{n}} \mathbf{w}_{\mathbf{i}} - \sum_{\mathbf{i}=1}^{\mathbf{n}} \mathbf{w}_{\mathbf{i}}^{2} / \sum_{\mathbf{i}=1}^{\mathbf{n}} \mathbf{w}_{\mathbf{i}}}$$

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

The standard error of $\theta_p^* = \ln(OR_p^*)$ is given by

$$\mathbf{SE}(\boldsymbol{\theta}_{\mathbf{p}}^{*}) = \sqrt{\frac{1}{\sum_{i=1}^{n} \mathbf{w}_{i}^{*}}}$$

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: $e^{\ln(x)}=x$)

$$\mathbf{OR}_{\mathbf{p}}^{*} = \mathbf{e}^{\mathbf{\theta}_{\mathbf{p}}^{*}} = \mathbf{e}^{\ln(\mathbf{OR}_{\mathbf{p}}^{*})}$$

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:

$$\left(\boldsymbol{e}^{\boldsymbol{\theta}_{p}^{*}-1.96*SE\left(\boldsymbol{\theta}_{p}^{*}\right)}, \boldsymbol{e}^{\boldsymbol{\theta}_{p}^{*}+1.96*SE\left(\boldsymbol{\theta}_{p}^{*}\right)}\right)$$

Example - MTHFR	C677T and	Breast	Cancer
-----------------	-----------	---------------	--------

	Ca	ses	Со	ntrols			RESULTS	
MTHFR C677T GAS	mt/mt	Total	mt/mt	Total	$\theta = \ln(OR)$	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

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

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

Example -	MTHFR	C677T	and	Breast	Cancer
-----------	--------------	--------------	-----	---------------	--------

	Ca	ses	Co	ntrols			RESULTS	
MTHFR C677T GAS	mt/mt	Total	mt/mt	Total	$\theta = \ln(OR)$	v	w=1/v	\mathbf{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
-----------------	------------------	---------------	--------

	Ca	ses	Co	ntrols			RESULTS	5
MTHFR C677T GAS	mt/mt	Total	mt/mt	Total	$\begin{array}{c} \theta \\ = \ln(OR) \end{array}$	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:

$$\theta_{p}^{*} = \frac{\sum_{i=1}^{11} w_{i}^{*} \theta_{i}}{\sum_{i=1}^{11} w_{i}^{*}} = \frac{w_{1}^{*} \theta_{1} + \dots + w_{11}^{*} \theta_{11}}{w_{1}^{*} + \dots + w_{11}^{*}} = \frac{3.66(0.21) \dots + 11.64(0.63)}{3.66 + \dots + 11.64} = 0.302$$

Example - MTHFR C677T and Breast Cancer

	Ca	ses	Co	ntrols			RESULTS	
MTHFR C677T GAS	mt/mt	Total	mt/mt	Total	$\begin{array}{c} \theta \\ = \ln(OR) \end{array}$	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(
$$\theta_{p}^{*} \stackrel{\checkmark}{=} \sqrt{\frac{1}{\sum_{i=1}^{11} w_{i}^{*}}} = \sqrt{\frac{1}{w_{1}^{*} + ... + w_{11}^{*}}} = \sqrt{\frac{1}{3.66 + ... + 11.64}} = 0.114$$

The 95% CI of $\theta_p^*=\ln(OR_p^*)$ is:

$$(\theta_p^* - 1.96 \text{se}_{\theta_p}^* , \theta_p^* + 1.96 \text{se}_{\theta_p}^*) =$$

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

=(0.079, 0.525)

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

$$\mathbf{OR}_{\mathbf{p}}^{*} = \mathbf{e}^{\mathbf{\theta}_{\mathbf{p}}^{*}} = \mathbf{e}^{\ln(\mathbf{OR}_{\mathbf{p}}^{*})} = \mathbf{e}^{0.302} = 1.352$$

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

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

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.

