APPENDIX

Methods – Additional details

Classification of outcomes
Outcome measures of individual trials were classified as observer-reported, patient-reported (via interviewer or directly recorded by patients), healthcare provider decision outcomes or as mixed (in cases where the outcome was a mixture of more than one category, e.g. both patient and observer-reported elements). Meta-analyses including trials classified as "mixed" were typically included at the screening stage as potentially eligible meta-analyses with patient reported outcomes, but it was determined at a later stage that the outcomes combined patient-reported and observer-reported elements.

Clinical events assessed by an adjudication committee were classified as observer reported (the "observers" being the adjudication committee members). Outcomes such as readmissions or need for conversion to open surgical procedure were classified as healthcare provider decision outcomes irrespective of how information on the outcome was procured (e.g. via hospital records or reported by patients).

Identification and inclusion of meta-analyses with observer-reported outcomes
Based on the screening using risk of bias scores we identified a potentially informative meta-analysis with an observer-reported outcome (i.e. potentially eligible for analysis (III)) from 226 reviews. Judging the workload involved in extracting data from all 226 analyses to be excessive, we only proceeded with data extraction from a random sub-sample of 120 meta-analyses. The size of this sub-sample was based on repeated random sampling until a number of meta-analyses had been reached (79 meta-analyses) which was well in excess of our pragmatic aim of a minimum of 30 meta-analyses still informative based on data extraction on blinding status from the individual trial publications.

Assessment of blinding status of patients, healthcare providers and outcome assessors
Our blinding algorithm entailed primarily basing blinding status on any explicit descriptions in publications, and only allowed passing of judgement on blinding status based on other information in certain specified situations. The algorithm comprised the following rules applied in the order stated:

If explicit description stated that some group (patients, healthcare providers, interviewers (if any) or outcome assessors) was blind/non-blind (e.g. "Patients were kept unaware of treatment status (…)" or "Theatre nurses were not blinded to treatment allocation") take as blind/non-blind for that group.

If explicit description stated that some group was blind and the trial used active control (e.g. "usual care") or no treatment as comparator, take as non-blind for groups for which blinding status was not explicitly stated.

If no indication of blinding (i.e. no mentioning of placebo/double dummy and not described as "double blind" or "single blind" etc.): take as non-blind for groups for which blinding status was not explicitly stated.

If any indication of blinding (i.e. mentioning of placebo/double dummy or described as "double blind" or "single blind" etc.): contact authors, UNLESS trial is a drug trial using placebo/double dummy AND described as "double blind"/"triple blind", in which case take as blind for groups for which blinding status was not explicitly stated.

Additional details of statistical analysis
Data management and graphics used Stata, version 14 (Stata Corp., Cary, North Carolina). Bayesian Markov chain Monte Carlo methods were used to fit the bias models in WinBUGS (MRC Biostatistics Unit, Cambridge, United Kingdom) (1). These models were based on the bias hierarchical model by Welton and colleagues (2), specifically model 3 which allows the treatment effect to vary, the average amount of bias

across meta-analyses to vary and additionally the study specific bias across trials to vary in a one stage approach. Vague priors were assumed with a modified Inverse Gamma (0.001, 0.001) prior on all variance components to allow increased weight on small values. This was chosen from the earlier BRANDO analysis by Savovic and colleagues (3) who found this prior to perform the best (with the lowest average mean squared error) having conducted a simulation study. It is well known with this type of modelling that variance components can be sensitive to the prior distributions (4). For each analysis, 2 parallel chains were run, with a burn-in of 250 000 iterations followed by at least a further 1 000 000 iterations, with a thinning of 5. Convergence was assessed by using history plots and checking that results from the 2 chains agreed. For location parameters (overall mean bias, baseline response rates, treatment effects), Normal (0, 1000) priors were assumed. The Welton et al hierarchical bias model assumes biases are broadly similar (exchangeability assumption) within a meta-analysis, and assumes the average bias is broadly similar (exchangeability assumption) across meta-analyses (2). The WinBUGS model code is given below.

Main analyses:

```
model {
        for (i in 1:Nb) {
        rc[i] ~ dbin(pc[i],nc[i])                       # likelihood for binary outcomes
        rt[i] ~ dbin(pt[i],nt[i])
        logit(pc[i]) <- mu[i]                                   # model for binary
outcomes (logit link)
        logit(pt[i]) <- mu[i] + delta[i]  + beta[i]*C[i]
         mu[i] ~ dnorm(0,.01)
        }

        for (i in Nb+1:Nc+Nb) {
        var[i] <- pow(se[i],2)   # calculate variances
        prec.smd[i] <- 1/var[i]     # set precisions
        lnor[i] ~ dnorm(nu[i],prec.smd[i])              # likelihood for continuous
outcomes on log odds ratio scale
        nu[i] <- delta[i]  + beta[i]*C[i]                       # model for
continuous outcomes (identity link)
        }

        for (i in 1:Nc+Nb) {
         beta[i]~dnorm(b[ma[i]],p.k2[ma[i]])I(-10,10)           # between study,
within MA, variation in bias
           delta[i]~dnorm(d[ma[i]],p.d[ma[i]])I(-10,10)         #RE for treatment
effect within meta-analysis

        }

        for (m in 1:N_ma) {
                d[m] ~ dnorm(0,.01)          # priors for true fixed (unrelated) treatment
effects
                b[m] ~ dnorm(b0,p.phi)
        #between meta-analysis variation in mean bias
                p.d1[m]~dgamma(.001,.001)
                p.d[m]<-p.d1[m]/(1-patom.d[m])
                patom.d[m]~dbeta(1,1)
        }
```

```
		b0 ~ dnorm(0,.001)				# vague prior for overall mean bias

		p.k1~dgamma(.001,.001)
		kappa <- pow(p.k,-0.5)
		p.k<-p.k1/(1-patom.k)
		patom.k~dbeta(1,1)
		for (m in 1:N_kappa_ok){
				p.k2[kappa_ok[m]]<-p.k
		}
		for (m in 1:N_kappa_cut){
				p.k2[kappa_cut[m]]<- cut(p.k)
		}

		p.phi1~dgamma(.001,.001)
		phi <- pow(p.phi,-0.5)
		p.phi<-p.phi1/(1-patom.phi)
		patom.phi~dbeta(1,1)
		b.new~dnorm(b0,p.phi)				#predictive distn for mean bias in new meta-
analysis
		beta.new~dnorm(b.new,p.k)	#predictive distn for bias in new study in new meta-analysis
		lkappa<-log(kappa)
		lphi<-log(phi)
		dum<-s[1]
}
```

