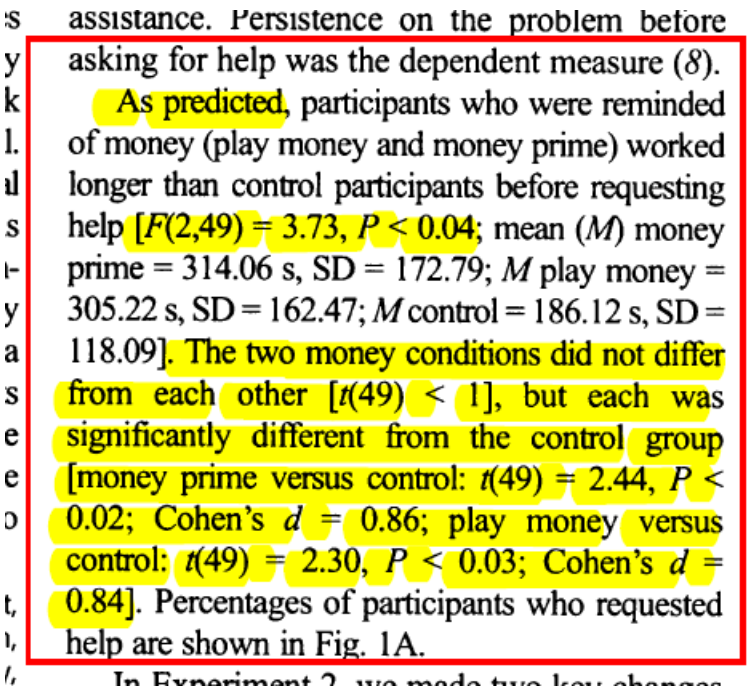
How to compute the R-Index?

I created an excel spreadsheet to assist with the computation of the R-Index.

To illustrate how to use the spreadsheet I use a Science article by Vohs, Mead, and Goode (2006). The article reports 9 independent studies. The spreadsheet can also be used for 9 articles with single studies, and multiple articles with multiple studies.

The first step is to find all statistical analysis that are reported for a single study and enter them in the spreadsheet.


assistance. Persistence on the problem before asking for help was the dependent measure (*8*).

As predicted, participants who were reminded of money (play money and money prime) worked longer than control participants before requesting help [$F(2,49) = 3.73$, $P < 0.04$; mean (M) money prime = 314.06 s, SD = 172.79; M play money = 305.22 s, SD = 162.47; M control = 186.12 s, SD = 118.09]. The two money conditions did not differ from each other [$t(49) < 1$], but each was significantly different from the control group [money prime versus control: $t(49) = 2.44$, $P < 0.02$; Cohen's $d = 0.86$; play money versus control: $t(49) = 2.30$, $P < 0.03$; Cohen's $d = 0.84$]. Percentages of participants who requested help are shown in Fig. 1A.

For Study 1, the article reports one F-test and three t-tests. These results are entered in the spreadsheet in the appropriate columns. The F-value of 3.73 is highlighted in the picture below.

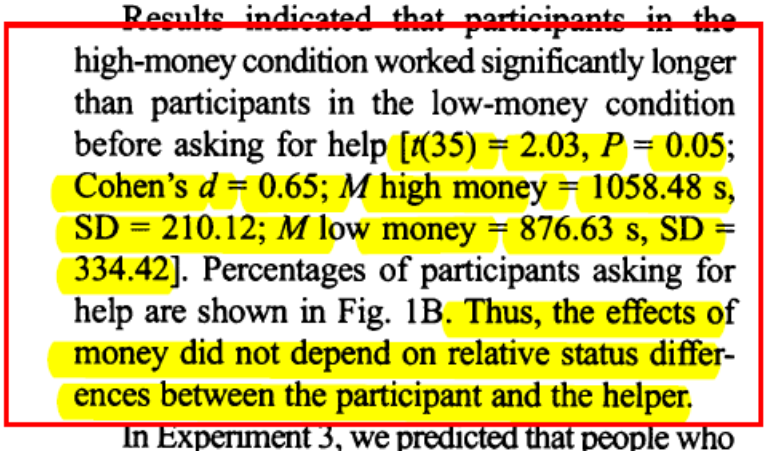
It is also important to specify the level of statistical significance for a two-tailed test. Typically, the value is .05. However, if the test is one-tailed, the level changes to .10 (.10 two-tailed equals .05 one-tailed). The spreadsheet uses this value to determine whether an observed p-value is considered significant (rejecting the null-hypothesis) or not (not rejecting the null-hypothesis). After entering the degrees of freedom, the spreadsheet computes the exact p-value for this test.