Icrosoft Excel - META_OR GAS																			
:	<u>F</u> ile <u>E</u> dit <u>V</u>	jew <u>I</u> nsert	F <u>o</u> rmat	<u>T</u> ools <u>D</u> a	ta <u>W</u> indov	v <u>S</u> tatsDire	ct <u>H</u> elp	Ado <u>b</u> e PDF									Type a que	stion for help	- 8 ×
-	i 🖆 🖄 🖾 🦉	o 🖄 🛛	6 0	🖦 😥 i V	Reply with	n Changes	End Review	N 📕 🗄 🗋	🖻 🔛 I	> 🖪 🔒		ABC 🛍	አ 🖻 💦	- 🍼 🔄	- (21 - 1	😣 Σ -	≜↓ Ž↓ 🏨 📣 90%	• 🕜 📘	
				Times New	Roman	- 11 -	в 7 т		= =	\$ %	4.0 .00 ↓ 4			A		35			
	AE25 -	£.		, mes nem	- Comon		D 1 <u>(</u>	2 = =		φ /0 /	.00 ->.0 =		· <u>· · ·</u>	- <u>-</u>					
			C	D	F	F	G	н			к	1	M	N	0	P	v	ΔF	
											K	-							
1	n=	11																	
2			Ca	ises	Con	itrols		RESULTS		ou	TPUT to b	e used for M	feta-regres	sion and P	ublication b	bias			
	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		
3																	(If t^2<0 then t^2=0)		
4	1	study1	15	06	11	5/	1.23	0.51	2.95	0.21	0.199	5.03	0.446	2.242	2.242	0.464			
5	2	study2 etudy3	45	100	12	103	2.20	1 11	4.47	0.37	0.075	7.41	0.274	2 722	2 722	2 250			
7	4	study5	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						FLOD	1.52	1.25	1.51			304.97					0.075		
16						FE OR	1.55	1.3/	1.71										
18		-				KL UK	1.55	1.00	1.09										
19																			
20																			
21																			
22	FE OR	1.530																	
23	95% LL 05% TT	1.50/																	
24	9376 OL	1./11																	_
26	RE OR	1.352																	—÷
27	95% LL	1.081																	
28	95% UL	1.692																	
29	D and the fam O	0.002																	
30	P-value for Q	0.002																	
32																			
33																			
34																			
35																			
36																			_
Image: A state of the state																			
$ $ Draw = $ _{\mathcal{O}}$ AutoShapes = \times \times \square \bigcirc 🖂 🐗 \bigcirc \bigcirc \bigcirc \bigcirc \bigcirc \checkmark \checkmark \bigcirc																			
Read	dy																		
đ) s	Start	SMOTE In	🕒 Inbox	- Micro	🚞 PMS GAS		Meta-analy	/si 🕱 🖡	licrosoft E.	. 💌 Micro	soft Exc	🦉 untitled	d - Paint	h, 🖸 🖪	0 🗳 🛞	🔁 🔀 🌆	∂ ₩₩₩₩ ₽ ≶(2469 🔊 🎊 💭	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.



OR (95% CI)