Supplementary analysis, continuous outcomes:

```
model {
		for (i in 1:N) {
		var[i] <- pow(smd.se[i],2)   # calculate variances
		prec.smd[i] <- 1/var[i]    # set precisions
		smd[i] ~ dnorm(nu[i],prec.smd[i])		# likelihood
		nu[i] <- delta[i]  + beta[i]*C[i]		# model
		 beta[i]~dnorm(b[ma[i]],p.k2[ma[i]])I(-10,10)

		  delta[i]~dnorm(d[ma[i]],p.d[ma[i]])I(-10,10)

		}
				for (m in 1:N_ma) {
				d[m] ~ dnorm(0,.01)
				# priors for true fixed (unrelated) treatment effects
	b[m] ~ dnorm(b0,p.phi)
		#between meta-analysis variation in mean bias

				p.d1[m]~dgamma(.001,.001)
				p.d[m]<-p.d1[m]/(1-patom.d[m])
				patom.d[m]~dbeta(1,1)
		}

		b0 ~ dnorm(0,.001)				# vague prior for overall mean bias
```

```
                p.k1~dgamma(.001,.001)
                kappa <- pow(p.k,-0.5)
                p.k<-p.k1/(1-patom.k)
                patom.k~dbeta(1,1)
                for (m in 1:N_kappa_ok){
                                p.k2[kappa_ok[m]]<-p.k
                }
                for (m in 1:N_kappa_cut){
                                p.k2[kappa_cut[m]]<- cut(p.k)
                }

                p.phi1~dgamma(.001,.001)
                phi <- pow(p.phi,-0.5)
                p.phi<-p.phi1/(1-patom.phi)
                patom.phi~dbeta(1,1)

                b.new~dnorm(b0,p.phi)
                                                #predictive distn for mean bias in new meta-analysis
                beta.new~dnorm(b.new,p.k)
                 #predictive distn for bias in new study in new meta-analysis

                lkappa<-log(kappa)
                lphi<-log(phi)
                dum<-s[1]
}
```

Supplementary analysis using label invariant model (5):

```
 model {

                for (i in 1:Nb) {
                rc[i] ~ dbin(pc[i],nc[i])
                rt[i] ~ dbin(pt[i],nt[i])
                                logit(pc[i]) <- mu[i]
                logit(pt[i]) <- mu[i] + delta[i]  + beta[i]*C[i]
                 mu[i] ~ dnorm(0,.01)
                }

                for (i in Nb+1:Nc+Nb) {
                var[i] <- pow(se[i],2)   # calculate variances
                prec.smd[i] <- 1/var[i]     # set precisions
                lnor[i] ~ dnorm(nu[i],prec.smd[i])              # likelihood
                nu[i] <- delta[i]  + beta[i]*C[i]                  # model
                }

                for (i in 1:Nc+Nb) {
                 beta[i]~dnorm(b[ma[i]],p.k2[ma[i]])I(-10,10)

                   delta[i]~dnorm(d[ma[i]],p.d[ma[i]])I(-10,10)
                }
```

```
                for (m in 1:N_ma) {
                        d[m] ~ dnorm(0,.01)
                        # priors for true fixed (unrelated) treatment effects
        b[m] ~ dnorm(b0,p.phi)
                #between meta-analysis variation in mean bias

                var_d[m]~dlnorm(mu2,p.tau)              # log-normal distribution for between-study
variances
p.d[m] <- 1/var_d[m]
p.k2[m] <- equals(kappa_ok[m],1)/(var_d[m]*lambda)
+equals(kappa_ok[m],0)/(var_d[m]*cut(lambda))

                }
                lambda ~dlnorm(0,1)          # vague prior for change in between-study variation associated
with characteristic
                b0 ~ dnorm(0,.001)                      # vague prior for overall mean bias

                p.tau<-1/(sd.tau*sd.tau)
sd.tau~dunif(0,2)
mu2~dnorm(0,0.001)
log.tau2.new~dlnorm(mu2,p.tau)   # predictive distn for heterogeneity among studies without the
characteristic
tau2.new<-exp(log.tau2.new)

                p.phi1~dgamma(.001,.001)
                phi <- pow(p.phi,-0.5)
                p.phi<-p.phi1/(1-patom.phi)
                patom.phi~dbeta(1,1)

                lphi<-log(phi)
                dum<-s[1]
```

Appendix Table 1 Characteristics of included trials

		All (n=1153)		(Ia) n=132		(Ib) n=95		(IIa) n=173		(IIb) n=91		(III) n=397	
		n	%	n	%	n	%	n	%	n	%	n	%
Blinding status													
Patients blind	Definitely no	66	5.7	16	12.1	5	5.3	6	3.5	3	3.3	24	6.0
	Definitely yes	170	14.7	15	11.4	24	25.3	40	23.1	29	31.9	38	9.6
	Probably no	589	51.1	73	55.3	32	33.7	73	42.2	18	19.8	250	63.0
	Probably yes	274	23.8	18	13.6	33	34.7	46	26.6	41	45.1	69	17.4
	Unclear	54	4.7	10	7.6	1	1.1	8	4.6	0	0	16	4.0
Healthcare providers blind	Definitely no	94	8.2	20	15.2	8	8.4	8	4.6	8	8.8	32	8.1
	Definitely yes	100	8.7	6	4.5	11	11.6	40	23.1	14	15.4	22	5.5
	Probably no	591	51.3	78	59.1	37	38.9	64	37.0	21	23.1	248	62.5
	Probably yes	312	27.1	21	15.9	37	38.9	53	30.6	47	51.6	79	19.9
	Unclear	56	4.9	7	5.3	2	2.1	8	4.6	1	1.1	16	4.0
Outcome assessors blind	Definitely no	21	1.8	0	0	0	0	0	0	0	0	18	4.5
	Definitely yes	202	17.5	0	0	76	80.0	0	0	54	59.3	128	32.2
	Probably no	290	25.2	0	0	0	0	0	0	0	0	160	40.3
	Probably yes	181	15.7	0	0	19	20.0	0	0	19	20.9	71	17.9
	Unclear	38	3.3	0	0	0	0	0	0	0	0	20	5.0
	N/A	421	36.5	132	100.0	0	0	173	100.0	18	19.8	0	0
Double-blind explicitly mentioned	Yes	402	34.9	28	21.2	49	51.6	81	46.8	64	70.3	100	25.2
	No	750	65.0	103	78.0	46	48.4	92	53.2	27	29.7	297	74.8
	Unclear	1	0.1	1	0.8	0	0	0	0	0	0	0	0
All groups described as blind/ double-blind/ triple-blind	Yes	412	35.7	29	22.0	49	51.6	87	50.3	65	71.4	102	25.7
	No	740	64.2	102	77.3	46	48.4	86	49.7	26	28.6	295	74.3
	Unclear	1	0.1	1	0.8	0	0	0	0	0	0	0	0
Risk of bias													
Concealment of allocation	High risk	127	11.1	15	11.5	19	20.0	9	5.2	11	12.1	67	16.9
	Low risk	510	44.4	69	52.7	46	48.4	110	64.0	57	62.6	151	38.1
	Unclear	512	44.6	47	35.9	30	31.6	53	30.8	23	25.3	178	44.9
Incomplete outcome data	High risk	177	15.8	6	4.5	9	11.1	12	6.9	8	9.1	69	18.3
	Low risk	771	68.7	103	78.0	59	72.8	143	82.7	69	78.4	228	60.3
	Unclear	175	15.6	23	17.4	13	16.0	18	10.4	11	12.5	81	21.4
Drug trial*		753	65.3	48	36.4	77	81.1	127	73.4	81	89.0	205	51.6

Funding	Profit organisations	251	21.8	16	12.1	29	30.5	40	23.1	36	39.6	61	15.4
	Non- profit organisations	364	31.6	63	47.7	36	37.9	48	27.7	23	25.3	144	36.3
	Both	108	9.4	10	7.6	5	5.3	13	7.5	10	11.0	36	9.1
	Unclear	430	37.3	43	32.6	25	26.3	72	41.6	22	24.2	156	39.3
Trial design	Cluster randomisation	20	1.7	2	1.5	0	0	0	0	0	0	12	3.0
	Cross-over	7	0.6	0	0	0	0	0	0	0	0	4	1.0
	Cross-over trial used as parallel group trial in meta-analysis	9	0.8	3	2.3	3	3.2	0	0	0	0	7	1.8
	Parallel	1112	96.4	125	94.7	92	96.8	173	100.0	91	100.0	374	94.2
	Split body	0	0	0	0	0	0	0	0	0	0	0	0
	Unclear	5	0.4	2	1.5	0	0	0	0	0	0	0	0

*Trials in which interventions in trial arms differ only by the administration of one or more substances, including parenteral fluid and nutrition, vaccines and some interventions of biological origin, e.g. blood components. This classification was used when scoring blinding status, but the category "drug trial" is only partly overlapping with the categorisation of experimental interventions as "Pharmacologic" in the main Table 1.

