Infectivity of human norovirus in live challenge trials: a systematic review and meta-analysis

W. Zane Billings1, Anne Marie Dye2, and Andreas Handel1

**Introduction**: Human norovirus is the most common cause of acute gastroenteritis and food-borne illness in the United States. The wide variety of potential routes of exposure make estimates of transmissibility from outbreak data somewhat unreliable. In contrast, human challenge studies involve a controlled inoculation, where experimental settings and route of transmission can be controlled, among other factors. Our objective was to find all available norovirus challenge study data, and from these data, determine whether the infectivity of norovirus is consistent across human challenge studies.

**Methods**: We conducted a systematic review of the literature in order to find all studies which report using norovirus challenge data, and from these studies, we abstract data to obtain unique challenge cohort data. From unique cohorts, we conduct a meta-analysis of the proportion of individuals infected during each study, including subgroup analyses by study risk of bias, inoculum genogroup, and FUT2 participant genotype control, and *post hoc* outlier analyses to determine sources of heterogeneity. We additionally assessed the possibility of publication bias using a funnel plot.

**Results**: We identified 65 reports for inclusion, and from those we found 20 studies which reported original collection of challenge data. We find a high degree of heterogeneity among studies, which is somewhat explained by outliers, and not explained by any abstracted covariates. Removing identified outliers reduces heterogeneity, but the outliers have no distinguishing features which set them apart from the other studies. The summary analyses are not meaningful due to the high degree of unexplainable heterogeneity.

**Discussion**: Overall, we find that the attack rate of norovirus is not consistent across human challenge studies. Our results suggest that either proportion of infected individuals in a norovirus challenge study is naturally heterogeneous, or is influenced primarily by unmeasured covariates. As potential infection rates may be dependent on several characteristics of research protocols, norovirus inoculum, research staff, and study participants, our results seem reasonable.

The primary source of funding for the review was the University of Georgia Graduate School. The review was not registered.

1 Department of Epidemiology and Biostatistics, College of Public Health, University of Georgia  
2 Department of Environmental Health, College of Public Health, University of Georgia

Word count: 4870

Number of tables: 3

Number of figures: 18

# Introduction

Human norovirus (NoV) is a small, round-structured, unenveloped, positive-sense single-strand RNA virus with a linear, unsegmented genome, belonging to the caliciviridae family. NoV illness was first described in 1929 (as “winter vomiting disease”), first identified as an infectious agent in 1968, and visualized in 1972.1,2 Of course, NoV was likely an annoyance long before the advent of virology. The Centers for Disease Control (CDC) estimate that NoV is responsible for over 50% of foodborne illnesses, and is the most common single cause of gastroenteritis in the United States.3 The fecal-oral route is the primary mode of NoV transmission,1 as functional norovirus is shed in the feces of infected patients, both symptomatic,4 and asymptomatic,5 as well as by recovered patients for weeks or longer.6,7 The vomit-oral route (through either contact transmission or particle aerosolization) also plays a large role in confined settings, and surface particles are also likely to act as environmental reservoirs.8 Modeling studies also indicate that norovirus has a low infectious viral load, which could be as low as 10 viral particles.9 The combination of environmental persistence, high infectivity via shedding, and low viral load can make norovirus highly tranmissible, although estimates of transmissibility vary.10

Most norovirus infections are self-limiting, and transmission to an individual can be highly variable due to the number of possible routes of exposure.10 Study of norovirus transmission is further complicated by rapid evolution and differences in tramissibility by genotype,11 as well as heterogeneous host immune and genetic profiles.12,13 For example, natural variation in histo-blood group antigens affects norovirus susceptibility,14 and in particular, individuals who do not have a functional copy of the fucosyltransferase-2 (FUT2) gene are practically immune to the most common genotypes of norovirus,13 but not all genotypes.15 Based on these factors, estimates of transmissibility and infectivity can often be variable by outbreak. However, human challenge studies offer a controlled setting where viral properties of NoV can be examined. During a challenege study, healthy individuals are administered live norovirus after which subjects are closely monitored for the duration of inducted illness. Such challenge studies provide the opportunity to control factors like inoculum strain, inoculum dose, and participant histo-blood group. Recall bias is also less likely to influence the result of challenge studies. Thus, using challenge studies to explore norovirus infectivity can be less prone to certain biases than challenge studies, even if the results are not always generalizable to natural outbreaks.

However, there is no consensus on what human challenge data have been published and are accessible for secondary analysis. Some previous studies have incorporated data from multiple past challenge studies,9,16 but do not claim to span the set of all norovirus challenge studies which have ever been conducted. So, our first goal in this study is to compile a list of reports which reference human norovirus challenge studies, and then to determine which unique studies are reflected by these reports (as one study tends to produce multiple reports). Then, we can conduct a meta-analysis to determine if the attack rate of norovirus consistent across historical challenge studies? Since challenge studies are believed to be more homogeneous than natural outbreak data, we hypothesize that the attack rate should be consistent across challenge studies, or explainable by factors such as host immune heterogeneity or viral genotype.

# Methods

We conducted a systematic review of papers using data from norovirus challenge studies, and from the included reports, we identified unique studies. Then, we abstracted data on the proportion of infected individuals from each study, and conducted a meta-analysis of these data.

## Literature search

We searched two databases to find literature:

* PubMed [(https://www.ncbi.nlm.nih.gov/pubmed/)](https://www.ncbi.nlm.nih.gov/pubmed/) and
* Web of Science [(https://www.webofscience.com/)](https://www.webofscience.com/).

The search terms, included in Table 1, contained terms for norovirus, terms related to human studies, and terms related to the challenge (or volunteer) study design. The databases were searched on September 28, 2021, with no filter for date used during the search.

Reports were selected for inclusion if they referenced a human norovirus challenge study with controlled inoculation of participants. No other inclusion criteria (location, date, study design, etc.) were used. Reports were excluded from the review if they were not primary research articles (i.e. reviews or letters to the editor), not written in English, or not available to the authors. Two reviewers (WZB and AMD) independently reviewed titles and abstracts, with disagreements being resolved by a third reviewer (AH). The reviewers repeated the process for the full-text review.

Table 1: Search strings for the two databases searched.

| **database** | **search strategy** |
| --- | --- |
| PubMed | ("norovirus" [MeSH Major Topic]) AND ("norovirus" OR "Norwalk virus" OR "snow mountain virus" OR  "Norwalk agent" OR "nonbacterial gastroenteritis" OR  "viral gastroenteritis" [Title/Abstract]) AND (human OR challenge OR experimental OR infect\* OR volunteer OR  vaccin\* OR adult OR clinical OR randomized OR  individual [Title/Abstract]) NOT ("mouse" or "murine" or "mice" [Title]) |
| WoS | (TS=("norovirus" OR "Norwalk virus" OR "snow mountain virus" OR  "Norwalk agent" OR "nonbacterial gastroenteritis" OR  "viral gastroenteritis")) AND (TS=(human OR man OR adult OR volunteer)) AND (TS=(volunteer OR challenge OR experimental OR infect\* OR vaccin\* OR  inoculum)) |

From the included reports, the reviewers identified unique studies by examining citations for data sources. We created a directed acyclic graph (DAG) based on citations in reference sections of papers–in this format, the end notes of the DAG are studies with original populations. Only reports which appeared to discuss original data collection and did not cite their data as coming from a previous report were considered to be unique studies.

## Data abstraction

From each study (only reports which included original data collection), one author (WZB) abstracted the following information:

1. Reference information including name of the first author and the year of publication;
2. Study design for the challenge portion of the study (either case series or randomized trial);
3. Demographic information reported for each study including age range, percentage of male participants, percentage of white participants, and location of the study site;
4. Whether the study controlled for participant FUT2 genotype;
5. Any other eligibility criteria for study enrollment;
6. Inoculum dose and genotype;
7. Criteria for reporting infection and illness; and
8. For each combination of FUT2 status, dose, and other study sub-samples (e.g. vaccine vs. placebo in vaccine trials), the number of participants challenged, number of participants with confirmed infection, and the number of participants with confirmed illness.

Reported inoculum strains were standardized to modern nomenclature where possible, criteria for infection and illness were standardized, and inoculum dose was converted to genome equivalent copies (1 RT-PCR unit = 400 genome equivalent copies).17

## Study quality assessment

Study quality was assessed using a modification of the JBI critical appraisal tool for case series studies.18 All included studies were evaluated on this scale by one author (WZB). Studies were assessed on seven domains: clear inclusion criteria, reliability of infection assessment, validity of infection assessment, whether all participants were drawn from the same cohort, complete reporting of relevant demographic information, complete reporting of relevant clinical information, and reporting of study site information. The full rubric for assessing studies is included in the appendix.

## Statistical methods

We fit an overall meta-analysis model for the proportion of infected individuals (proportion infected) by pooling together all subgroups reported within each study. We used a generalized linear mixed-effected modeling approach with logit-transformed proportions (log-odds) as the outcome.19,20 The method of maximum likelihood was used to estimate (between-study heterogeneity). Confidence intervals for pooled effects were calculated using the Knapp-Hartung adjustment,21,22 which is typically sensible.20,23,24 The GLMM method with log-odds as the outcome has been previously recommended in the literature for the meta-analysis of proportions.25 When using a GLMM approach, no weights are estimated for each study, and (the estimate of between-study heterogeneity) can only be estimated through the method of maximum likehood, and a confidence interval for cannot be obtained.20

*A priori* subgroup analyses were conducted to examine the effect of study risk of bias (high risk of bias vs. other studies), norovirus genogroup (GI, GII, or unknown), and whether FUT2 was controlled for in the study. For subgroup analyses, we use the so-called “fixed-effects (plural)” model for between-subgroup differences.20,26,27 Estimated between-study heterogeneity () was assumed to be different for all subgroups, and models within-subgroups were fit used the random-effects GLMM method as previously described. Differences between subgroups were assessed visually and using differences in rather than by conducting any formal statistical test. The effect of year of publication was examined visually, rather than by conducting any additional analyses.