Practice in SPSS

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

🛄 Untitle	ed - SPSS Data	a Editor	alura Curali	- 11686 118	adam Uala													_ 8	×
	view Data		Taiyze Graph	s Utilities wi	ndow Help	<u>a</u>													—
13: gas		zRE	OR																-
1	gas	1_0		ul	var	var	var	var	var	var	var	va	Ir	var	var	var	var	var	
1	study1	1.23	.51	2.95			C	art1 - SPSS (hart Editor		•				1				1
2	study2	1.44	.84	2.47			File	Edit View G	allery Chart :	5eries Formal	t Analyze i	Graphs H	Help						
3	study3	2.29	1.11	4.69						ير ا محمد ا عد ا م		r 1 16		8	Category /	Axis		1	×I
4	study4	1.48	1.05	2.09			_ 💻				· •• ••			71 945	Display	avia line		OK	.
5	study5	1.59	.84	2.99			_											UK	!
0	study6	1.15	.40	2.75			_								Axis Title:	GAS		Cancel	
8	study/	71	47	2.12			_								Litle Justific	ation: Center	_	Help	
9	study9	.88	.54	1.44											- Axis Mark	(ers		Display labe	ls
10	xstudy10	.78	.32	1.91				study1			-				Tick n	narks 🔲 Grid	lines	Labels	i II
11	xstudy11	1.87	1.52	2.30				study2	•		╞╷╺								
12	zFE OR	1.53	1.37	1.71				study3 study4	1										
13	zRE OR	1.35	1.08	1.69				study5	ł .	\vdash	['] =				Categor	/ Axis: Labels		X	i I
14	Define 9	Simple High-L	ow-Close: Su	mmaries of S	eparate Va	iables I	×I	o study6 ≤ study7	1 ⊢				-1		- Display		I		
15			- Bars Repr	resent			1	0 study8	⊢	-	H !	- 1			• Alla	bels		Continue	
10	(#) OR [0	<u>r]</u>	П	iah: 🗛 uraa	1/052/11/15/10	ок]	study9	1						- C Ever	y 2 label	ls	Cancel	
18	∰ 95%LL	. (U) . Tull		I 🗰 MEAN	1(95%UL [ul])	Paste		tudy10	1	-					T 🖂 T	ick marks for sk	ipped labels	Help	
19				DW: 🏟 MEAN	I(95%LL [II])	Reset	i H	FEOR	•		│ ⊢∎-	1			Label T	ext			
20						Cancel	11	REOR	1			l			Labe	I: FEOR			Г
21			l 🕨 a	ose: 🛞 MEAN	I(OR [o_r])		1 F I		.3.5	.7 .9		2	4	4	Change	study10	_		
22				Change	e Summary	Help	1 []		.4	.6 .8	1		3	5		study11			
23										OR	(95% C	D				REOR	•		
24				ategory Axis:							(- Orienta	ion: Automati			
25			Template	e (and Iges)			H.								ononta	Automati			1
20	-		Use c	hart specificatio	ns from:		H											_	-
21	-		Eile	п ^і —			H								<u> </u>				-
29	- '						H					SPSS Pro	ocessor is	ready //					-
30	-			Tit	les Opt	ions									-				-
31																			1
32																			
I ► \Da	ta View 🖌 Va	riable View /																	Ľ
										SPSS Process	or is ready	ſ							_
🏄 Start	COSM	. 🕒 Εισερ	DMS G.	💌 Meta	. 🛗 Untitle	e 📧 Micros.	. 🔁 BEC	R 🛗 Out	ou 🏦 Cha	rt 🦉 unti	tle 🛛 🖪	C 🖸	ی 😂 🕹 اند (. 🛞 🚂 🖉	x x x	10 S 🗵 🤇	🕑 🄏 2827 🔛	10:35 /	AM

Practice

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

		-			AD	H2		
			*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

			Ca	ses	Cont	trols				
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.

X	Microsoft Excel -	META_OR GA	s														_	<u> a x</u>
: 2) <u>Fi</u> le <u>E</u> dit <u>V</u> iev	w <u>I</u> nsert F	F <u>o</u> rmat <u>T</u> oo	ols <u>D</u> ata	<u>W</u> indow <u>S</u> ta	atsDirect <u>H</u> el	p Ado <u>b</u> e i	PDF								Type	a question for help 🛛 🗸	- 8 ×
1	ı 🖆 🖄 🖾 👒	N 13	5 🔊 🖷	n ver Re	eply with <u>C</u> han	ges E <u>n</u> d Rev	view	D 💕 月			ABC 🛍 🕽	i 🗈 🖺 -	I - (*	(° - 1 😣	$\Sigma - \frac{A}{Z} \downarrow \frac{Z}{A}$	1 🛍 🛷	98% 🔻 🕜 📘	
			: Tin	nes New Rom	an 🚽 11	- B Z	U E		\$ % ,	÷.0 .00 ₹		- 🗞 - A						
	M27 -	fx							4 70 7	.00 9 .0 =;			Ŧ					
	A	B	С	D	E	F	G	Н	1	J	K	L	М	N	0	Р	Y	-
1		3			1													
2			Ca	ises	Con	trols		RESULTS		OU	IPUT to be	e used for M	leta-regres	sion and P	ublication	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	у=ө√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		
+	/ 11	study/	343	1312	196	1104	1.74	1.43	2.12	0.55	0.010	99.14	0.100	9.957	9.957	5.018		
9		study11	551	1400	100	1104	1.07	1.04	2.00	0.05	0.011	237.77	0.100	7.450	2.450	5.710	0.000	
10						FE OR	1.76	1.55	2.00									
11						RE OR	1.76	1.55	2.00									
12																		
13																		
14																		
16	FE OR	1 758																
17	95% LL	1.548																
18	95% UL	1.996																
19																		
20	RE OR	1.758																
21	95% LL	1.046																
23																		
24	P-value for Q	0.749																
25																		
26																		
27																		
29																		
30																		
31																		
32	() N Sheet1										14							
D	aw • 🔓 AutoSh	apes 🔹 🔪 👌		A 🗘	8 🔏 🗳	• - <u>- A</u>	• = =	로 🛯 🗐									I	
Rea	dv																	
d)	Start	10TE In	🕒 Inbox - Mic	2ro 🔁 P	MS GAS	Micros	oft E	Microsoft Ex	кс 📃 🖭 Ме	ta-analysi	MTHFR (and B	N 🖸 🏟 🜖	#08		<u>i</u> 1 2 1 2 1 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3	🔀 🕝 🔏 2514 🔛 🥘 10):09 PM

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.