Appendix Table 2 Associations between reported study characteristics

Study characteristic 1	**Study characteristic 2**	**All trials (n, %)**	**(Ia) (n, %)**	**(Ib) (n, %)**	**(IIa) (n, %)**	**(IIb) (n, %)**	**(III) (n, %)**
Patients	**Healthcare provider**						
Blinded	Blinded	399 (34.6)	26 (19.7)	48 (50.5)	84 (48.6)	61 (67.0)	99 (24.9)
Blinded	Non-blinded	45 (3.9)	7 (5.3)	9 (9.5)	2 (1.2)	9 (9.9)	8 (2.0)
Non-blinded	Blinded	13 (1.1)	1 (0.8)	0 (0)	9 (5.2)	0 (0)	2 (0.5)
Non-blinded	Non-blinded	696 (60.4)	98 (74.2)	38 (40.0)	78 (45.1)	21 (23.1)	288 (72.5)
OR (95%)		474.7 (253.0 to 890.7)	364.0 (42.8 to 3092.2)	388.0 (21.9 to 6880.2)	364.0 (76.3 to 1737.2)	271.9 (15.2 to 4876.1)	1782.0 (372.1 to 8533.4)
Patients	**Outcome assessor**						
Blinded	Blinded	264 (35.9)	0 (0)	57 (60.0)	0 (0)	52 (71.2)	103 (25.9)
Blinded	Non-blinded	10 (1.4)	0 (0)	0 (0)	0 (0)	0 (0)	4 (1.0)
Non-blinded	Blinded	120 (16.3)	0 (0)	38 (40.0)	0 (0)	21 (28.8)	96 (24.2)

Non-blinded	Non-blinded	341 (46.4)	0 (0)	0 (0)	0 (0)	0 (0)	194 (48.9)
OR (95%)		75.0 (38.6 to 145.8)		1.5 (0.0 to 76.9)		2.4 (0.0 to 127.1)	52.0 (18.6 to 145.5)
Patients	**Allocation concealment**						
Blinded	Yes	261 (22.7)	23 (17.4)	38 (40.0)	68 (39.3)	52 (57.1)	56 (14.1)
Blinded	No	183 (15.9)	10 (7.6)	19 (20.0)	18 (10.4)	18 (19.8)	51 (12.8)
Non-blinded	Yes	249 (21.7)	46 (34.8)	8 (8.4)	42 (24.3)	5 (5.5)	95 (23.9)
Non-blinded	No	456 (39.7)	53 (40.2)	30 (31.6)	45 (26.0)	16 (17.6)	195 (49.1)
OR (95%)		2.6 (2.0 to 3.3)	2.7 (1.1 to 6.1)	7.5 (2.9 to 19.5)	4.0 (2.1 to 7.9)	9.2 (3.0 to 28.9)	2.3 (1.4 to 3.5)
Patients	**Incomplete outcome data**						
Blinded	Complete	338 (30.1)	28 (21.2)	44 (46.3)	74 (42.8)	56 (61.5)	78 (19.6)
Blinded	Incomplete	100 (8.9)	5 (3.8)	13 (13.7)	12 (6.9)	14 (15.4)	29 (7.3)
Non-blinded	Complete	433 (38.6)	75 (56.8)	15 (15.8)	69 (39.9)	13 (14.3)	150 (37.8)
Non-blinded	Incomplete	252 (22.4)	24 (18.2)	23 (24.2)	18 (10.4)	8 (8.8)	140 (35.3)
OR (95%)		2.0 (1.5 to 2.6)	1.8 (0.6 to 5.2)	5.2 (2.1 to 12.7)	1.6 (0.7 to 3.6)	2.5 (0.9 to 7.1)	2.5 (1.5 to 4.1)
Healthcare provider	**Outcome assessor**						
Blinded	Blinded	248 (33.7)	0 (0)	48 (50.5)	0 (0)	48 (65.8)	100 (25.2)
Blinded	Non-blinded	1 (0.1)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0.3)
Non-blinded	Blinded	136 (18.5)	0 (0)	47 (49.5)	0 (0)	25 (34.2)	99 (24.9)
Non-blinded	Non-blinded	350 (47.6)	0 (0)	0 (0)	0 (0)	0 (0)	197 (49.6)
OR (95%)		638.2 (88.7 to 4594.5)		1.0 (0.0 to 52.5)		1.9 (0.0 to 98.7)	199.0 (27.4 to 1447.8)
Healthcare provider	**Allocation concealment**						
Blinded	Yes	243 (21.1)	16 (12.1)	33 (34.7)	74 (42.8)	45 (49.5)	54 (13.6)
Blinded	No	168 (14.6)	11 (8.3)	15 (15.8)	19 (11.0)	16 (17.6)	47 (11.8)
Non-blinded	Yes	267 (23.2)	53 (40.2)	13 (13.7)	36 (20.8)	12 (13.2)	97 (24.4)
Non-blinded	No	471 (41.0)	52 (39.4)	34 (35.8)	44 (25.4)	18 (19.8)	199 (50.1)

OR (95%)		2.6 (2.0 to 3.3)	1.4 (0.6 to 3.4)	5.8 (2.4 to 13.9)	4.8 (2.4 to 9.3)	4.2 (1.7 to 10.7)	2.4 (1.5 to 3.7)
Healthcare provider	**Incomplete outcome data**						
Blinded	Complete	319 (28.4)	23 (17.4)	40 (42.1)	81 (46.8)	49 (53.8)	77 (19.4)
Blinded	Incomplete	88 (7.8)	4 (3.0)	8 (8.4)	12 (6.9)	12 (13.2)	24 (6.0)
Non-blinded	Complete	452 (40.2)	80 (60.6)	19 (20.0)	62 (35.8)	20 (22.0)	151 (38.0)
Non-blinded	Incomplete	264 (23.5)	25 (18.9)	28 (29.5)	18 (10.4)	10 (11.0)	145 (36.5)
OR (95%)		2.1 (1.6 to 2.8)	1.8 (0.6 to 5.7)	7.4 (2.8 to 19.2)	2.0 (0.9 to 4.4)	2.0 (0.8 to 5.5)	3.1 (1.8 to 5.1)
Outcome assessor	**Allocation concealment**						
Blinded	Yes	200 (27.2)	0 (0)	46 (48.4)	0 (0)	41 (56.2)	99 (24.9)
Blinded	No	184 (25.1)	0 (0)	49 (51.6)	0 (0)	32 (43.8)	100 (25.2)
Non-blinded	Yes	94 (12.8)	0 (0)	0 (0)	0 (0)	0 (0)	52 (13.1)
Non-blinded	No	256 (34.9)	0 (0)	0 (0)	0 (0)	0 (0)	146 (36.8)
OR (95%)		3.0 (2.2 to 4.0)		0.9 (0.0 to 48.3)		1.3 (0.0 to 66.1)	2.8 (1.8 to 4.2)
Outcome assessor	**Incomplete outcome data**						
Blinded	Complete	267 (37.7)	0 (0)	59 (62.1)	0 (0)	55 (75.3)	126 (31.7)
Blinded	Incomplete	103 (14.5)	0 (0)	36 (37.9)	0 (0)	18 (24.7)	73 (18.4)
Non-blinded	Complete	190 (26.8)	0 (0)	0 (0)	0 (0)	0 (0)	102 (25.7)
Non-blinded	Incomplete	149 (21.0)	0 (0)	0 (0)	0 (0)	0 (0)	96 (24.2)
OR (95%)		2.0 (1.5 to 2.8)		1.6 (0.0 to 84.0)		3.0 (0.1 to 156.6)	1.6 (1.1 to 2.4)