Influence of individual studies on the overall result was analyzed *post-hoc* using three methods. First, we used a simple method which classifies studies as outliers if the estimate confidence interval for the individual study does not overlap with the confidence interval for the pooled estimate.20 Second, we used a leave-one-out approach and manually identified outliers using a Baujat plot28 and diagnostics.29 Third, we used the GOSH (Graphical display Of Study Heterogeneity) method, wherein we fit 1,000,000 models with random subsets of studies included. Then, we plotted the estimated heterogeneity vs. the estimated effect size of all random subset analyses.30 From the GOSH results, we applied three unsupervised clustering algorithms: -means,31 DBSCAN,32 and Gaussian mixture modeling.33 Study over-representation within clusters is used to determine which studies have an undue effect on heterogeneity.20 Overall, outliers were detected by consensus–if any two of the three methods flagged a particular study, that study was designated as an outlier.

Finally, since we have more than 10 studies,34 we assessed publication bias graphically using a contour-enhanced funnel plot35 and numerically using Peters’ test.35 Peters’ test accounts for dependence between the effect size and standard error for effect sizes based on binary outcome data. The common method, Egger’s test,36 does not, so Peters’ test has a lower false positive rate in comparison.20,35

## Software

Reference management was conducted using both EndNote37 for deduplication and searching for missing reference fields, and Zotero38 for archival purposes. Data abstraction and review of reports was conducted using Microsoft Excel 365 (Microsoft Corporation, Santa Rosa, CA, USA) and Google Sheets (Google, Mountain View, CA, USA).

All analyses were conducted using R version 4.1.1.39 The packages meta,40,41 metafor,42,43 and dmetar20,44 were used for meta-analysis. Figures were generated using the analysis packages, robvis,45 and PRISMA2020.46–48 Tables were generated using the package flextable.49 This report was generated using R Markdown with the packages rmarkdown,50–52 knitr,53–55 and bookdown.56,57 Several additional packages were used for data cleaning and wrangling, and for miscellaneous tasks.58–63 A complete printout of the R session information can be found in the Appendix.

# Results

## Identification of studies

Overall, 5301 non-duplicate records were obtained from searching PubMed and Web of Science. After reviewing the titles and abstracts of records, 98 reports were sought for retrieval. Of those, two reports could not be obtained, and so 96 reports were retrieved for full-text review. After applying the eligibility criteria, we selected 65 reports for inclusion in the study. The PRISMA diagram for study inclusion is shown in Figure 1.

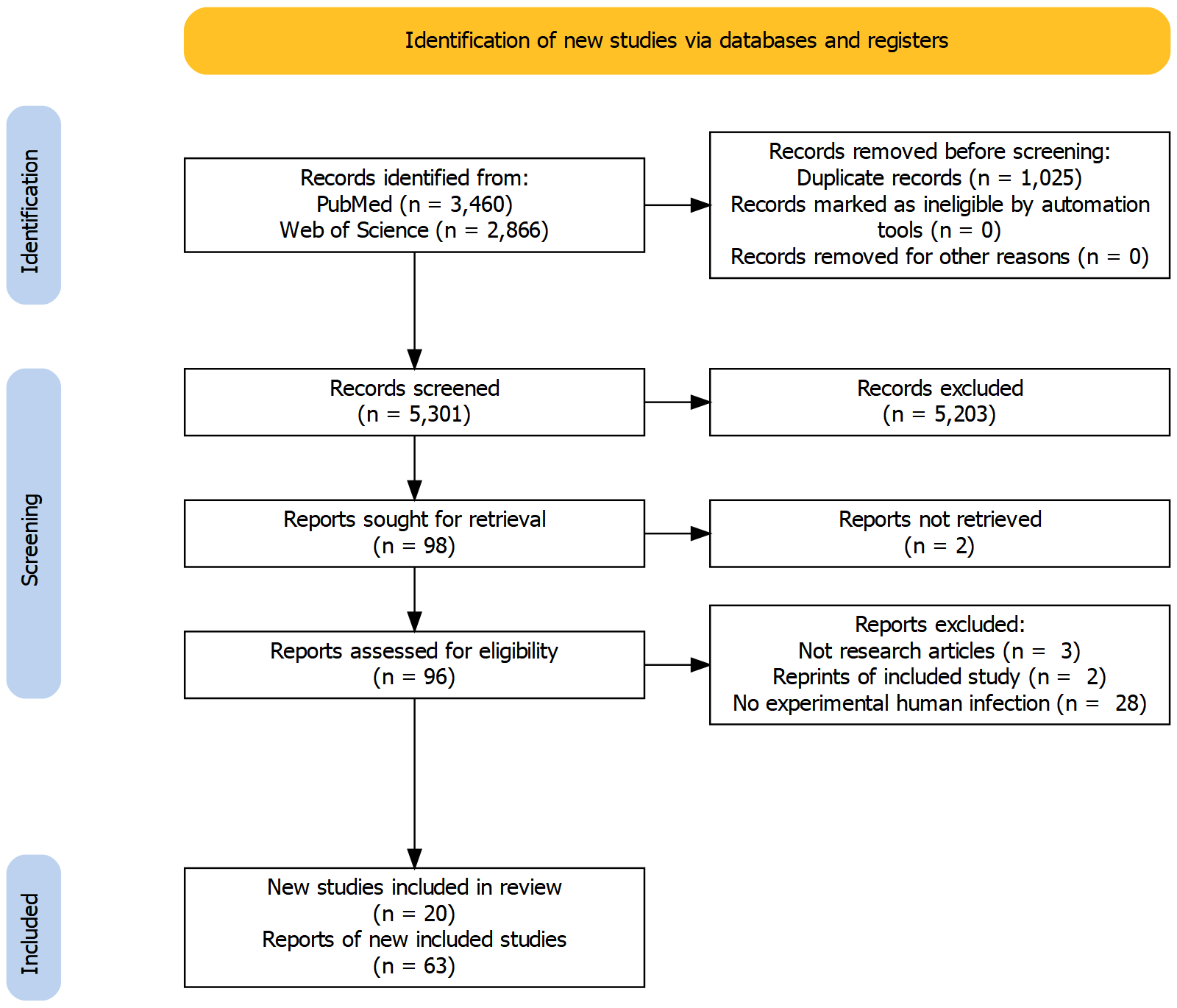


Figure 1: PRISMA 2020 flow diagram showing how many records and reports were identified at each stage of the review process.

From the list of included reports, we determined the reports which reflect original data collection by examining the methods reported in the paper. A DAG showing citations pointing to original data sources is shown in Figure 2. In total, 20 papers reflecting original sources of data were selected for inclusion in synthesis results.6,9,13,17,64–79 Of the 65 reports included,6,9,13,16,17,64–123 7 appeared to reflect original sources of data, but either did not provide the results necessary for synthesis, or provided the results in a format that was not usable for data extraction.93,94,104,106,107,111,115

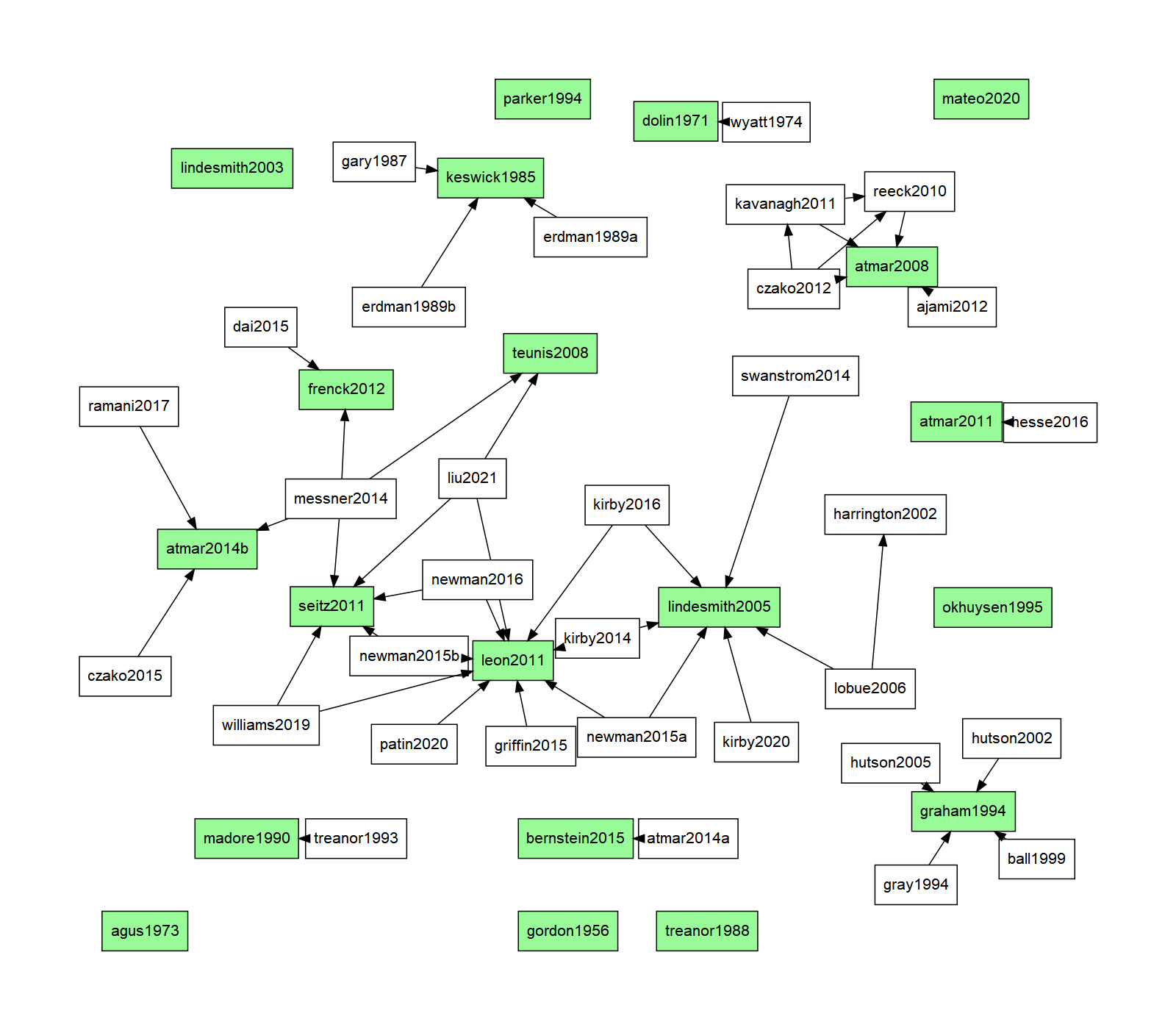


Figure 2: DAG showing citations which indicated the terminal study as the source of the data. The end notes colored in green are the reports which were identified as original data collection studies.