🔤 Mi	icrosoft Excel - I	META_OR GA	s															_ 8 ×
:	<u>File E</u> dit <u>V</u> iev	w <u>I</u> nsert f	F <u>o</u> rmat <u>T</u> oo	ols <u>D</u> ata	<u>W</u> indow <u>S</u> ta	atsDirect <u>H</u> e	lp Ado <u>b</u> e i	PDF								Type	a question for help \bullet	_ 8 ×
1	🔁 🖄 🖾 👒	N B	5 🛛 🔊 🖷	n P R P P R R R R R R R R R R	eply with <u>C</u> han	ges E <u>n</u> d Re	view 📘	🗋 💕 🛃	> 🖪 🕯		at 🕄 🗱 🖓	(🗈 🖺 •	I 🔊 🛨	(* - 19)	$\Sigma - \frac{A}{Z} \downarrow \frac{Z}{A}$	1 🛍 🍕	98% 🗸 🕜 📘	
			i Tir	nes New Rom	an 🖌 11	- B Z	u∣≣		\$ % ?	÷.0 .00		- 🗞 - A	-				Zoom	
	F20 -	fx					[4 /0 /	.00 -0 -;			Ŧ					
	A	B	С	D	E	F	G	Н	1	J	К	L	М	N	0	Р	Y	
1	n=	0			-													
2			Ca	ises	Con	trols		RESULTS		OU	IPUT to be	e used for M	leta-regres	ssion and P	ublication	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	у=ө√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	0	study6	12	88	24	342	1.15	0.48	2.75	0.14	0.200	5.00	0.447	2.230	2.230	0.303		
/	8	study8	43	318	74 50	410	0.71	0.47	1.07	-0.34	0.043	23.05	0.208	4.801	4.801	-1.045		_
9	10	study9 study10	8	233	13	298	0.88	0.34	1.44	-0.15	0.002	4.76	0.458	2.183	2.183	-0.544		
10		studyro		200		200	0.110					67.20				0.0.11	0.001	
11						FE OR	0.94	0.74	1.19									
12						RE OR	0.94	0.74	1.19									
13																		
14																		
15																		
17 1	FE OR	0.935																_
18 9	95% LL	0.736																
19	95% UL	1.188																
20	PE OP	0.026																
21 1	95% LL	0.930																
23	95% UL	1.191																
24																		
25]	P-value for Q	0.409																
26																		
27																		
29																		
30																		
31																		
32	• • • Sheet1										14					<u> </u>	 	
1					a . A	- <i>a</i> - A		z 🗖 🖻			12							
Dead	w w A	apes • 🔨		ં ≪યારું		• • • •	• = •••	₩ ■ ■	Ŧ									
Ready	y • • • • • • • • • • • • • • • • • • •					inda	[[150 ····			6000	····		N (2) 10 1500 N (2)	

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

		Ca	ses	Cont	rols				
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	7.4 1
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).