t-values are entered in the t-value column. The spreadsheet converts t-values into F-values. After entering degrees of freedom, it computes the exact p-value. The article also reports that two conditions do not differ from each other, $t(49) < 1$. This comparison is theoretically not important. It can be included or excluded. For the computation of the R-Index it is important to enter information about the

way a significance test is used. There are multiple options (e.g., MC for manipulation check). Most important is whether a test is considered to test a theoretically important hypothesis or not. Tests that are considered to be theoretically important will be included in the R-Index and are labeled FHT (focal hypothesis test). The spreadsheet shows that all three tests produced significant results.

		Average Median STD DEV									
Title	Article #	Study #	Design	Type	Crit 2t	N	t	F	df1	df2	F p-value
Vohs Science	1	1	BS	FHT	0.05	51		3.73	2	49	0.031056503
Vohs Science	1	1.01	BS	FHT	0.05	51	2.44	5.9536	1	49	0.018351452
Vohs Science	1	1.02	BS	FHT	0.05	51	2.3	5.29	1	49	0.025747297

In this example, the F-test and the two t-tests that compare the experimental condition to two control conditions are all considered theoretically important and labeled FHT.


Results indicated that participants in the high-money condition worked significantly longer than participants in the low-money condition before asking for help [$t(35) = 2.03$, $P = 0.05$; Cohen's $d = 0.65$; M high money = 1058.48 s, SD = 210.12; M low money = 876.63 s, SD = 334.42]. Percentages of participants asking for help are shown in Fig. 1B. Thus, the effects of money did not depend on relative status differences between the participant and the helper.

In Experiment 3, we predicted that people who

The second study reports a t-value of 2.03 with 35 degrees of freedom. It also reports that p equals .05. Entering this information into the spreadsheet reveals that the exact p-value is .05001, which is above .05. The spreadsheet would treat this finding as a non-significant result. However, the result is considered as significant in the article. This interpretation implies that researchers are willing to accept a slightly higher type-I error rate. For this reason, I set the significance level at .06. This way the result is treated as significant and power increases slightly because power changes with the significance criterion. The adjusted significance criterion is highlighted in red.

		Average Median STD DEV										
Title	Article #	Study #	Design	Type	Crit 2t	N	t	F	df1	df2	F p-value	r/beta
Vohs Science	1	1	BS	FHT	0.05	51		3.73	2	49	0.031056503	
Vohs Science	1	1.01	BS	FHT	0.05	51	2.44	5.9536	1	49	0.018351452	
Vohs Science	1	1.02	BS	FHT	0.05	51	2.3	5.29	1	49	0.025747297	
Vohs Science	1	2.00	BS	FHT	0.06	37	2.03	4.1209	1	35	0.050011529	

The process continues for all studies. Statistical tests from the same study are separated and I highlight this separation manually with a yellow line (sorry, not automated yet).

The next picture shows the data entry for Study 9. Study 9 reported three chi-square tests. These are entered in the columns for the chi-square test.

		Average Median STD DEV									
Title	Article #	Study #	t p-value	z	p-value	chi-square	chi-df	chi p	CI low	CI high	CI avg	CI t
Vohs Science	1	8.02										
Vohs Science	1	9.00				10.1	2	0.006409				
Vohs Science	1	9.00				7	1	0.008151				
Vohs Science	1	9.00				8.22	1	0.004143				

When all data are entered, you can scroll over to the right.