## Study characteristics

The abstracted study characteristics for the 20 included papers are shown in Table 2. The earliest study was conducted in 1956, and the most recent study was conducted in 2020, with 15/20 of the studies conducted in 1990 or later, and 11/20 of the studies conducted in 2000 or later. All studies were conducted in the United States, in at least 8 different states. Most studies (15/20) reported information on the age range of participants, and all studies only included adults, as expected for challenge studies. The three oldest studies only included male subjects, but later studies tended to include an equal balance of male and female participants, where reported. Racial diversity in cohorts was reported in less than half of the included studies, and varied significantly by study. Half of the included studies (10/20) either reported results stratified by FUT2 (secretor) genotype or only recruited participants with FUT2+ genotype. Three studies only included participants with A or O blood types. Only one study (Frenck 2012) reported low pre-challenge titer as an eligibility criterion for subject recruitment. Note that the two oldest studies recruited participants at least partially from incarcerated populations–while the studies claim that the subjects are volunteers, neither study reports methods for preventing coercive recruiting, nor mentions how incarcerated subjects were compensated. These studies likely fail to meet modern ethical guidelines, but we chose to include the data in our synthesis in the hopes that some good can come out of past scientific misconduct.

Table 2: Abstracted demographic information and other study characteristics for the 20 unique original data collection studies which were identified.

| **study** | **n1** | **e2** | **RoB** | **age** | **% male** | **% white** | **setting** | **FUT23** | **other** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Gordon 1956 | 42 | 31 | High | 21 - 30 | 100 |  | NY | No | Incarcerated subjects |
| Dolin 1971 | 16 | 10 | High | 18 - 45 | 100 |  | MD | No | Incarcerated subjects |
| Agus 1973 | 7 | 4 | Med | 19 - 21 | 100 |  | MD | No |  |
| Keswick 1985 | 32 | 19 | High |  |  |  | TX | No |  |
| Treanor 1988 | 10 | 9 | High | adults |  |  | NY | No |  |
| Madore 1990 | 61 | 51 | High | 18 - 35 |  |  | NY | No |  |
| Graham 1994 | 50 | 41 | High | 19 - 39 | 41 | 84 | TX | No |  |
| Parker 1994 | 14 | 11 | High | adults |  |  | multiple | Yes |  |
| Okhuysen 1995 | 72 | 47 | High | adults |  |  | TX | No |  |
| Lindesmith 2003 | 15 | 9 | High | 20 - 49 | 49 | 71 | NC | Yes |  |
| Lindesmith 2005 | 77 | 34 | High | 21 - 54 | 47 | 73 | NC | Yes |  |
| Atmar 2008 | 84 | 62 | High | 18 - 50 |  |  | TX | Only + |  |
| Teunis 2008 | 108 | 40 | High |  |  |  | NY | No |  |
| Atmar 2011 | 16 | 16 | Low | 18 - 50 | 59 |  | TX | Only + | Blood group O/A only |
| Leon 2011 | 44 | 16 | Low | 18 - 48 | 48 | 34 | GA | No |  |
| Seitz 2011 | 13 | 10 | Med | 18 - 50 |  |  | GA | Yes |  |
| Frenck 2012 | 40 | 17 | Low | 19 - 48 | 48 | 23 | OH | Yes | < 1600 reciprocal titer at start |
| Atmar 2014 | 49 | 31 | Low | 20 - 50 | 54 | 53 | TX | Only + | Blood group O/A only |
| Bernstein 2015 | 132 | 57 | Med | 18 - 49 | 52 | 25 | multiple | Only + |  |
| Mateo 2020 | 16 | 11 | Low | 18 - 49 | 47 | 47 | CA | Only + | Blood group O/A only |
| 1Sample size | | | | | | | | | |
| 2Number of events | | | | | | | | | |
| 3Whether subject FUT2 genotype status was measured | | | | | | | | | |

The majority of the studies (11/20) were judged to be at high risk of bias. A summary plot of the risk of bias for the included studies is shown in Figure 3. The most common issue was whether the recruited participants were from the same underlying cohort (or if the study stratified results by cohort) and whether inclusion criteria were clearly reported. Note that there are no high risk of bias studies which were conducted after 2010–many older studies lacked the same standardized reporting which is now commonplace, and so have a higher risk of bias. Furthermore, older reports often tended to report the results of multiple challenge studies simultaneously, without stratifying results by cohort. Risk of bias domain assessments for each study are shown in Figure 4.

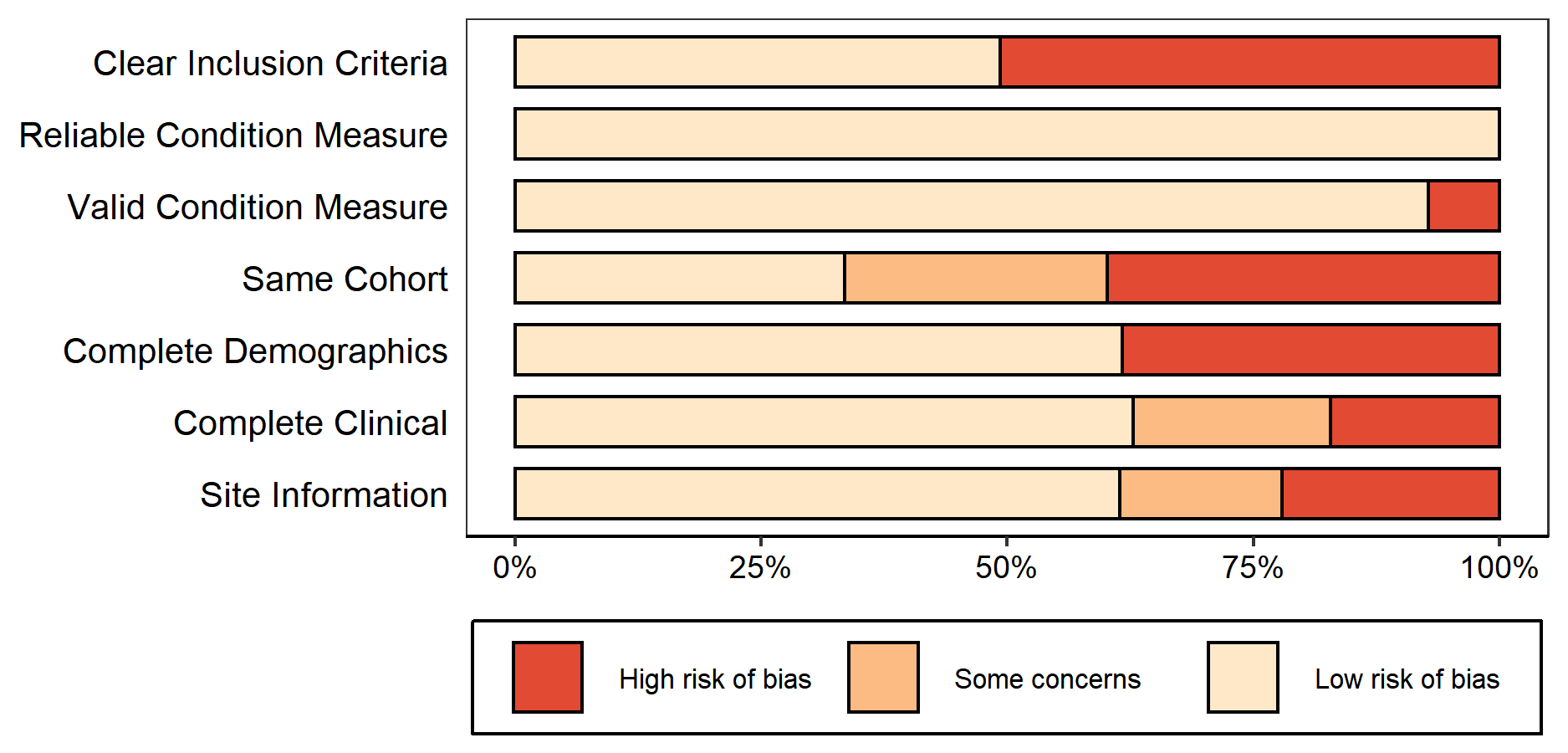


Figure 3: Summary risk of bias estimates for each domain across the entire set of included studies. Percentages are weighted by study size.