2	Microsoft Excel - I	META_OR G/	s															
: 2	<u>Eile E</u> dit <u>V</u> iev	v <u>I</u> nsert I	F <u>o</u> rmat <u>T</u> oo	ols <u>D</u> ata	Window St	atsDirect <u>H</u> el	p Ado <u>b</u> e i	PDF								Туре	a question for help 🛛 🗸	.8×
-	- • 🗪 🖘 🖾 🖏		5 👔 🖦	in I ₩2 Re	oly with Chan	aes End Rei	view	n na 🗖			ABC 11	(🗈 🙈 -	I 🗐 🚽	(H + 1 Q)	$\Sigma = \frac{1}{2} \begin{bmatrix} z \\ z \end{bmatrix}$	1 🌆 🔜	98% 🗸 👩 📕	
-					.pry mar <u>a</u> nan						▼ B ≫0 0 ■ .=1==	0 - 1 - 1 - 1		· 69	Z Y A	•	Ţ	
			2 I II	nes New Rom	an • 11	• B 1	⊻∣≣	= = 📑	\$%,	.00 -00		• 🥙 • 🗛	- -					
	H25 👻	f _x																
	A	В	С	D	E	F	G	Н		J	K	L	M	N	0	P	Y	
1	n=	10																
-		10																
2			Ca	ises	Cor	trols		RESULTS		OUT	FPUT to be	e used for N	Meta-regre	ss <mark>ion and P</mark>	ublication	bias		
	aa	Study	mt/mt	Total	mt/mt	Total	OR	95%LL	95%UL	$\theta = \ln(OR)$	v	w=1/v	SE=√v	1/SE	x=√w	y=θ√w	t^2	
3																	(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		
0	3	study3	25	190	12	193	2.29	1.11	4.69	0.83	0.135	7.41	0.30/	2.722	2.722	2.250		
/	4	study4	94 7	514	24	495	1.48	0.48	2.09	0.39	0.031	52.54	0.1/5	2 236	2 236	0.303		
9	7	study0	343	1312	196	1160	1.13	1.43	2.73	0.14	0.010	00 14	0.100	0.057	0.057	5 521		
10	8	study/	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									
10	P					KE OK	1.33	1.05	1.69									
18	1																	
19																		
20)																	
21	FE OR	1.528																
22	95% LL	1.363																
23	93% UL	1./12																_
25	RE OR	1.331							1									
26	95% LL	1.046							ă.									
27	95% UL	1.692																
28	D 1 0 0	0.001																
29	P-value for Q	0.001																
31	/																	
32																		
l∎	♦ ► ► ► ► ► ► ► ► ► ► ► ► ► ► ► ► ► ► ►	, ,							-		•							
D	aw 🕶 🔓 A <u>u</u> toSha	apes 🔹 🔨 `		🔺 🖪 🛟	🚨 🔝 🗳	<u>• - 🚄 - A</u>	• = =	₹ ∎ 🗊	Ŧ									
Rea	ady																	
	Start COSM	10 🙆 In	box 📔	PMS GAS	🖳 Meta-ar	🖭 МТНЕ	R 📳	MAGASS	Book1	HWE PM	IS 📳 MI	ETA	B 🖸 🕅 🛈	₩ 08	* * *	<u>,</u>	🔀 🕝 🇞 🎦 🔛 🕘 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		Cont	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).

2	Microsoft Excel -	META_OR G	AS															- 8 ×
: 🛛	<u>Eile E</u> dit <u>V</u> iev	w <u>I</u> nsert	F <u>o</u> rmat <u>T</u> oe	ols <u>D</u> ata	<u>W</u> indow <u>S</u> ta	atsDirect <u>H</u> e	lp Ado <u>b</u> e I	PDF								Туре	a question for help 🛛 👻	- 8 ×
1	. 🖢 🖄 🖾 👒		5 📝 🖷	∭⊡ W ∛Re	ply with <u>C</u> han	ges E <u>n</u> d Re	view	🗋 📂 🛃	> 🖪 🔒		ar 🕄 🖓	6 🗈 🛍 •	I 🔊 🗸	(* - 1)	$\Sigma \rightarrow \begin{array}{c} A \\ Z \end{array} \downarrow \begin{array}{c} Z \\ A \end{array}$	1 🛄 🛷	98% 🗸 🕜 📮	
			j Tir	mes New Roma	an 🔹 11	- B I	U E	프 프 🔤	\$%,	€.0 .00 €.00 €.00.		- 🖄 - <u>A</u>	-			Chart	Wizard	
	123 -	fx										_						
	A	В	С	D	E	F	G	Н	1	J	К	L	M	N	0	Р	Y	
1	n=	8																
2			C	ises	Con	trols		RESULTS		OU	TPUT to be	e used for N	Meta-regres	sion and P	ublication	bias		
-	aa	Study	mt/mt	Total	mt/mt	Total	OR	95%LL	95%UL	$\theta = \ln(OR)$	v	w=1/v	SE=\v	1/SE	x=√w	v=θ√w	t^2	
3																· ·	(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	3.047	3.647	1.339		
	5	study3	20	190	12	193	2.29	1.11	4.09	0.83	0.135	7.41	0.307	2.722	2.722	2.250		<u> </u>
0	6	study5	7	88	24	342	1.59	0.04	2.99	0.40	0.104	5.00	0.323	2 236	2 236	0.303		
9	8	study0	43	318	74	410	0.71	0.40	1.07	-0.34	0.043	23.05	0.447	4 801	4 801	-1 645		
10	0	study0	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	1											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	FF OR	1.075																
20	95% LL	0.868																
21	95% UL	1.330																
22																		
23	RE OR	1.134]								
24	95% LL	0.845								T								
25	95% UL	1.522																
20	P-value for O	0 099																
28	1 Faller for Q	0.000																
29	1																	<u> </u>
30																		
31																		
<u>32</u>	♦ ► ► ► Sheet1 /										•							FIL
1 p						- <i>A</i> - A	- =	z 🛛 🖻										
100	aw K Autosh	apes · \	.00	ા જાય દુર		• 🛎 • 4		÷ = _	F									
Rea	ady		,															
a tr	Start COSN	40 🕒 In	ibox 📔	PMS GAS	Meta-ar	1 👜 МТН	FR 🖭	MAGASS	🛃 Book1	HWE PN	15 🕙 M	ETA	B 🖸 😭 🚺	#38	x x x	💃 🖓 🍳 😵	🔀 🕝 🔏 2609 🔛 🥘 10):40 PM