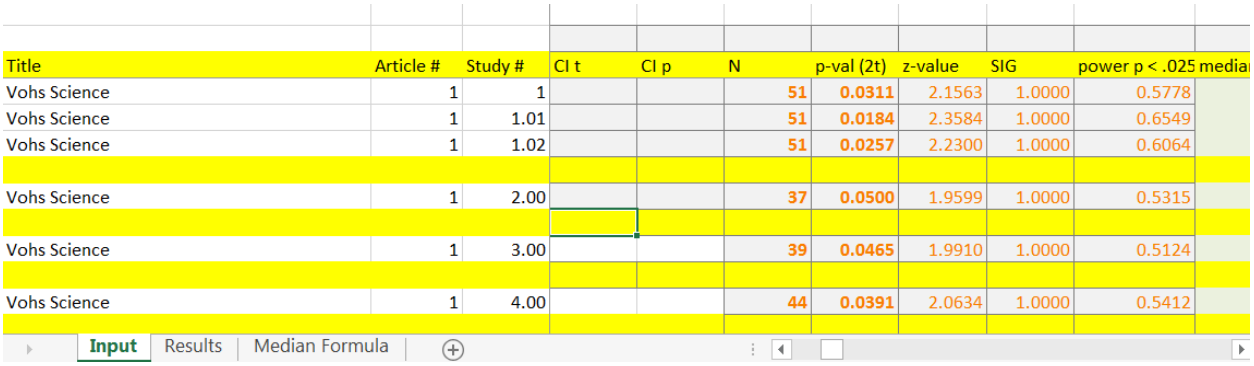
		Average Median STD DEV			0.0258 0.0248 0.0147	2.3015 2.2453 0.2791	1.0000 1.0000 0.0000	0.6302 0.6123 0.0978	0.0556	0.0000	0.9444
Title	Article #	Study #	CI p	N	p-val (2t)	z-value	SIG	power p < .025	P > .8	P < .5	.5 < p < .8
Vohs Science	1	1		51	0.0311	2.1563	1.0000	0.5778	0	0	1
Vohs Science	1	1.01		51	0.0184	2.3584	1.0000	0.6549	0	0	1
Vohs Science	1	1.02		51	0.0257	2.2300	1.0000	0.6064	0	0	1
Vohs Science	1	2.00		37	0.0500	1.9599	1.0000	0.5315	0	0	1

The first column in this section shows the sample size (N) based on the sample sizes entered earlier. The second column shows the p-values from the various statistical tests that were entered earlier (F, t, chi-square, etc.). The next column converts the two-tailed p-values into z-scores for one-tailed tests (p / 2). These z-scores can be interpreted as non-centrality parameters. The z-score is then used to compute observed power. The Significance column (SIG) records whether a p-value is below the criterion value (1) or not (0).

Summary statistics are presented above the rows for data entry. For significant results, the average shows the success rate of significance tests. The success rate is 100%. The median of observed power values provides an unbiased estimate of median power (the average can be biased, but often shows very similar results). Median power is 61%.

It can be informative to look at these columns that are based on all statistical tests. However, this information is not used for the computation of the R-Index because some tests from the same study are not independent. For this reason, I compute average (mean or median) statistics for statistical tests included in the same study. Unfortunately, this step is not automated yet because the number of rows varies across studies.

To complete this step, you need to switch to the "median formula" sheet by clicking on it at the bottom of the screen.