Appendix Table 3 Combinations of blinding status of patients, healthcare providers and outcome assessors in main analyses

i) Analysis Ia. Number of trials with different combinations of blinding status of trial groups (irrespective of meta-analysis) (N=132)

	Patients blind	Patients non-blind
Healthcare providers blind	26	1
Healthcare providers non-blind	7	98
Total	33	99

ii) Analysis Ib (N=95)

	Patients blind	Patients non-blind
Healthcare providers blind and outcome assessors blind	48	0
Healthcare providers non-blind and outcome assessors blind	9	38
Healthcare providers blind and outcome assessors non-blind	0	0
Healthcare providers non-blind and outcome assessors non-blind	0	0
Total	57	38

iii) Analysis IIa (N=173)

	Healthcare providers blind	Healthcare providers non-blind
Patients blind	84	2
Patients non-blind	9	78
Total	93	80

iv) Analysis IIb (N=91)

		Healthcare providers blind	Healthcare providers non-blind
Patient reported	Patients blind	13	5

outcomes	Patients non-blind	0	0
Observer reported outcomes	Patients blind and outcome assessors blind	48	4
	Patients non-blind and outcome assessors blind	0	21
	Patients blind and outcome assessors non-blind	0	0
	Patients non-blind and outcome assessors non-blind	0	0
Total		61	30

v) Analysis III (N=397)

	Outcome assessors blind	Outcome assessors non-blind
Patients blind and healthcare providers blind	98	1
Patients blind and healthcare providers non-blind	5	3
Patients non-blind and healthcare providers blind	2	0
Patients non-blind and healthcare providers non-blind	94	194
Total	199	198

Appendix Table 4 Calculated 95% ranges in underlying distributions across meta-analyses, and across trials within a meta-analysis, of difference in effect estimates between blinded and non-blinded trials.

Analysis	Overall, average, ROR (95% CrI)	Between–meta-analysis variability in average bias, SD (φ) (95% CrI)	Estimated 95% range in distribution of bias (ROR) across meta-analyses*	Between-trial within-meta-analysis variability in bias (Average increase in between-trial heterogeneity), SD (κ) (95% CrI)	Estimated 95% range in distribution of bias (ROR) across trials within a meta-analysis**
(Ia) The effect of blinding patients in trials with patient-reported outcomes	0.91 (0.61, 1.34)	0.20 (0.01, 0.74)	0.61, 1.36	0.22 (CrI, 0.02 to 0.60)	0.58, 1.41
(Ib) The effect of blinding patients in trials with blinded observer-reported outcomes	0.98 (0.69, 1.39)	0.11 (0.01, 0.55)	0.79, 1.22	0.10 (CrI, 0.01 to 0.60)	0.80, 1.18
(IIa) The effect of blinding healthcare providers in trials with healthcare provider decision outcomes	1.01 (0.84, 1.19)	0.06 (0.01, 0.26)	0.90, 1.14	0.06 (CrI, 0.01 to 0.30)	0.89, 1.14
(IIb) The effect of blinding healthcare providers in trials with blinded observers/patients assessing the outcome	0.97 (0.64, 1.45)	0.13 (0.01, 0.82)	0.74, 1.25	0.10 (CrI, 0.01 to 0.59)	0.79, 1.18
(III) The effect of blinding outcome assessors (i.e. observers) in trials with subjective outcomes	1.01 (0.86, 1.18)	0.09 (0.01, 0.31)	0.84, 1.21	0.05 (CrI, 0.01 to 0.22)	0.91, 1.12
The effect of double-blinding BRANDO study*	0.87 (0.79, 0.96)	0.14 (0.03, 0.28)	0.66, 1.14	0.14 (0.02, 0.30)	0.66, 1.14
The effect of high or unclear risk of bias for the domain	0.87 (0.80, 0.93)	0.12 (0.02, 0.24)	0.69, 1.10	0.10 (0.02, 0.25)	0.72, 1.06

blinding ROBES study*					

* Limits around the overall average within which 95% of values of average within-meta-analysis bias (ROR) are estimated to lie, across meta-analyses.
** Limits around the meta-analysis average bias within which 95% of values of bias in individual trials are estimated to lie, for a meta-analysis set to have the average bias equal to the overall, average, bias, as an example.
***Values from the BRANDO study analysis on the impact of lack of or unclear double blinding (3).
****Values from the ROBES study analysis on the impact of high or unclear risk of bias for the domain blinding (6).

Appendix Table 5 Additional secondary analyses

	N (MA, trial)	ROR (95% CrI)	φ* (95% CrI)	κ** (95% CrI)
The effect of blinding patients in trials with the following outcomes:				
Private patient-reported outcomes	(14, 120)	1.06 (0.67, 1.69)	0.22 (0.02, 0.85)	0.32 (0.02, 0.63)
Patient and Observer-reported outcomes (blinded) with mixed outcomes	(34, 277)	0.94 (0.74, 1.19)	0.11 (0.01, 0.48)	0.12 (0.01, 0.52)
Patient and Observer-reported outcomes (blinded) without mixed outcomes	(32, 267)	0.95 (0.76, 1.21)	0.11 (0.01, 0.44)	0.13 (0.01, 0.52)
The effect of blinding healthcare providers in trials with the following outcomes:				
Observer-reported outcomes assessed by blind observers	(11, 78)	1.05 (0.56, 1.58)	0.11 (0.01, 0.67)	0.11 (0.01, 0.61)
All outcomes jointly including mixed	(42, 250)	1.01 (0.86, 1.19)	0.06 (0.01, 0.26)	0.06 (0.01, 0.26)
The effect of blinding outcome assessors in trials with the following outcomes:				
Any objective outcomes	(15, 207)	0.94 (0.61, 1.26)	0.23 (0.02, 0.82)	0.13 (0.02, 0.39)
All-cause mortality	(11, 168)	0.91 (0.51, 1.31)	0.29 (0.02, 1.15)	0.10 (0.02, 0.32)
Subjective interactive	(15, 145)	1.22 (0.94, 1.58)	0.08 (0.01, 0.39)	0.16 (0.01, 0.53)
Subjective pure observation	(31, 252)	0.92 (0.76, 1.12)	0.10 (0.01,, 0.39)	0.05 (0.01, 0.20)
Observer-reported outcomes without mixed outcomes	(61, 604)	1.01 (0.88, 1.14)	0.10 (0.01, 0.33)	0.08 (0,01, 0.22)
Observer-reported outcomes including mixed outcomes	(65, 624)	1.01 (0.89, 1.14)	0.09 (0.01, 0.30)	0.08 (0.01, 0.21)