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 metaanalysis 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 nonsignificant 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 In(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		Cont	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:

File Edit	l - SPSS Dat View Data	a Editor Transform A	nalyze Graph	ıs Utilities V	Vindow Help	-	-	
	3 🖳 🗠		L [? #	<u>*</u> [[]	1	0		
4 :								
	theta	precisio	var	var	var	var	var	
1	.21	2.24						File Edit View Insert Format Analyze Graphs Utilities Window Help
2	.37	3.65						
	.03	5.71						
5	.46	3.10						Lenge Output
6	.14	2.24						E Cranh
7	.55	9.96						Notes Stupin
8	34	4.80						
9	13	4.01						
11	20	9 44						
12								
13		Simple Scatte	rplot				×	
14				Y Axi	s:	0	к	
15					1/SE [precisio]	Pa	ste	
16				X Axi	s:			
17)=In(OR) [theta]		sel	
19						Car	icel	
20			L C	→ Set M	larkers by:	H	alp	
21			L.					
22				Label	Cases by:			
23								
24		emplate				-		
26	E	Use chart spe	ecifications from	n:				.4 .2 0.0 .2 .4 .6 .8 1.0
27		File						
28						_		In(OR)
29				Titles	. Options			I I I I I I I I I I I I I I I I I I
30					-			
32								
	View V	riable View 1						
, , , , Dat								SPSS Processor is ready
👌 Start	Co,]	🖸 In 🍋	РМ [💷) м	1e 🎬 Un.	🛛 🔀 Mic	CP CF [1	i un 👩 M	Ma 🕅 un 🖂 Pa 🖾 Mic 🗱 On 📘 🖸 🕢 🖬 🖧 🗴 🍕 🕼 🖏 🖏 🖉 🖉 🖓 🖉 🖓 🕅 🖓 💷 🕅 🖓 🚛 👔
							· · · · · · · · · · · · · · · · · · ·	

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:

 $\mathbf{y}_i = \mathbf{a} + \mathbf{b} \mathbf{x}_i$ for i=1 to n, where n is the number of studies

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

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

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

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

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).

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

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 tdistribution with n-2 df, there is no indication of publication bias.

Example - MTHFR C677T and Breast Cancer

		Ca	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 \geq 0.05.

Therefore, there is no significant publication bias in the metaanalysis.

Practice

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

					AD	H2		
			*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 metaanalyses 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.

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.

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

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	ТС	CC	TT	ТС	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.