Title	Article #	Study #	CI t	CI p	N	p-val (2t)	z-value	SIG	power p < .025 media
Vohs Science	1	1			51	0.0311	2.1563	1.0000	0.5778
Vohs Science	1	1.01			51	0.0184	2.3584	1.0000	0.6549
Vohs Science	1	1.02			51	0.0257	2.2300	1.0000	0.6064
Vohs Science	1	2.00			37	0.0500	1.9599	1.0000	0.5315
Vohs Science	1	3.00			39	0.0465	1.9910	1.0000	0.5124
Vohs Science	1	4.00			44	0.0391	2.0634	1.0000	0.5412

Input | Results | Median Formula

On this sheet highlight the row with the correct number of statistical tests. For example, for Study 1 there are three statistical tests. So you need to highlight the row 1.02 (counting from 1.00, 1.01, 1.02) and copy it.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1													
2		N	p-val (2t)	z-value	SIG	power p <	median N	median p	median z	avg sig	med pow	med d*	
3	1.00	10	0.05	2	1	0.5	10	0.05	2	1	0.5	1.458445	1.00
4	1.01	10	0.05	2	1	0.5	10	0.05	2	1	0.5	1.458445	1.01
5	1.02	10	0.05	2	1	.5	10	0.05	2	1	0.5	1.458445	1.02
6	1.03	10	0.05	2	1	0.5	10	0.05	2	1	0.5	1.458445	1.03
7	1.04	10	0.05	2	1	0.5	10	0.05	2	1	0.5	1.458445	1.04
8	1.05	10	0.05	2	1	0.5	10	0.05	2	1	0.5	1.458445	1.05
9	1.06	10	0.05	2	1	0.5	10	0.05	2	1	0.5	1.458445	1.06
10	1.07	10	0.05	2	1	0.5	10	0.05	2	1	0.5	1.458445	1.07
11	1.08	10	0.05	2	1	0.5	10	0.05	2	1	0.5	1.458445	1.08
12	1.09	10	0.05	2	1	0.5	10	0.05	2	1	0.5	1.458445	1.09
13	1.10	10	0.05	2	1	0.5	10	0.05	2	1	0.5	1.458445	1.10
14	1.11	10	0.05	2	1	0.5	10	0.05	2	1	0.5	1.458445	1.11
15	1.12	10	0.05	2	1	0.5	10	0.05	2	1	0.5	1.458445	1.12
16	1.13	10	0.05	2	1	0.5	10	0.05	2	1	0.5	1.458445	1.13

Then switch back to the input sheet and past it in the last row of a study.

		Average	0.6302	42.8333	0.0237	2.3031	1.0000	0.6311	0.7489	
		Median	0.6123	38.5000	0.0243	2.2536	1.0000	0.6155	0.7313	
		STD DEV	0.0978	10.4195	0.0107	0.2241	0.0000	0.0793	0.1323	
Title	Article #	Study #	power p < .025	median N	median p	median z	avg sig	med pow	med d*	
Vohs Science	1	1	0.5778							
Vohs Science	1	1.01	0.6549							
Vohs Science	1	1.02	0.6064	51	0.025747	2.230002	1	0.606435	0.644129	
Vohs Science	1	2.00	0.5315							
Vohs Science	1	3.00	0.5124							
Vohs Science	1	4.00	0.5412							

You need to repeat this procedure for every study. The rows above the data entry rows (Average, Median, STD DEV) show summary statistics. This information can be interested, but it is not used for the R-Index because it includes statistical tests that were expected to produce non-significant results and tests that are not theoretically important (e.g., manipulation checks).

The next columns show pseudo-effect sizes. These are real effect sizes for simple between-subject designs (r converted to d). However, for complex designs these effect sizes DO NOT correspond to actual d-values. For example, in a within-subject design the effect size may be small (d = .20) but the study may have high power due to a strong intraclass correlation. In this case, pseudo-d can be much higher.

		Correlation between Sample Size and Effect Size			
		Median p r(N,d)	All r(N,d)	FHT r(N,d)	FHT BS r(,d)
		-0.6272	-0.6292	-0.63	-0.63
			0.7536	0.7536	0.7536	
5969	0.7181		0.7536	0.7536	0.7536	42.1176
5062	0.6675		0.7100	0.7100	0.7100	37.0000
0814	0.1145		0.1357	0.135671	0.135671	9.0061
ow	med d*		all d*	FHT d*	BS FHT d*	N
			0.9432	0.943242	0.943242	34.0000
2785	0.829242		0.7306	0.730584	0.730584	34.0000
3566	0.673565		0.6736	0.673565	0.673565	40.0000
			0.8012	0.801206	0.801206	35.0000
0621	0.788947		0.7775	0.777542	0.777542	35.0000

Above the pseudo-d columns are correlations between sample size and pseudo-d for median ds, all individual rows, just rows that tested focal hypotheses and rows that tested focal hypotheses in a between-subject design. Negative correlations show that smaller samples have larger effects. This may reveal questionable research practices, but it can also show that researchers allocate more resources to test hypotheses about weaker effects. Thus, negative correlations have to be interpreted WITH CAUTION.