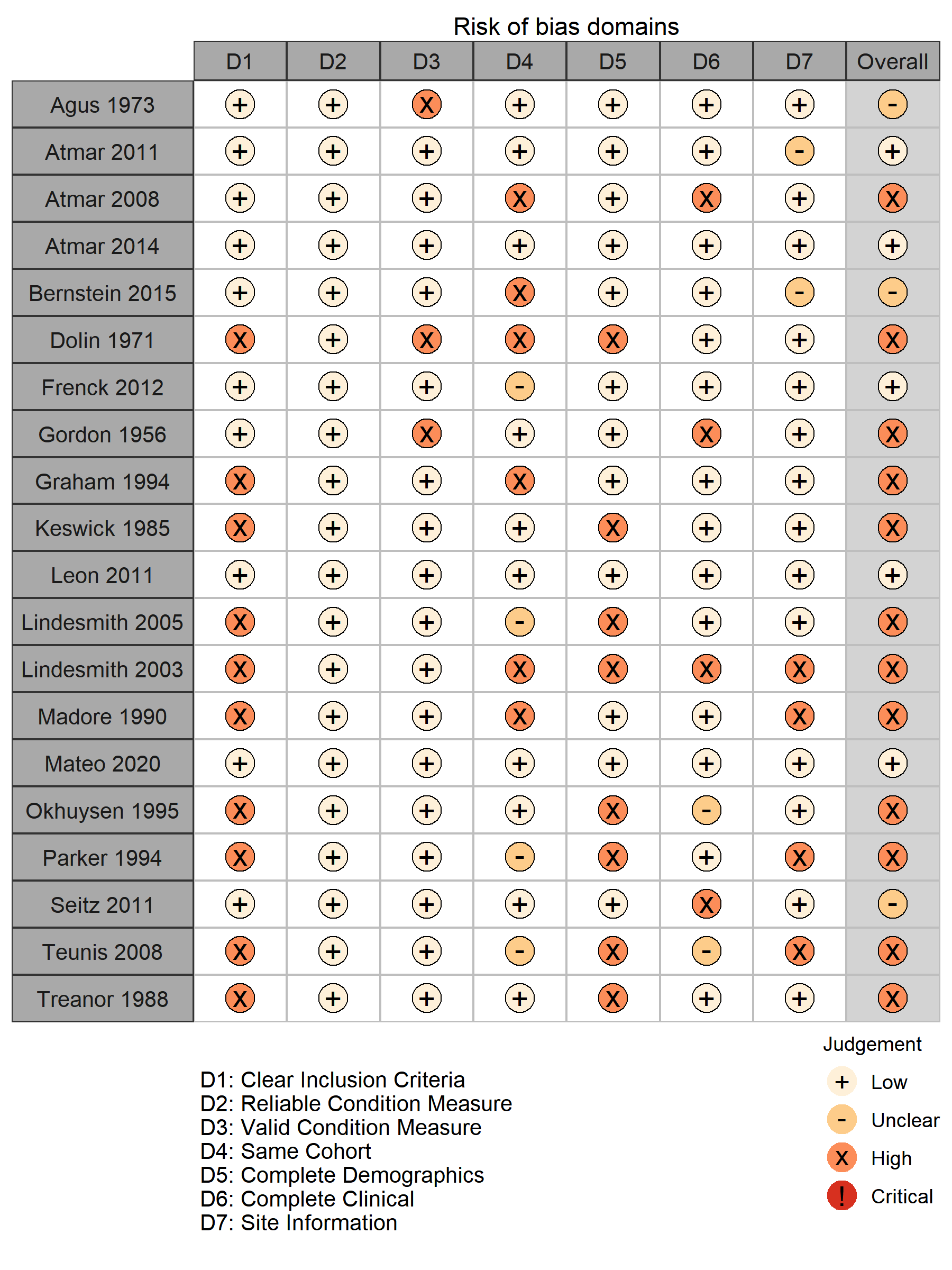


Figure 4: Individual risk of bias assessments in each domain for all included studies.

## Meta-analysis

Estimates of the estimated proportion and 95% confidence interval for each study, along with the estimated random-effects summary estimate are shown in Figure 5. Individual point estimates for each study range from 0.36 to 1.00, with 95% confidence limits ranging from 0.22 at the lowest to 1.00 at the highest. The heterogeneity among studies is extremely high, estimated at , and the 95% prediction interval indicates a low level of precision in the summary estimate. The summary estimate is unlikely to be trustworthy due to the high degree of heterogeneity present.

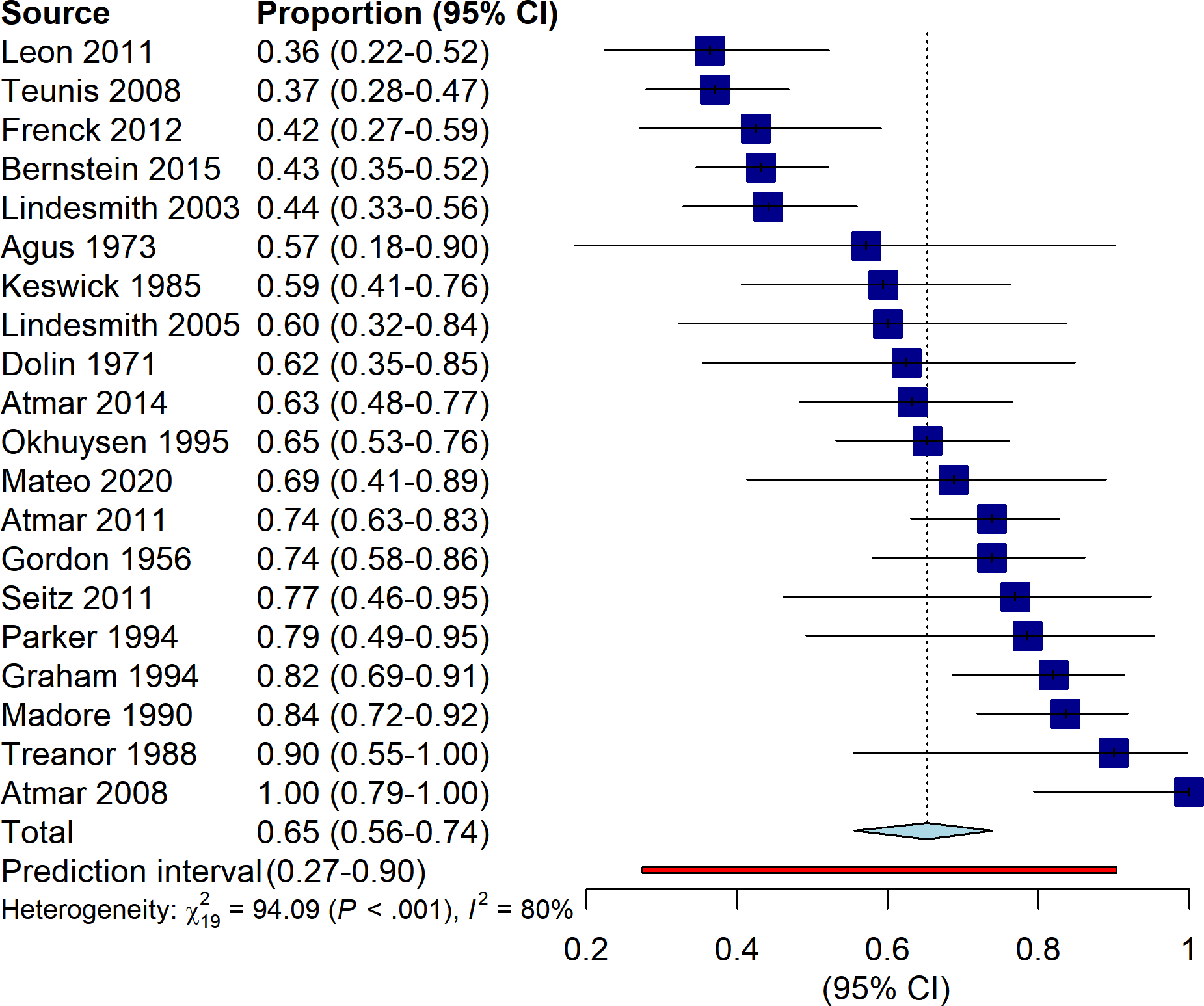


Figure 5: Forest plot showing estimates for individual studies, along with the random-effects summary estimate.

In order to examine the heterogeneity between studies, we examined several *a priori* subgroup analyses. The first subgroup analysis we considered was the effect of stratifying by FUT2 genogroup, shown in Figure 6. Stratifying by FUT2 control status minimally reduces the within-subgroup heterogeneity as well as the overall heterogeneity. For both subgroups (FUT2 not controlled and FUT2 status controlled for), the summary estimates and values are similar to the overall pooled measure, indicating little difference between subgroups. The 95% prediction interval for the overall effect is slightly smaller, as is the overall value, but controlling for FUT2 appears to explain only a slight fraction of heterogeneity in the outcome.

Next, we conducted a subgroup analysis using norovirus genogroup as the stratifying variable. The results are shown in Figure 7. Interestingly, we notice that the studies which do not report the inoculum genogroup (Genogroup unknown) are quite homogeneous, , although the summary estimate and prediction interval are wide. However, the more interesting comparison is between studies reporting a genogroup I inoculum and studies reporting a genogroup II inoculum. Among these two groups, stratifying by genogroup does not reduce heterogeneity at all compared to the overall analysis.

The final planned subgroup analysis we conducted was stratified by risk of bias (high risk of bias studies compared to low/medium risk of bias studies, which were pooled), shown in Figure 8. Again, we can see a great deal of variation within both subgroups, and stratifying by risk of bias does not appear to have a noticeable effect on heterogeneity. High risk of bias studies appear to have the same amount of spread in results as do medium and low risk of bias studies.

In order to determine if there was an effect of publication year, we resorted the forest plot for the overall meta-analysis by publication year rather than by effect size in Figure 9. Based solely on visual inspection of the forest plot, there seems to be no clear pattern in either effect size measurement or study precision over time, and so no further investigation of the potential effect of publication year was considered.

Since the planned subgroup analyses provided no insight into heterogeneity in the synthesis results, several post-hoc outlier analyses were conducted. Using a basic method for outlier detection (as described previously), five studies were detected as outliers: Leon 2011, Teunis 2008, Bernstein 2015, Lindesmith 2005, and Atmar 2008. While removing these outliers has a large effect on the heterogeneity (updated ), we proceeded with other methods of outlier identification before making any further decisions.