* Between meta-analysis standard deviation in mean bias, ** Standard deviation increase in between trial heterogeneity

The following secondary analyses were planned but were not conducted because there were less than 10 meta-analyses:

The effect of blinding patients in trials with the following outcomes:

- Non-private patient-reported outcomes (4 MAs, 12 trials)
- Mixed outcomes (2 MAs, 7 trials)

The effect of blinding healthcare providers in trials with the following outcomes:

- Patient-reported outcomes by blinded patients (3 MAs, 18 trials)
- Mixed outcomes (2 MAs, 7 trials)

The effect of blinding outcome assessors in trials with the following outcomes:

- Other objective outcomes (4 MAs, 39 trials)
- Mixed outcomes (4 MAs, 12 trials)

The main analyses looking at only the trials scored as "definitely yes" vs trials scored as "definitely no":

- (Ia) "definitely yes" vs. "definitely no" (6 MAs, 23 trials)
- (Ib) "definitely yes" vs. "definitely no" (4 MAs, 15 trials)
- (IIa) "definitely yes" vs. "definitely no" (3 MAs, 13 trials)
- (IIb) "definitely yes" vs. "definitely no" (4 MAs, 16 trials)
- (III) "definitely yes" vs. "definitely no" (8 MAs, 54 trials)

The main analyses, hypothesis of harm:

- (Ia) (3 MAs, 23 trials)
- (Ib) 0
- (IIa) (6 MAs, 53 trials)
- (IIb) (1 MA, 5 trials)
- (III) (5 MAs, 16 trials)

The main analyses, by binary or continuous outcomes

- (Ia) (Binary: 9 MAs, 42 trials vs. Continuous: 9 MAs, 90 trials)
- (Ib) (Binary: 11 MAs, 78 trials vs. Continuous: 3 MAs, 17 trials)
- (IIa) (Binary: 25 MAs, 151 trials vs. Continuous: 4 MAs, 22 trials)
- (IIb) (Binary: 11 MAs, 82 trials vs. Continuous: 2 MAs, 9 trials)

Commentary: There was no clear difference according to type of outcome (binary vs. continuous) in main analysis III, in which there were more than 10 meta-analyses with continuous and binary outcomes, respectively, as reported in the main paper. In her PhD dissertation (7), Gemma Clayton analyzed the issue further based on Sterne's two-step model. Of the remaining four analyses, with less than 10 MA per analysis, there was a difference by type of outcome in one. In the analysis of patient blinding in trials with patient-reported outcomes, 9 meta-analyses with binary outcomes showed a large impact of blinding, and 9 meta-analyses with continuous outcomes showed no statistically significant effect of blinding, and a point estimate > 1 (indicating lower effects in non-blinded trials). We interpret this as a random event. (7).

The following post hoc analyses were considered but not conducted because there were less than 10 meta-analyses:

The main analyses by type of comparator (active control vs. inactive control (placebo/no treatment/standard care)):

- Ia (Active: 6 MAs, 20 trials vs. Inactive: 12 MAs, 112 trials)

- Ib (Active: 6 MAs, 29 trials vs. Inactive: 8 MAs, 66 trials)

- IIa (Active: 8 MAs, 42 trials vs. Inactive: 21 MAs, 131 trials)

- IIb (Active: 6 MAs, 29 trials vs. Inactive: 7 MAs, 62 trials)

Appendix Table 6 Main analyses based on the label-invariant model by Rhodes and colleagues

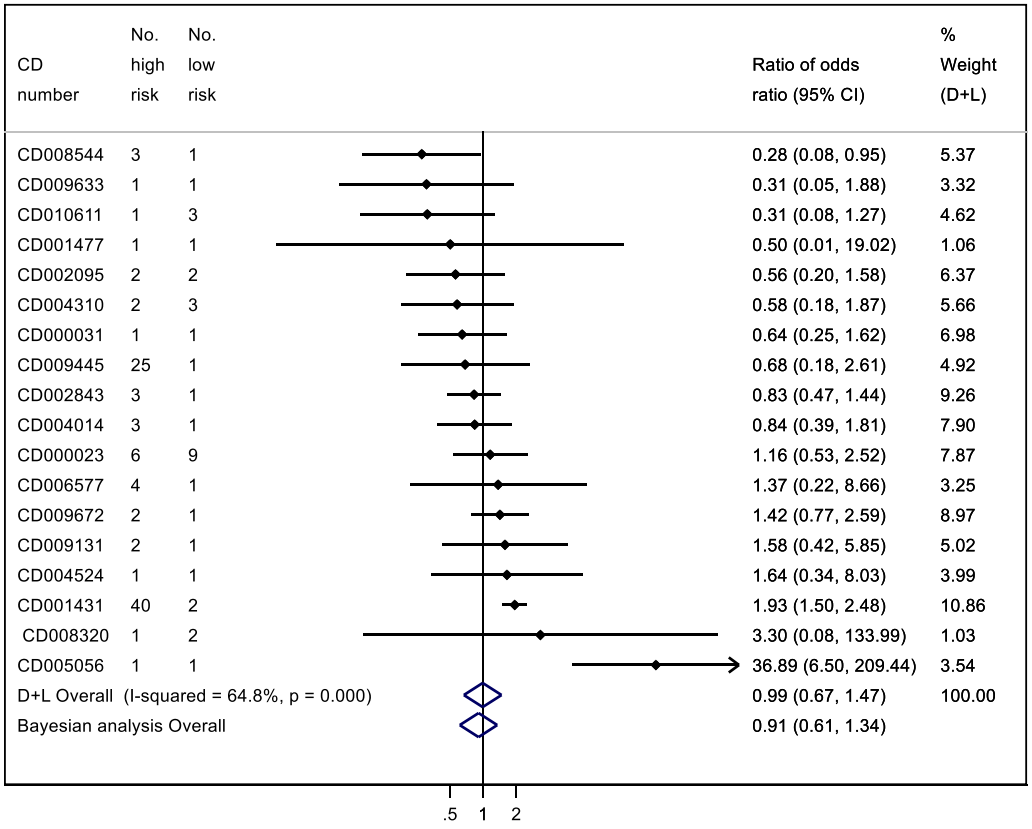
Analysis	Proportion of MAs with 1 low risk study	Parameters	Welton model	Label-invariant model
Ia (18, 132)	12/18=67%	ROR (95% CrI)	0.91 (0.61, 1.34)	0.99 (0.67, 1.46)
		Phi (95% CrI)	0.20 (0.01, 0.74)	0.17 (0.01, 0.80)
		Kappa (95% CrI)	0.22 (0.02 to 0.60)	
		Lambda (95% CrI)		0.56 (0.10, 2.6)
Ib (14,95)	6/14=43%	ROR (95% CrI)	0.98 (0.69, 1.39)	0.98 (0.71, 1.40)
		Phi (95% CrI)	0.11 (0.01, 0.55)	0.10 (0.01, 0.53)
		Kappa (95% CrI)	0.10 (0.01 to 0.60)	
		Lambda (95% CrI)		0.84 (0.13, 5.15)
IIa (29, 173)	15/29=52%	ROR (95% CrI)	1.01 (0.84, 1.19)	0.96 (0.77, 1.15)
		Phi (95% CrI)	0.06 (0.01, 0.26)	0.07 (0.01, 0.22)
		Kappa (95% CrI)	0.06 (0.01 to 0.30)	
		Lambda (95% CrI)		0.75 (0.12, 4.20)
IIb (13, 91)	4/13=31%	ROR (95% CrI)	0.97 (0.64, 1.45)	0.98 (0.64, 1.46)
		Phi (95% CrI)	0.13 (0.01, 0.82)	0.12 (0.01, 0.76)
		Kappa (95% CrI)	0.10 (0.01 to 0.59)	
		Lambda (95% CrI)		0.56 (0.10, 2.75)
III (46, 397)	15/46=33%	ROR (95% CrI)	1.01 (0.86, 1.18)	1.01 (0.86, 1.21)
		Phi (95% CrI)	0.09 (0.01, 0.31)	0.09 (0.01, 0.33)
		Kappa (95% CrI)	0.05 (0.01 to 0.22)	
		Lambda (95% CrI)		0.41 (0.09, 1.49)