In the following columns the information about sample size, z-scores, p-values, significance, observed power, and pseudo-effect sizes is presented again, but ONLY FOR ROWS that have a focused hypothesis test (FTH) designation. As a result, the success rate and median power can be quite different from the values when all statistical tests are included. In this case, the values are the same because the spreadsheet contains only FTH rows.

			Size								Median FHT r(N,d) -0.7422				
												Success Rate	Obs. Powe	Inflation Rate	R-Index
												1.0000	0.5969	0.4031	0.1937
		Average	43.1111	0.0258	2.3015	1.0000	0.6302	42	0.0308	2.2037	1.0000	0.5969	0.7181		
		Median	38.0000	0.0248	2.2453	1.0000	0.6123	39	0.0258	2.2295	1.0000	0.6062	0.6675		
		STD DEV	9.7008	0.0147	0.2791	0.0000	0.0978	9	0.0140	0.2331	0.0000	0.0814	0.1145		
Title	Article #	Study #	N	FHT p	z-value	SIG	OBS Power	median N	median p	median z	avg sig	med pow	med d*		
Vohs Science	1	1.02	51.0000	0.0257	2.2300	1.0000	0.6064	51	0.025747	2.230002	1	0.606434571	0.644129		
Vohs Science	1	2.00	37.0000	0.0500	1.9599	1.0000	0.5315	37	0.050012	1.959865	1	0.531512225	0.66746		
Vohs Science	1	3.00	39.0000	0.0465	1.9910	1.0000	0.5124	39	0.046484	1.99097	1	0.512367569	0.659728		
Vohs Science	1	4.00	44.0000	0.0391	2.0634	1.0000	0.5412	44	0.039074	2.063407	1	0.541194277	0.642219		
Vohs Science	1	5.00	34.0000	0.0215	2.2992	1.0000	0.6328								
Vohs Science	1	5.01	34.0000	0.0097	2.5856	1.0000	0.7342								
Vohs Science	1	5.02	34.0000	0.0410	2.0440	1.0000	0.5335	34	0.021493	2.299204	1	0.632785447	0.829242		
Vohs Science	1	6.00	40.0000	0.0397	2.0568	1.0000	0.5386	40	0.039707	2.056787	1	0.53856646	0.673565		

To get the median values, you can just copy the median cells from the previous section into the identical section here.

The red cells show the final results. The success rate is 100%. Median observed power is 60%. The inflation rate (success rate – observed power) is 40% and the R-Index is 20%.

The R-Index can be compared to the following scenarios.

An R-Index of 22% is consistent with a set of studies in which the null-hypothesis is true and a researcher reported only significant results.

An R-Index of 40% is consistent with a set of studies with 30% power and the non-significant results are not published. In this case, observed power is 70%, the success rate is 100%, and inflation rate is 30%.

In conclusion, the R-Index in this example suggests that even replication studies with much larger samples would fail and that the true effect size is close to zero.

If you use this spreadsheet in a publication, please cite it as follows.

R-Index.org (2014). R-Index 1.0. www.r-Index.org.

If you would like to share the results of an R-Index analysis, please submit the spreadsheet with completed data and a brief description of the research question (e.g., computed R-Index for a specific journal, a particular research topic, or a researcher, including yourself) to www.r-index.org. The results will be posted with the research community on the R-Index.org website. All results will be published anonymously. The spreadsheets will be openly accessible and errors will be corrected following the model of Wikipedia.