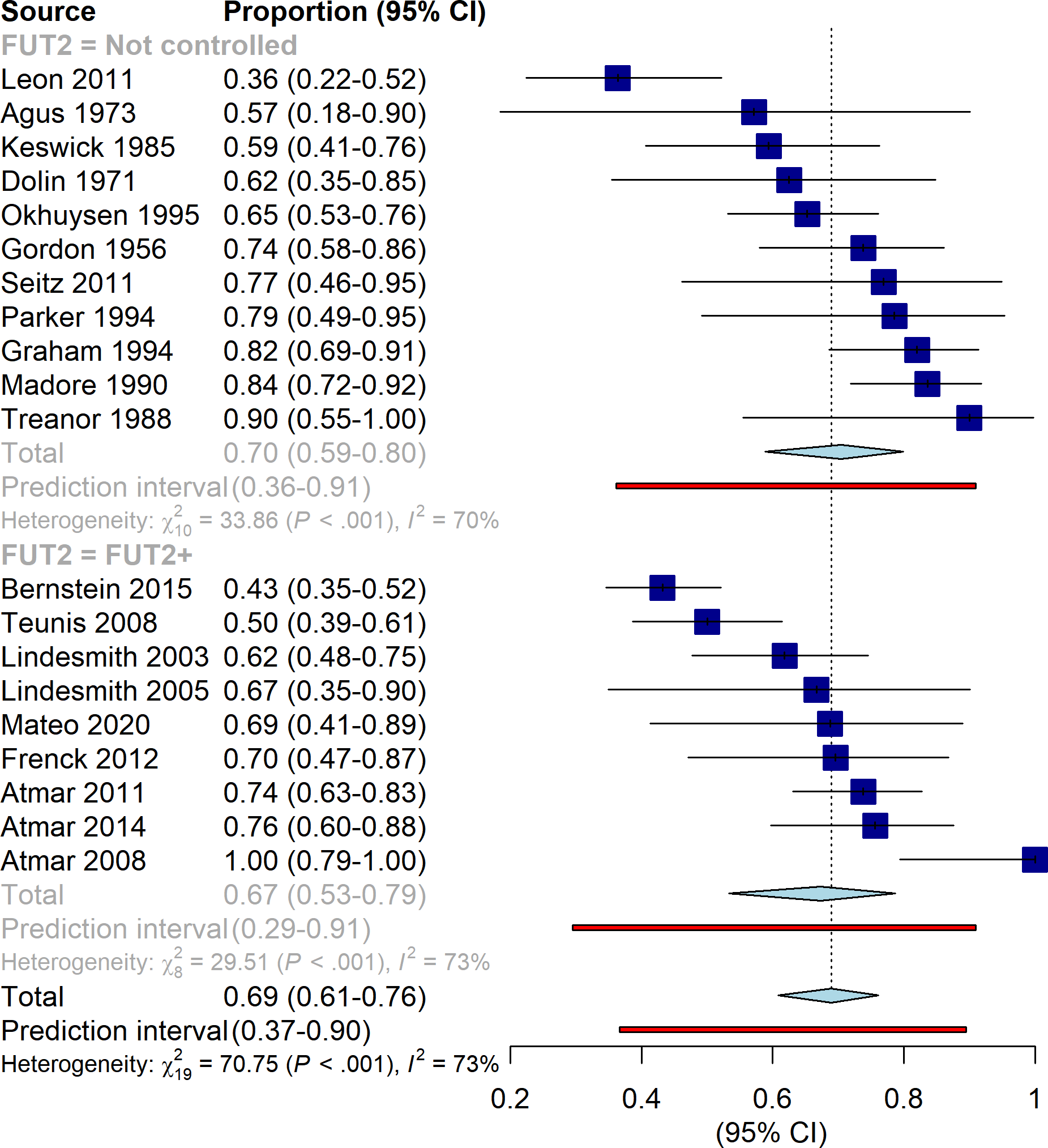


Figure 6: Forest plot showing the subgroup analysis stratified by whether studies controlled for FUT2 status or not. In this analysis, participants who were FUT2- were excluded in order to determine if including FUT2- participants in the study would bias the results. There appear to be few differences between the subgroups.

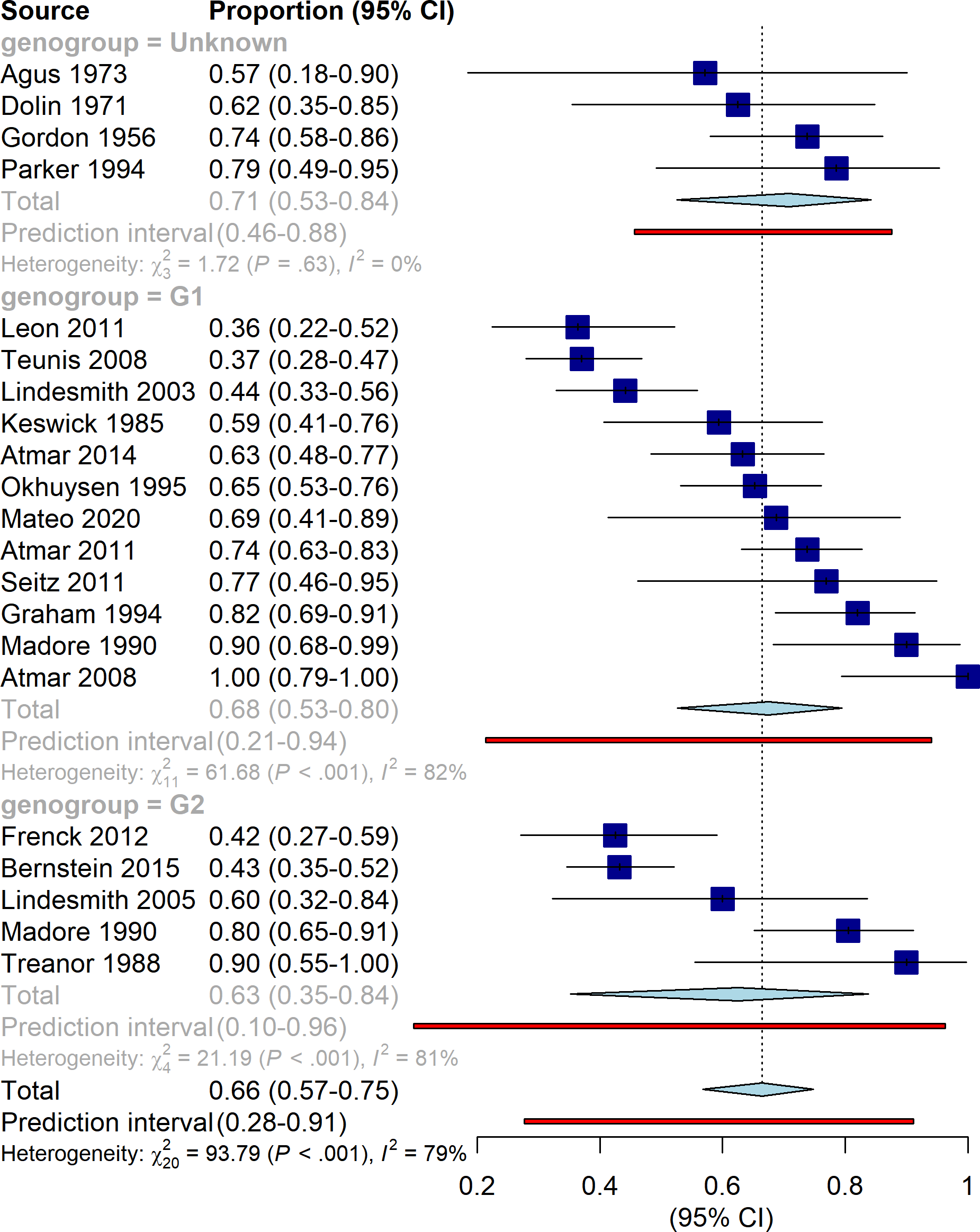


Figure 7: Forest plot showing the subgroup analysis stratified by inoculum genogroup. Interestingly the studies which failed to report a genogroup are homogeneous, but differentiating between GI and GII inocula explains little of the heterogeneity within the overall analysis.

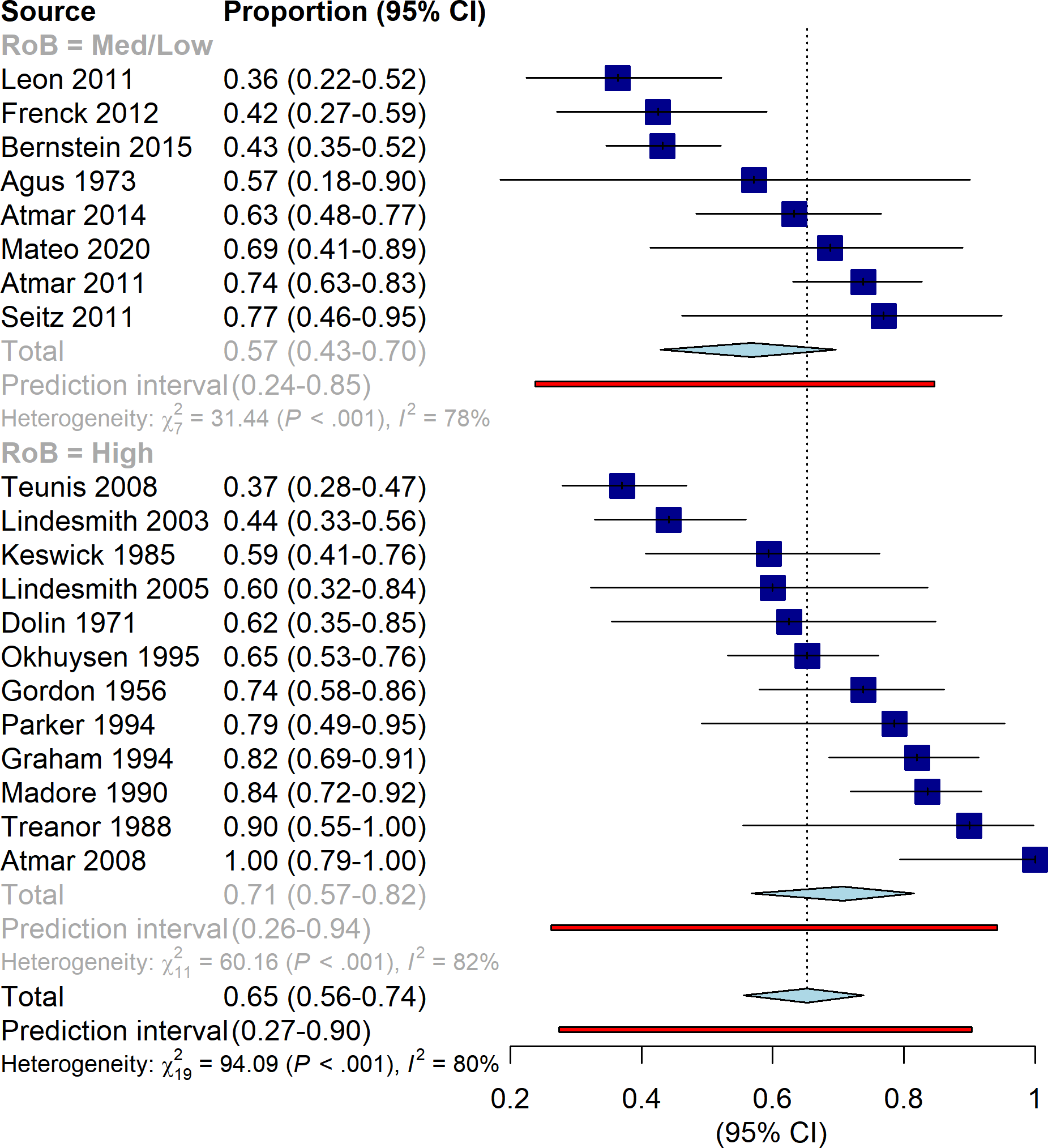


Figure 8: Forest plot showing the subgroup analysis stratified by risk of bias. There are no visible differences between the high risk of bias and low/med risk of bias groups.