Interpretation/comment

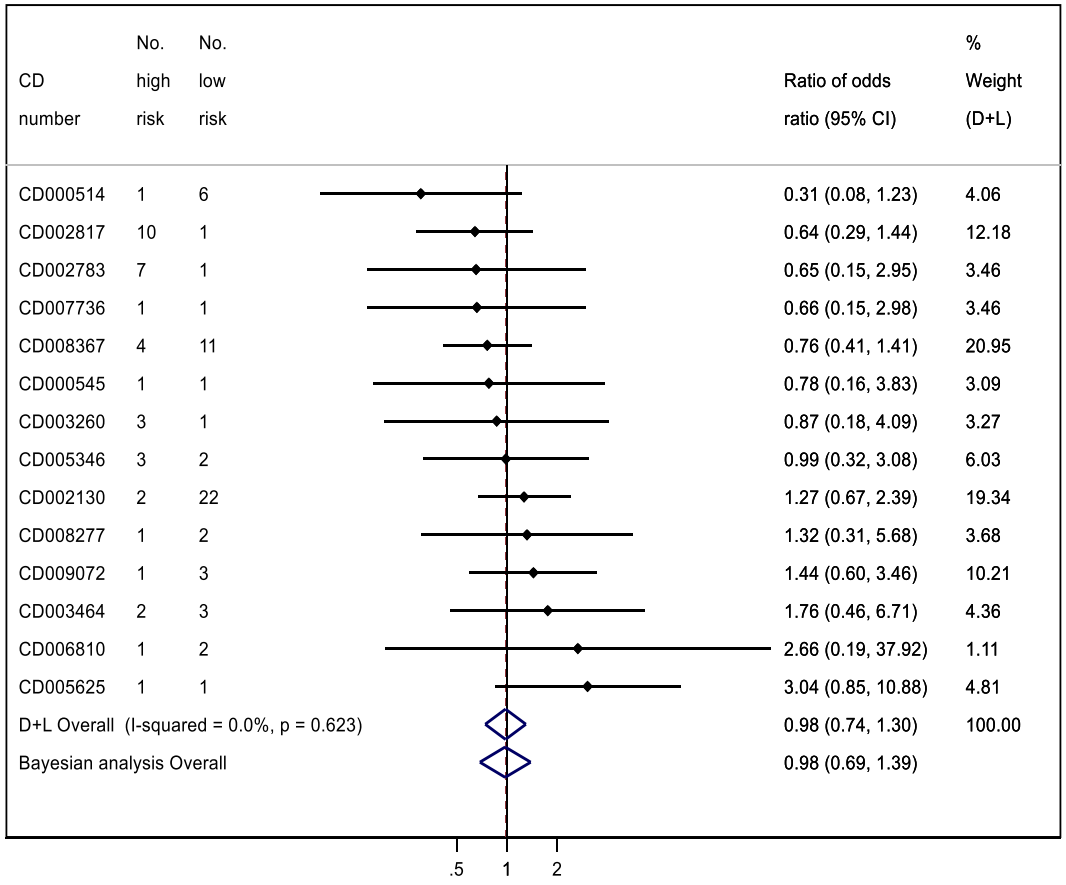
This comparison between the Welton model (which was defined in our protocol) and the Rhodes model (which was not published until after we planned our study) was a post-hoc analysis. The estimates of ROR and of between-meta-analyses heterogeneity in bias from both models were very similar. The estimates of Lambda in the Rhodes model (proportion between SDs) are not directly comparable with Kappa in the Welton model (increase in SD), but all five point estimates were below 1, indicating a possible reduction in heterogeneity among non-blinded trials. However, Lambda was estimated with very low precision, and the upper credibility limits are consistent with a significant increase.

Appendix Figure 1 RORs from individual meta-analyses and from analyses combined across all meta-analyses. Results for individual meta-analyses are frequentist estimates with confidence intervals, based on comparing the summary odds ratio from studies with the study characteristic of interest with the summary odds ratio from studies without the characteristic. The overall estimates of RORs marked Bayesian analysis Overall are results based on the Bayesian hierarchical model (Welton model) described in the main text. CD numbers are identifiers of individual Cochrane reviews, from the Cochrane Database of Systematic Reviews.