DRG 🚺	genotypes	- Notepad					🗠 C:\WINDOW5\system32\cmd.exe
File Ed	lit Format	View Help					For analysing an individual GAS, type: 1
25 118 94 215	21 92 87 215	11 23 12 65	28 110 80 178	23 162 85 242	15 43 25 94	<u>~</u>	For analysing GASs and perfrorming a meta-nalaysis of them, type: 2 2
50 173 387 140	80 145 577 196	17 24 196 74	58 43 274 135	96 38 695 140	32 7 343 43		You are performing an analysis of GAS and a meta-analysis of them Create an input file with the name "ORGgentotypes.txt" and
181 440	259 104 509	50 13 155	134 362	141 91 786	27 8 351		locate the file at the same directory as the current software.
I						V	For each study first enter the genotype frequencies of the control group (from wild types (wt) to more mutants (mt) genotypes) followed by the respective
Micro	coft Evcol	Poold					(e.g. for two alleles the order of entering
- Sellero	Solt Excel-	BOOKI	French	Taala Data	ter-de-		the frequencies is: 1st) wt/wt of control, 2nd) wt/mt
	Adoba DDE	w <u>i</u> nsert	Format	Tools Data	window		5th) wt/mt of cases, 6th) mt/mt of cases).
		> 👌	🛃 🄊	- Σ - <u>A</u> ↓	🏭 75%	• • •	the frequencies of the control group, followed by the less severe group, the more severe group, etc.
Arial		• 10 •	BI	토 클 클	· .	• <u> - A</u> - 関	The results are shown in the file "outputORG.txt"
М	8 👻	f _x					Enter the number of studies
1	A B	С	D	E F	G	H I	11
2		Breast Cancer	r	Helathy cor	itrols		Number of studies= 11 0000
3	cc	тс	TT	сс тс	TT		
5	25	21	11	28 23	15		Enter the number of groups (up to 20) (e.g. for a simple case-control study, enter=2,
6	118	92	23	110 162	43		for disease progression where the groups are:
7	94	87	12	80 85	25		controls, diseased, diseased with complication,
8	215	215	65	178 242	94		
9	50	80	17	58 90	32		2
10	1/3	145	24	43 38	242		
12	140	106	74	135 140	43		Number of groups – 2
13	242	259	50	166 141	27		Enter the number of genotypes (up to 200), e.g. for two alleles is 3
14	181	104	13	134 91	8		
15	440	509	155	362 786	351		3
16							Number of genotypes= 3
H 4 🕨	N ∖ Sheet1 ,	(Sheet2)	Sheet3/				Specify the number to add to zero frequencies
D <u>r</u> aw •	AutoSh	apes 🔹 🔨	$\mathbf{X} \square ($) 🔚 🖪 🕄	: 🚨 🚵	🖄 • 🚄 • 🍟	
Ready							▼

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

🖬 C:\WINDOWS\system32\cmd.exe			🚺 outputORG - Not	tepad			
GENETIC ASSOCIATION RESULTS		A	File Edit Format	View Help			
Study ORG 95% LL 1 1.100957 0.6086171 2 1.718735 1.267359 3 1.399888 0.9761942 4 1.409944 1.139179 5 1.234500 0.8445081 6 1.075926 0.7006444 7 1.722928 1.501016 8 0.7273660 0.5646600 9 0.8166907 0.6378981 10 1.112093 0.7977763 11 1.874497 1.635324	95% UL 1.991576 2.330869 2.007477 1.745065 1.804589 1.652218 1.977649 0.9369556 1.045596 1.550248 2.148649		Study 1 2 3 4 5 6 7 8 9 10 11	ORG 1.100957 1.718735 1.399888 1.409944 1.234500 1.075926 1.722928 0.7273660 0.8166907 1.112093 1.874497	95% LL 0.6086171 1.267359 0.9761942 1.139179 0.8445081 0.7006444 1.501016 0.5646600 0.6378981 0.7977763 1.635324	95% UL 1.991576 2.330869 2.007477 1.745065 1.804589 1.652218 1.977649 0.9369556 1.045596 1.550248 2.148649	Ă
META_ANALYSIS RESULTS			***META-ANAL	YSIS RESULTS*	* *		
Heterogeneity metrics Q= 75.88856 P-value for Q= 0. I^2(%)= 86.82278	.0000000E+00		Heterogeneity Q= 75	y metrics .88856 P-\ .82278	value for Q=	0.000000E+00	
Fixed Effects model ORG= 1.433720 95% Lower Limit= 1.336540 95% Upper Limit= 1.537965			Fixed Effect:	.02270 s model 1 4337	20		
Random Effects model ORG= 1.255769 95% Lower Limit= 1.017186			95% Lower Lin 95% Upper Lin Random Effect	mit= 1.33654 mit= 1.53790 ts model	40 55		
C:\Documents and Settings\user\Desktop\PMS	S GAS\ORGASSMA>		ORG= 95% Lower Lin 95% Upper Lin	1.2557 mit= 1.01718 mit= 1.5503	59 86 11		V
			T				

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 nondiseased) 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



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)



	2	3	4	5	6	
7	8	9	10	11	12	
13	14	15	16	17	18	
19	20	21	22			

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

					Ch	romo	son	ne N	luml	ber											
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	2
				E.							2										
					•			•													
10 bins	10	<mark>8</mark> bins	8	<mark>6</mark> bins	6	6	6	6	6	6	6	4 bins	4	4	4	4	4	4	4	2 2	
	2	0		40		1	6	0			80				10	0				1	

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=ΣRi/s

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

The method



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.