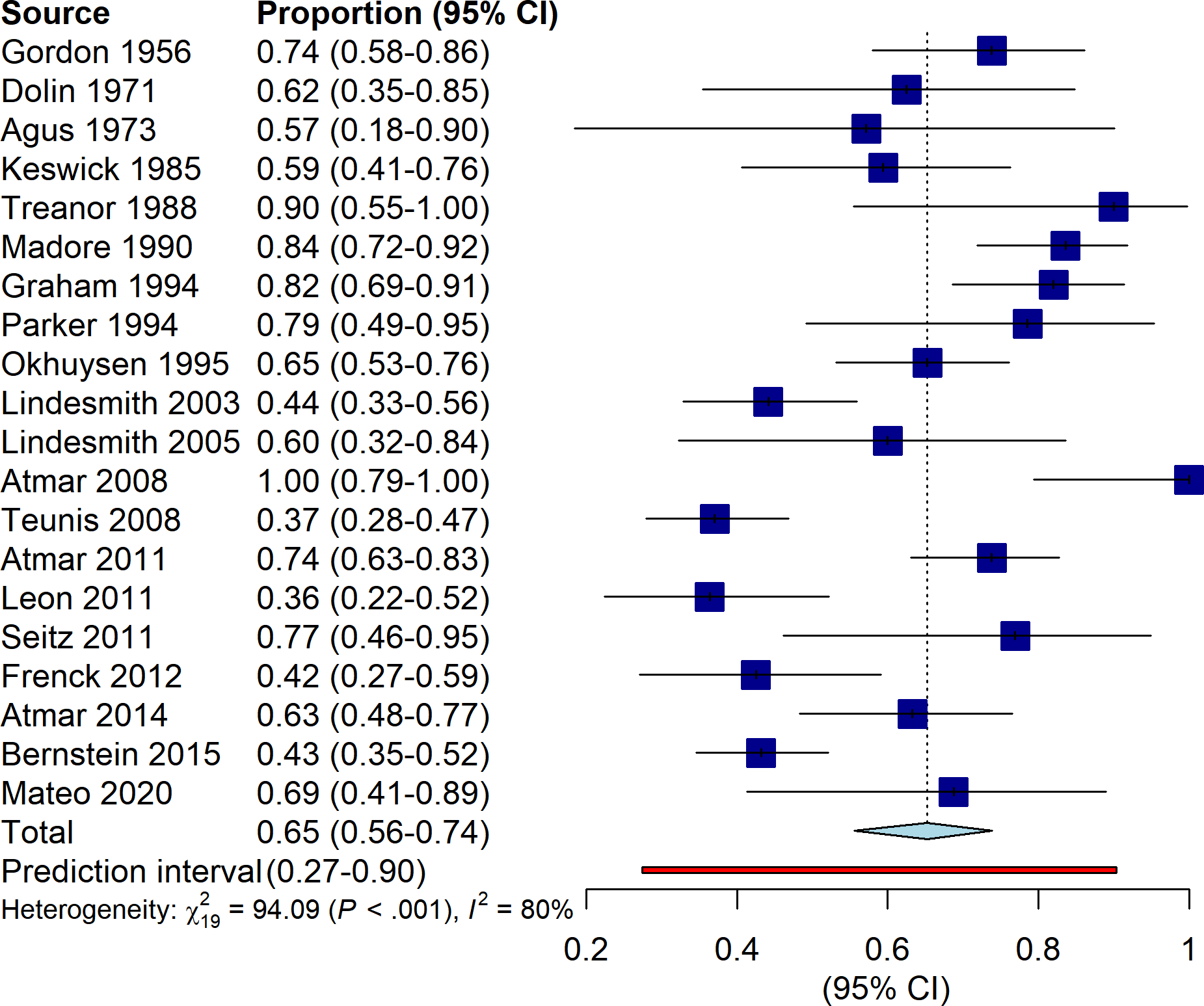


Figure 9: Forest plot of the overall analysis sorted by the year of publication. No patterns are immediately visible, and it appears that there is no trend between the estimated effect size and the year of publication.

Next, we performed a leave-one-out (LOO) influence analysis. The overall effect of removing each study is shown in Figure 10. Removing any one study appears to have little effect on the , indicating that removing multiple studies is likely necessary to obtain a trustworthy summary estimate. From the LOO results, a Baujat plot was created, shown in Figure 11. Based on the Baujat plot and LOO diagnostics (shown in the appendix), we identified three studies as outliers: Bernstein 2015, Teunis 2008, and Madore 1990. Removing these three studies slightly decreases the heterogeneity (updated ).



Figure 10: Forest plot showing the resulting effect size when studies are removed one at a time from the overall analysis. The studies are sorted by the decrease in heterogeneity associated with leaving that study out.

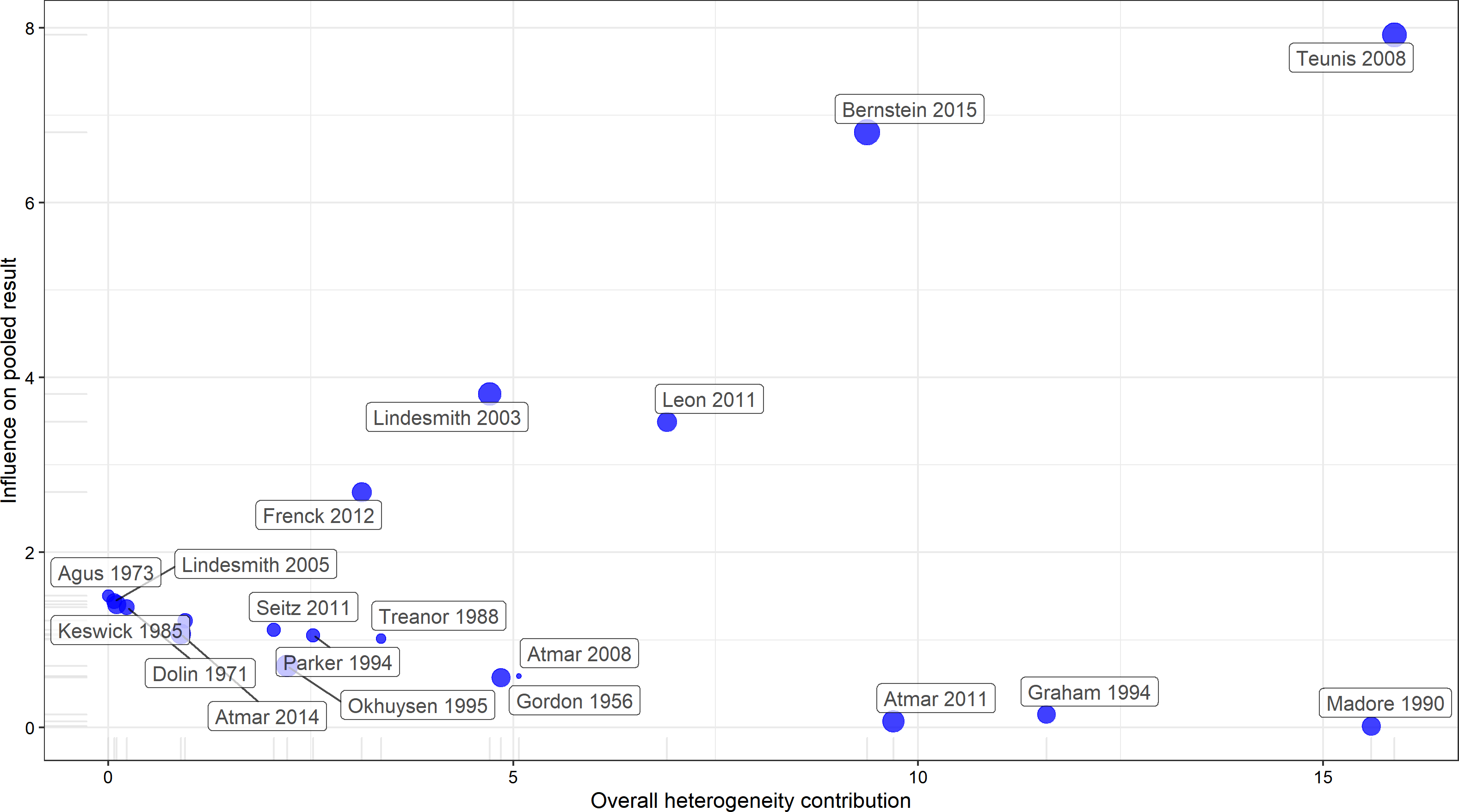


Figure 11: Baujat plot showing the change in effect size vs. the challenge in heterogeneity estimate associated with leaving out each study. Each point on the plot represents leaving out one study from the overall analysis, and the size of the point reflects that study’s sample size.

The final *post hoc* method for heterogeneity we considered was the GOSH method. The overall GOSH results are shown in Figure 12. There are a few remarkable patterns which can be observed in the GOSH plot. Notably, the majority of effect size estimates (log-odds, on the -axis) fit a roughly normal distribution around a log-odds estimate of (corresponding to an estimated proportion infected of ), while the heterogeneity estimates of the random subset models have a skewed distribution. Most of the random subsets have a high estimate, around . This indicates that attempting to deal with heterogeneity may be a fruitless pursuit, as the majority of the random subsets are quite heterogeneous. However, we can also see a number of random subset models with , spread over a wide range of effect sizes. These subsets with are likely not meaningful or easy to find, however.