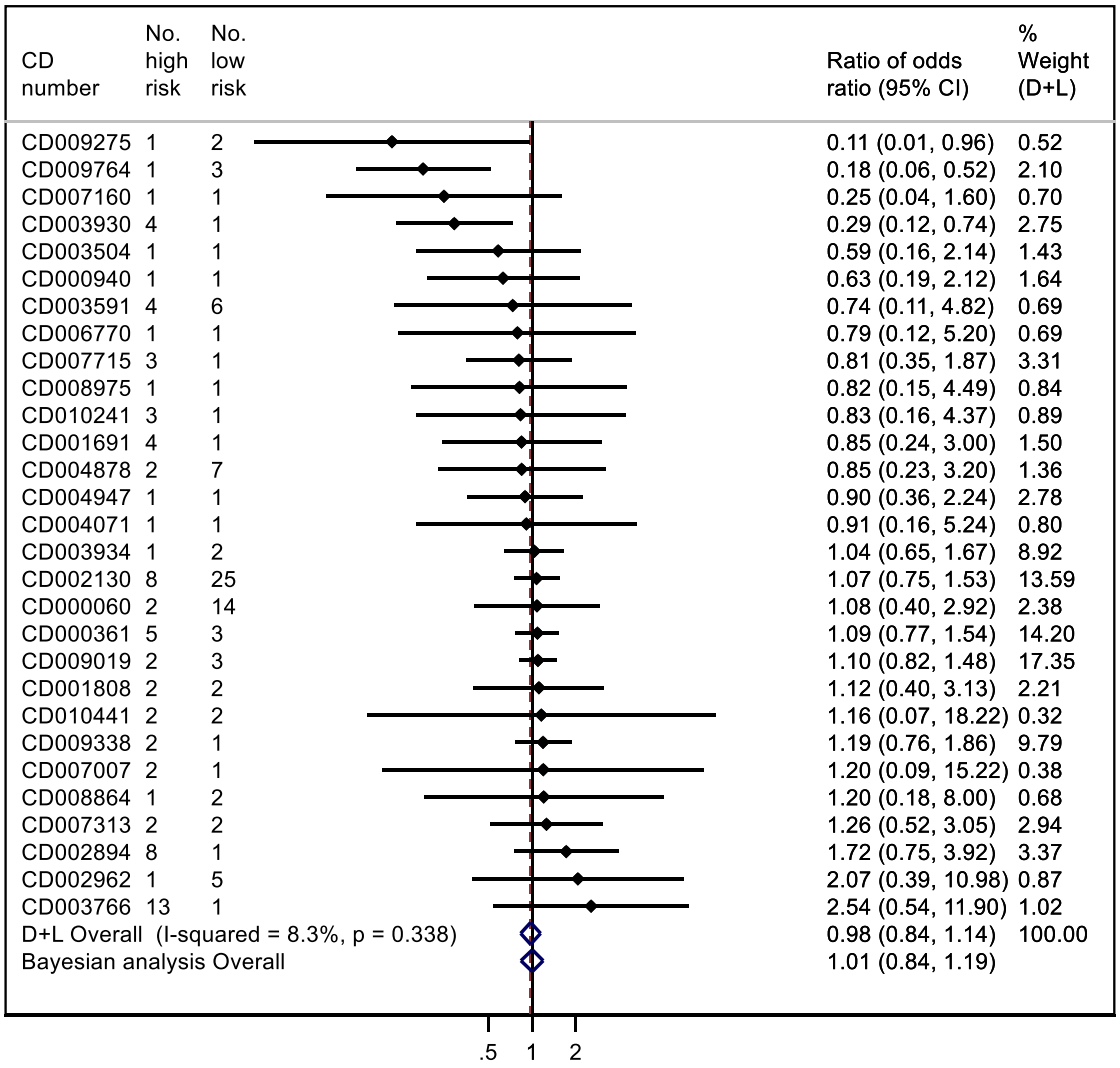
(Ia) The effect of blinding patients in trials with patient-reported outcomes



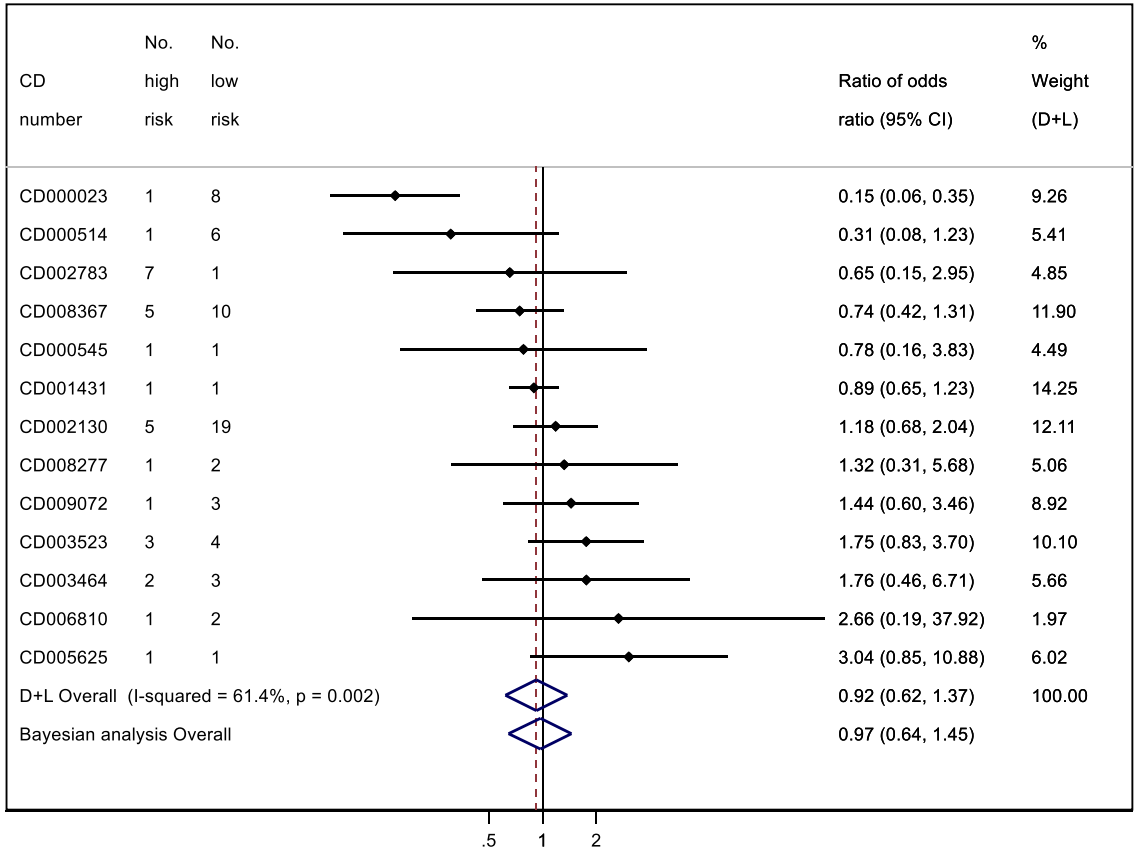
(Ib) The effect of blinding patients in trials with blinded observer-reported outcomes



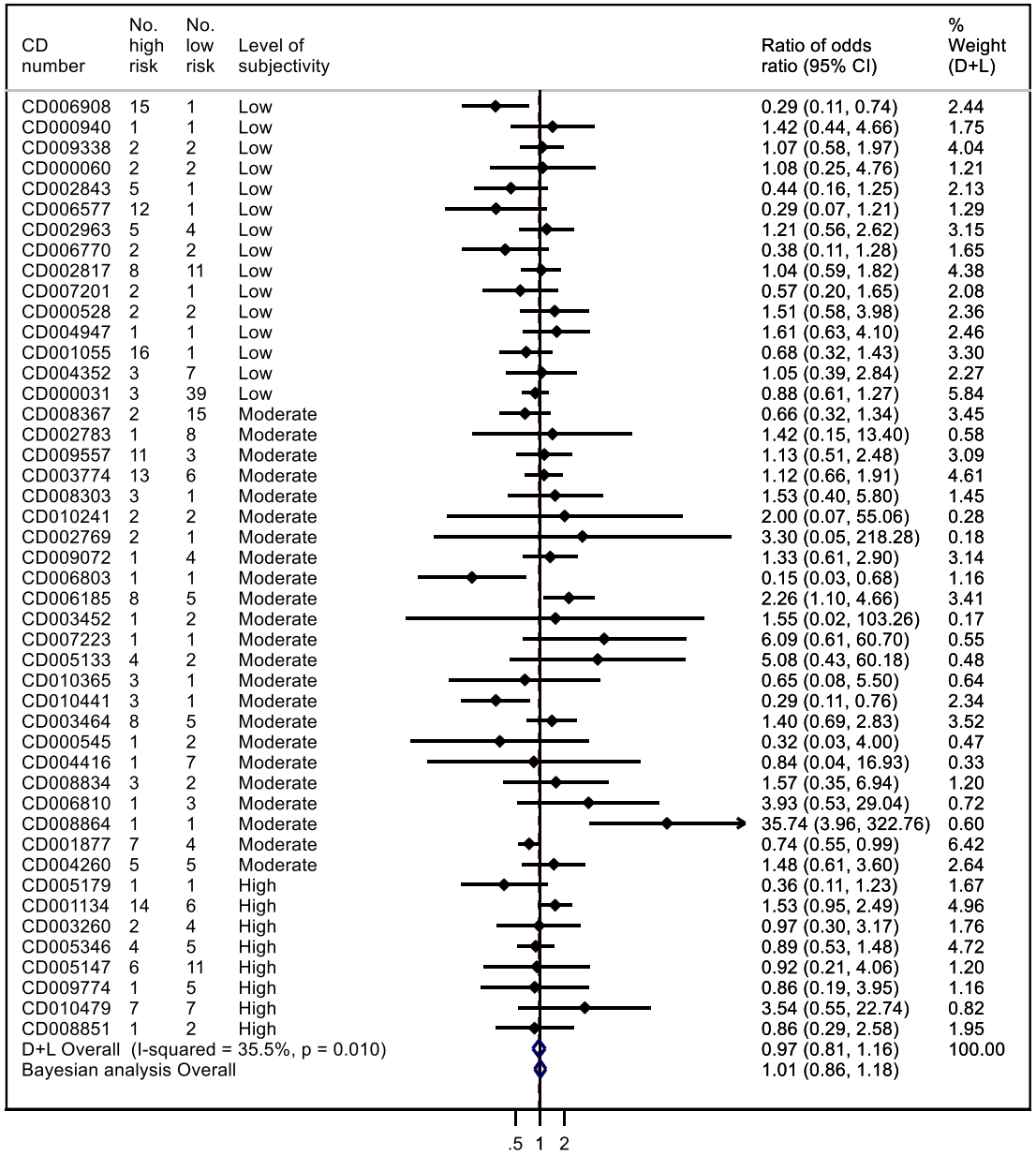
(IIa) The effect of blinding healthcare providers in trials with healthcare provider decision outcomes