1st simulation



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,	Ethnicity	Pedigrees	Age (years)	Markers	Weighting	Location with
publication year						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

Execute the program "HEGESMA v2.0.exe" - 🗆 × C:\WINDOWS\system32\cmd.exe - HEGESMA v2.0.exe ٠ C:\Documents and Settings\user\Desktop\PMS GAS>HEGESMA_v2.0.exe *************HEGESMA v.2************* By Elias Zintzaras and John Ioannidis For software enquires mail Elias Zintzaras, e-mail: zintza@med.uth.gr The program caclulates the average rank, the Q, Ha, and B metrics, and their corresponding P-values. In addition, The program caclulates the metrics and the P-values for a specific bin, restricted to ranks +/-2the average rank of the bin, as described by Zintzaras and Ioannidis in Genet Epidemiol (2004) 8;28(2):123-137. In the curent director put a txt file with the data. This file has the format of a matrix with the columns being the studies and the rows being the ranks of each bin. Name the input file: xxx.dat The study weights are in a txt file, named: weights.dat The output file is in the current directory, named:monte_weight How many studies ? studies= 3.000000 How many bins ? 120 bins= 120 How many runs ? 3000 3000 runs= Main analyis ? if not enter 0 if unweighted enter 1 if weighted enter 2

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



The output file is named "monte_unweight"

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

四 1	monte_	_unw	eight	- Micro	osoft V	Vord									_	
ΞE	jie <u>E</u> d	lit <u>V</u>	liew	<u>I</u> nsert	F <u>o</u> rm	at]	<u>T</u> ools	T <u>a</u> ble	<u>M</u> a	thType	<u>W</u> ine	dow	<u>H</u> elp	- 🐼		×
Ec	ourier N	zw		10 •	в	I	υ	ΞΞ		🗸 aby	- A	- P	E 🖷	100%		P
Ė			_				- 1			_	_	Ŧ				
뜨	- Å.	1 * 1		2 • 1	3	· 4 ·	1 • 5	6		7 • • • 8	3 * 1 *	9 • •	• 10 •	1 • 11 •	12 -	<u>''</u> –
			-		n		<i>.</i>									
12		<u>, 190</u>	411		P-val	lue	IOL	mean								
		0	.789	9000												
ģ		ŏ	.709	9000												
14		0	.724	1000												
4		0	.45	7000												
12		0	.068	3000												
i.		0	.25	7000												
끈		0	.768	3000												
		0	.599	9000												
12		0	.743	3000												
12		0	.714	1000												
Ġ.		0	.09	1000												
11		0	414	5000												
i.		ŏ	.488	3000												
7		0	.088	3000												
		0	.47	5000												
1		0	.391	1000												
I - I		0	.571	7000												
1 R		0	.010	0000												
11		1	.000	0000												
i 🛓		0	.371	1000												
!?		0	.800	0000												-
		0	.830	2000												±
R.		0	361	2000												•
<u> </u> -		1	.000	0000												Ŧ
=	G 🗉 🔅	: 🕸 🖣	•													►
∄ <u>D</u> r	aw 🕶 🕻	6 A	<u>u</u> toSh	apes 🕶	1 1		0	A	: :	8	1 🖉	•4	2 - 1	<u>A</u> - =		₽₽
Pa	ge 9	Se	ec 1		9/20	At	14.90	n Ln 3	32 C	ol 13	RE	C TR	K EX	T OVR	Englis	sh (U

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:

- transformation of the data (e.g. logtransformation 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).

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 <u>class</u> 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<0.000037*	P<0.001	P<0.000037*	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	Hypoth	nesis-free	Hypothesis-driven			
Approaches	WGS	VGS GWAS				
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)	17 genes (based on meta-analyses)				
Replicated findings across approaches	<u>9p21.3</u> (Chr9 at 18-48cM) with 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, He 2007, WTCCC 2007)					

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


(Kitsios and Zintzaras 2009 Ann Hum Genet)