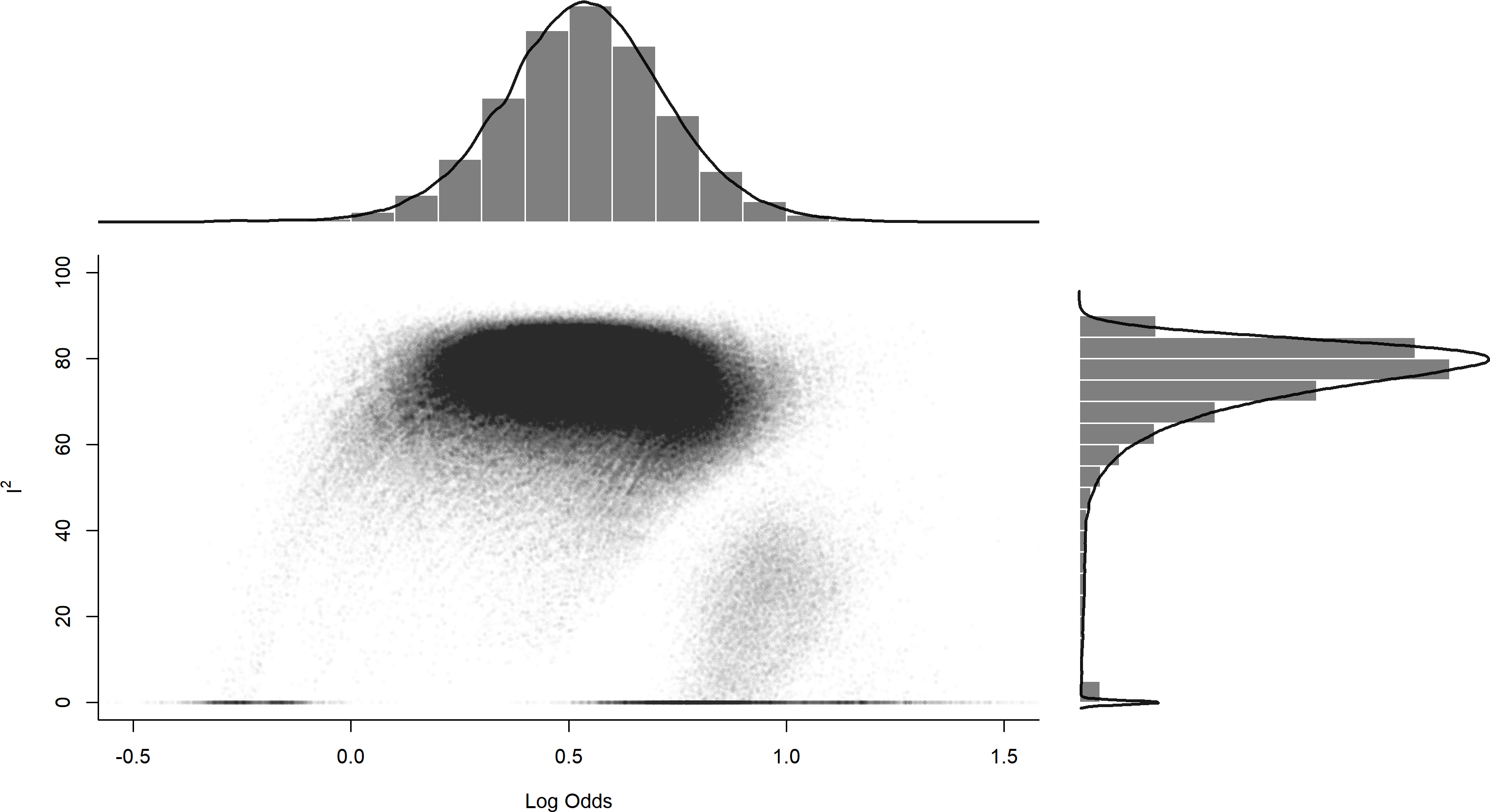


Figure 12: GOSH plot: each point on the plot represents one random subset analysis, where the inverse-variance meta analysis model was fit to a random subset of studies. The y-axis is the estimated heterogeneity, and the x-axis is the estimated effect size. The marginal histograms and density curves show the overall estimated distribution of the results.

We also used a *post hoc* clustering approach on the results of the GOSH analysis to identify outliers. Three clustering methods were applied, and the raw clustering results are shown in the appendix. The clustering methods identified four studies are potential outliers: Bernstein 2015, Leon 2011, Madore 1990, and Teunis 2008. Note that these detected outliers are the same as the LOO method, with the addition of the Leon study. Removing all four of these studies simultaneously lowers the estimate to .

Using a consensus of the three outlier detection methods (a study is declared an outlier if any two of the three detection methods identified it), the studies we identify as outliers are: Bernstein 2015, Leon 2011, Madore 1990, and Teunis 2008. The forest plot with these four studies excluded is shown in Figure 13. The with the outliers removed is , and the 95% prediction interval remains quite wide. Notably, the consensus method failed to reduce heterogeneity any better than the simple method did, and the four studies identified as outliers have no superficial similarities which were not already examined (see Table 2).

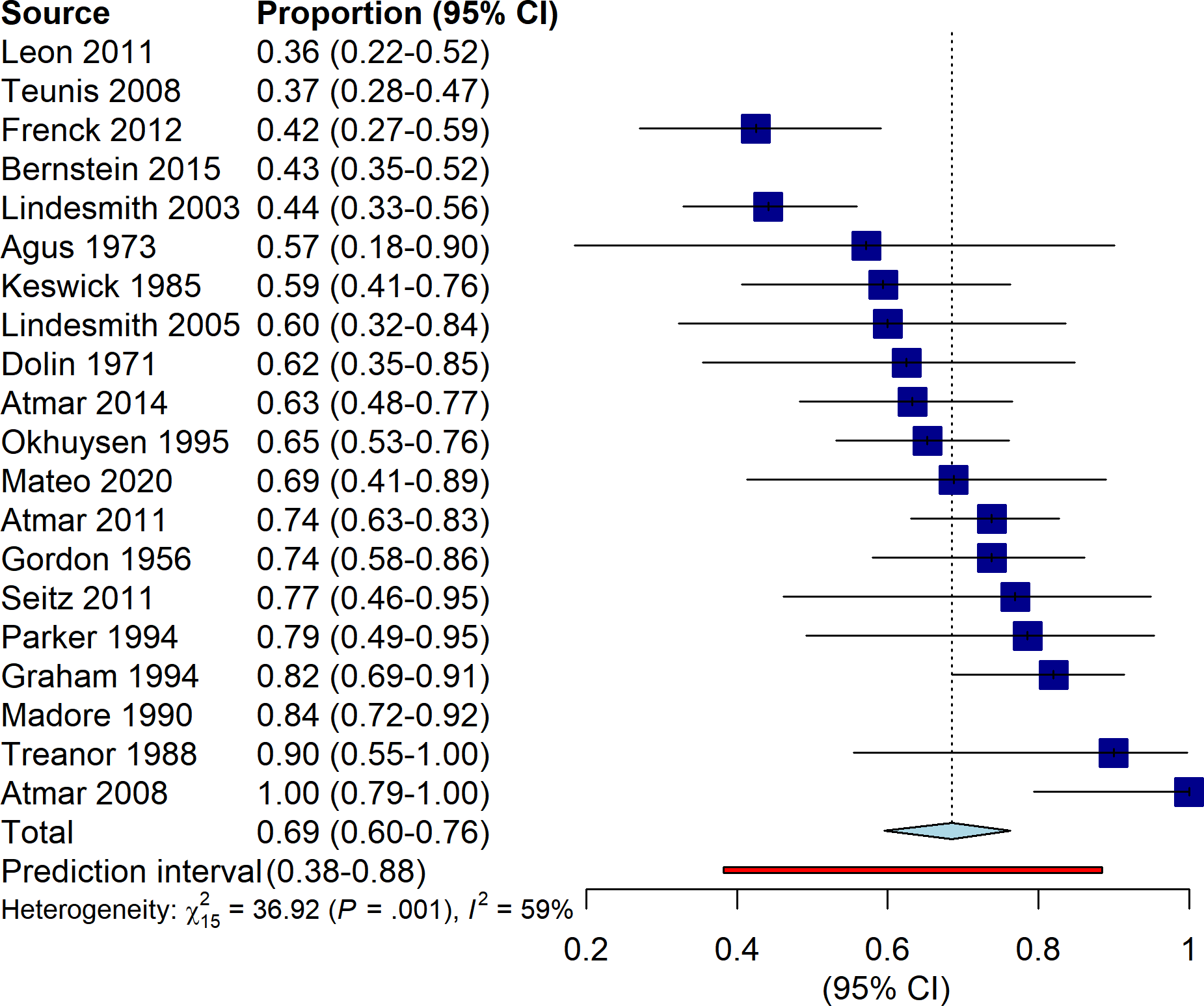


Figure 13: Forest plot showing the individual study estimates and overall random effects summary estimate when the four outliers detected using the consensus method are removed.

Table 3 shows summary statistics of the different methods we used to examine heterogeneity.

Table 3: Summary estimates and heterogeneity for all synthesis analyses conducted.

| **analysis** | **prop1** | **CI2** | **PI3** | **I^24** |
| --- | --- | --- | --- | --- |
| Overall | 0.65 | (0.56, 0.74) | (0.27, 0.90) | 80% (70%, 87%) |
| Simple outliers | 0.72 | (0.64, 0.78) | (0.45, 0.89) | 49% (09%, 71%) |
| Consensus outliers | 0.69 | (0.60, 0.76) | (0.38, 0.88) | 59% (39% - 77%) |
| FUT2 | 0.69 | (0.61, 0.76) | (0.37, 0.90) | 73% (58%, 83%) |
| Genogroup | 0.66 | (0.57, 0.75) | (0.28, 0.91) | 79% (68%, 86%) |
| Risk of bias | 0.65 | (0.56, 0.74) | (0.27, 0.90) | 80% (70%, 87%) |
| 1Estimated proportion from random-effects model | | | | |
| 295% confidence interval for estimate | | | | |
| 395% prediction interval for estimate | | | | |
| 4I^2 estimate with 95% confidence interval | | | | |

Finally, we assessed publication bias to determine if small-study effects or missing literature could explain the lack of coherent synthesis results. The results are shown visually in Figure 14. Some patterns are noticeable by visual inspection: notably, there is a cluster of 5 studies in the upper left with relatively high sample sizes which appear to estimate lower outcomes than the other studies. There also appear to be more studies which estimate high proportions than there are studies estimating low proportions as outcomes, even though many of these studies are not significant.