(IIb) The effect of blinding healthcare providers in trials with blinded observers/patients assessing the outcome



(III) The effect of blinding outcome assessors (i.e. observers) in trials with subjective outcomes


CD number
No. high risk
No. low risk
Level of subjectivity
Ratio of odds ratio (95% CI)
% Weight (D+L)
CD006908 15 1 Low 0.29 (0.11, 0.74) 2.44
CD000940 1 1 Low 1.42 (0.44, 4.66) 1.75
CD009338 2 2 Low 1.07 (0.58, 1.97) 4.04
CD000060 2 2 Low 1.08 (0.25, 4.76) 1.21
CD002843 5 1 Low 0.44 (0.16, 1.25) 2.13
CD006577 12 1 Low 0.29 (0.07, 1.21) 1.29
CD002963 5 4 Low 1.21 (0.56, 2.62) 3.15
CD006770 2 2 Low 0.38 (0.11, 1.28) 1.65
CD002817 8 11 Low 1.04 (0.59, 1.82) 4.38
CD007201 2 1 Low 0.57 (0.20, 1.65) 2.08
CD000528 2 2 Low 1.51 (0.58, 3.98) 2.36
CD004947 1 1 Low 1.61 (0.63, 4.10) 2.46
CD001055 16 1 Low 0.68 (0.32, 1.43) 3.30
CD004352 3 7 Low 1.05 (0.39, 2.84) 2.27
CD000031 3 39 Low 0.88 (0.61, 1.27) 5.84
CD008367 2 15 Moderate 0.66 (0.32, 1.34) 3.45
CD002783 1 8 Moderate 1.42 (0.15, 13.40) 0.58
CD009557 11 3 Moderate 1.13 (0.51, 2.48) 3.09
CD003774 13 6 Moderate 1.12 (0.66, 1.91) 4.61
CD008303 3 1 Moderate 1.53 (0.40, 5.80) 1.45
CD010241 2 2 Moderate 2.00 (0.07, 55.06) 0.28
CD002769 2 1 Moderate 3.30 (0.05, 218.28) 0.18
CD009072 1 4 Moderate 1.33 (0.61, 2.90) 3.14
CD006803 1 1 Moderate 0.15 (0.03, 0.68) 1.16
CD006185 8 5 Moderate 2.26 (1.10, 4.66) 3.41
CD003452 1 2 Moderate 1.55 (0.02, 103.26) 0.17
CD007223 1 1 Moderate 6.09 (0.61, 60.70) 0.55
CD005133 4 2 Moderate 5.08 (0.43, 60.18) 0.48
CD010365 3 1 Moderate 0.65 (0.08, 5.50) 0.64
CD010441 3 1 Moderate 0.29 (0.11, 0.76) 2.34
CD003464 8 5 Moderate 1.40 (0.69, 2.83) 3.52
CD000545 1 2 Moderate 0.32 (0.03, 4.00) 0.47
CD004416 1 7 Moderate 0.84 (0.04, 16.93) 0.33
CD008834 3 2 Moderate 1.57 (0.35, 6.94) 1.20
CD006810 1 3 Moderate 3.93 (0.53, 29.04) 0.72
CD008864 1 1 Moderate 35.74 (3.96, 322.76) 0.60
CD001877 7 4 Moderate 0.74 (0.55, 0.99) 6.42
CD004260 5 5 Moderate 1.48 (0.61, 3.60) 2.64
CD005179 1 1 High 0.36 (0.11, 1.23) 1.67
CD001134 14 6 High 1.53 (0.95, 2.49) 4.96
CD003260 2 4 High 0.97 (0.30, 3.17) 1.76
CD005346 4 5 High 0.89 (0.53, 1.48) 4.72
CD005147 6 11 High 0.92 (0.21, 4.06) 1.20
CD009774 1 5 High 0.86 (0.19, 3.95) 1.16
CD010479 7 7 High 3.54 (0.55, 22.74) 0.82
CD008851 1 2 High 0.86 (0.29, 2.58) 1.95
D+L Overall (I-squared = 35.5%, p = 0.010) 0.97 (0.81, 1.16) 100.00
Bayesian analysis Overall 1.01 (0.86, 1.18)
.5
1
2

Appendix References

1. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and computing. 2000;10(4):325-37.
2. Welton NJ, Ades AE, Carlin JB, Altman DG, Sterne JAC. Models for potentially biased evidence in meta-analysis using empirically based priors. Journal of the Royal Statistical Society: Series A (Statistics in Society). 2009;172(1):119-36.
3. Savovic J, Jones HE, Altman DG, Harris RJ, Juni P, Pildal J, et al. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. Ann Intern Med. 2012;157(6):429-38.
4. Lambert PC, Sutton AJ, Burton PR, Abrams KR, Jones DR. How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. Statistics in medicine. 2005;24(15):2401-28.
5. Rhodes KM, Mawdsley D, Turner, RM, Jones HE, Savović J, Higgins JPT. Label-invariant models for the analysis of meta-epidemiological data. Stat Med. 2018;37(1):60-70).
6. Savović J, Turner RM, Mawdsley D, et al. Association Between Risk-of-Bias Assessments and Results of Randomized Trials in Cochrane Reviews: The ROBES Meta-Epidemiologic Study. Am J Epidemiol. 2018;187(5):1113-22.
7. Clayton GL. External Evidence Synthesis in the Design and Analysis of Trials. A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of doctor of philosophy in the Faculty of Health Sciences, Bristol Medical School, 2019.