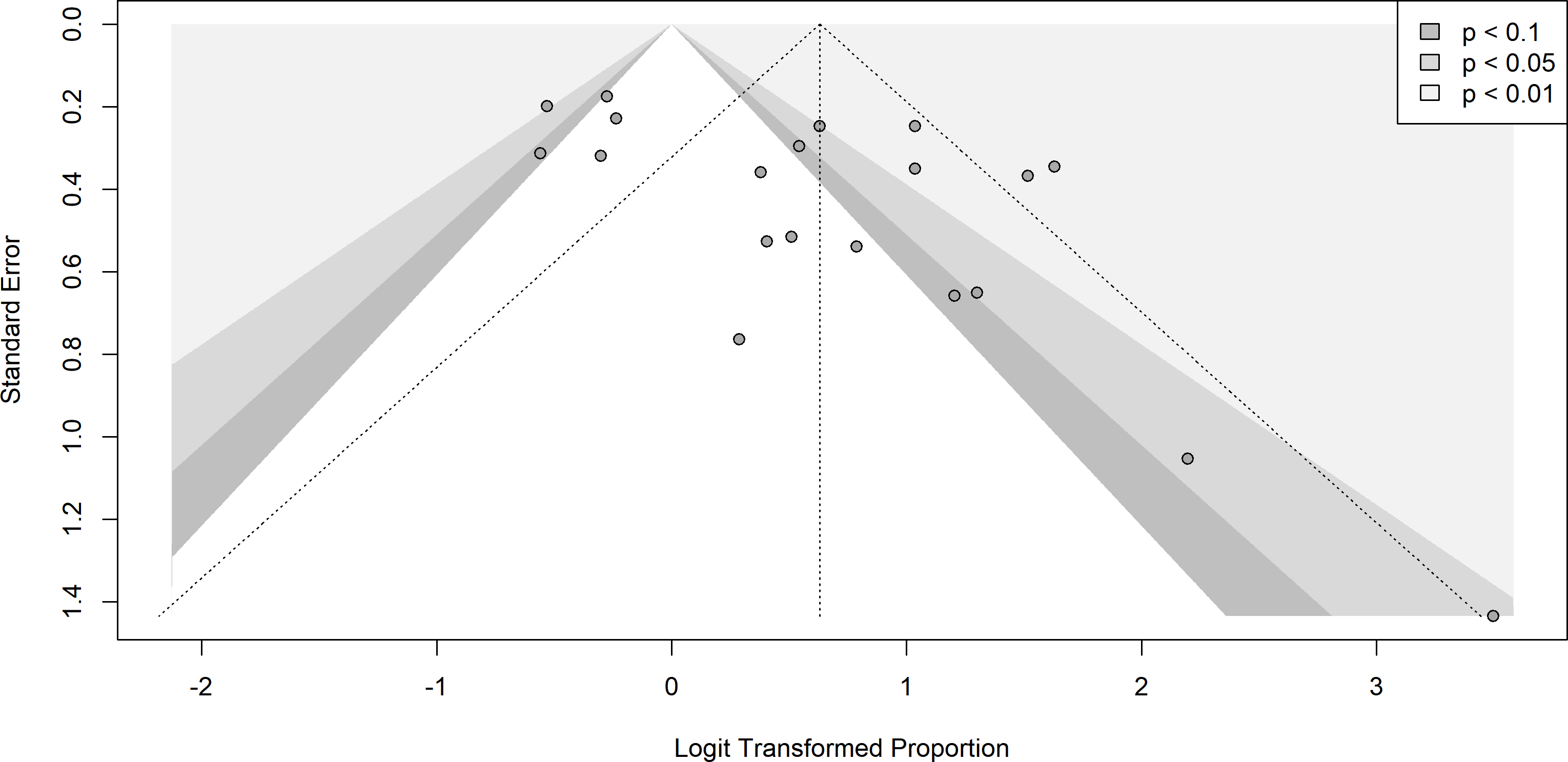


Figure 14: Contour-enhanced funnel plot. Each point represents an included study. Contour shading indicates that the study result is significant at the shaded alpha level. The study in the bottom left is Atmar 2008, which estimated a proportion of 1.

We also assessed the asymmetry of the funnel plot using Peters’ test (note that Peters’ test for asymmetry is based on the slope of the regression line, rather than the intercept). The estimated slope was with an approximate 95% confidence interval of . The test was not significant at a 95% level of confidence (). The test is also shown visually in Figure 15.

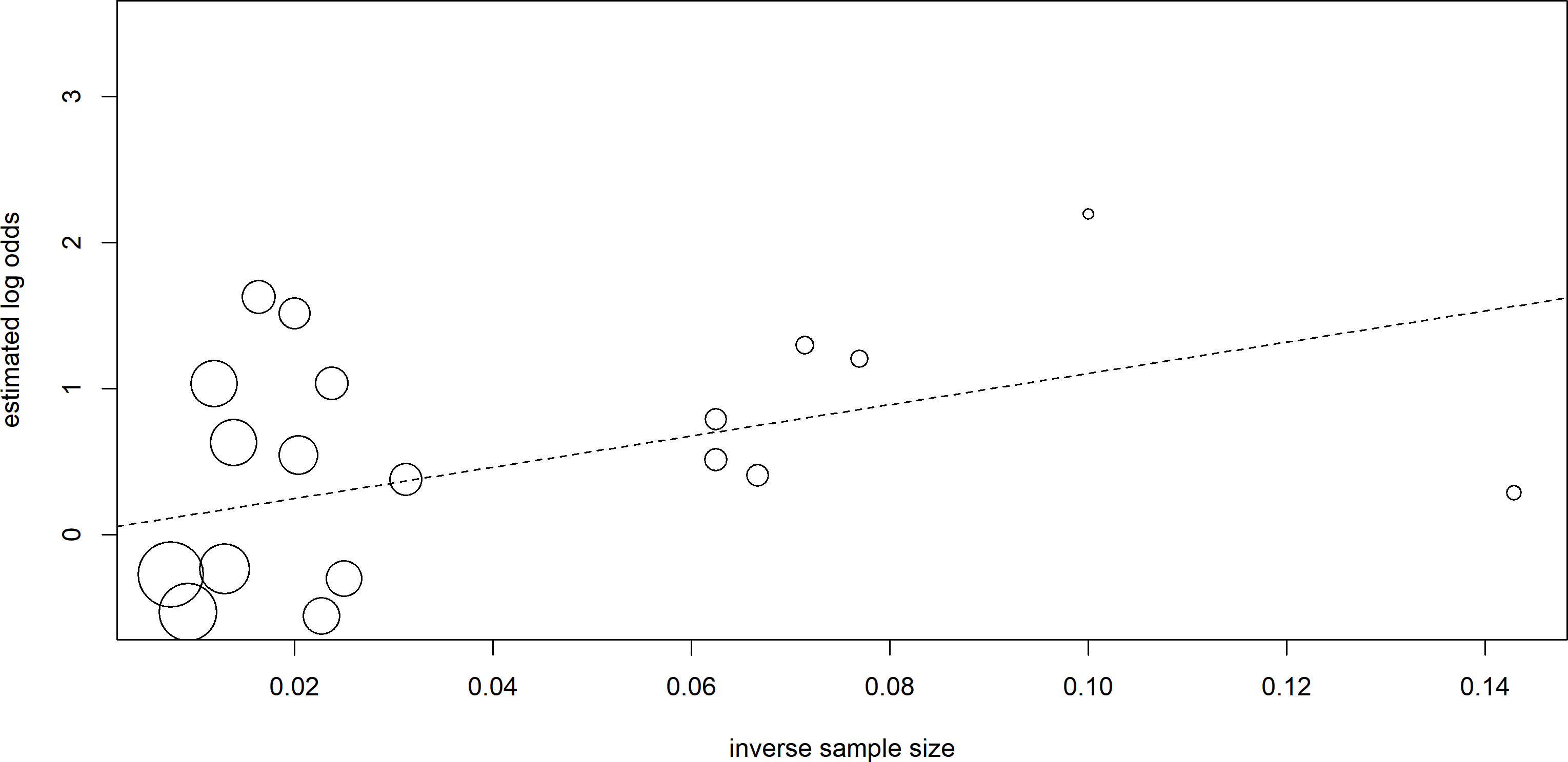


Figure 15: Visual representation of the Peters’ test. The test fits a linear regression model with the estimated log-odds as the outcome and inverse sample size as the predictor, using inverse variance weighting, depicted as the size of the point. The regression line appears to have an intercept close to zero.

From the test, we cannot conclude that there is definite evidence of asymmetry. However, visual inspection of the data does suggest interesting patterns which may be unrelated to problems that cause asymmetry.

# Discussion

We identified 20 unique studies which contributed unique norovirus challenge data to the literature among the 65 studies we included in our review. Synthesizing data from these studies, we found significant heterogeneity between the proportion of infected individuals in norovirus challenges studies, which could not be explained by study risk of bias, inoculum genogroup, or exclusion of FUT2- participants. *Post hoc* automated analyses also detected a subset of outlying studies which had little in common. The funnel plot also provided visual evidence of heterogeneity in proportion estimates across studies. Taken together, our results indicate a degree of heterogeneity in infectivity across challenge studies, contrary to what we hypothesized.

There is room for bias in our systematic review results, however. Primarily, we may not have captured all published norovirus challenge studies. Searching Google Scholar or medRχiv to identify gray literature and examining references of included studies may reveal additional sources of data that are missing from our review. Due to the breadth of records captured and identified from the two databases, the included studies seem likely to be representative of all norovirus human challenge studies which generated original infection data, but our data may not necessarily be reflective of unpublished studies (though there was little statistical evidence of funnel plot asymmetry in our results). Furthermore, our study risk of bias assessment tool is a prototype, as the JBI case series quality assessment tool is not a perfect fit for the typical challenge study design, and to the authors’ knowledge, no such tool has been verified by challenge study experts. That being said, several of our studies are old and there are numerous issues with the reporting of case series in general, so finding several studies at high risk of bias may not be surprising. Incomplete study capture or errors in using the risk of bias assessment tool may also contribute to the unexplainable heterogeneity in our results.

However, between-study heterogeneity may also be explained by unmeasured (and potentially unmeasurable) differences between challenge studies. Challenge studies are likely to capture different cohorts, and minor variations in study administration and protocols could potentially lead to variation in the proportion of challenged individuals who were affected. Potential differences with the inocula used, such as length of cold storage, original source and number of times passaged, genotypic differences despite the reported genogroup, and inoculum dose could all play a large role in norovirus infectivity. Given the high probability of norovirus infection when even a small number of viral particles are administered,9,16 inoculum dose and genogroup are not expected to play a large role in challenge study infectivity, but surprinsing effects or slight variation in protocols could lead to observed heterogeneity across studies.

Overall, we find that norovirus infectivity appears to be heterogeneous across challenge studies. We conducted a comprehensive systematic review and meta analysis, and so feel fairly confident that this result is representative of norovirus challenge studies in general. We expected norovirus challenge studies to have less variability in infectivity, but the observed variation could be due to differences in protocols and study administration, differences in inoculum characteristics, or differences in sampled cohorts which reflect the variability of norovirus attack rate in natural outbreaks due to heterogeneous host populations.

# References

1. Robilotti E, Deresinski S, Pinsky BA. Norovirus. *Clinical Microbiology Reviews*. 2015;28(1):134-164. doi:[10.1128/CMR.00075-14](https://doi.org/10.1128/CMR.00075-14)

2. Ludwig-Begall LF, Mauroy A, Thiry E. Noroviruses—The State of the Art, Nearly Fifty Years after Their Initial Discovery. *Viruses*. 2021;13(8, 8):1541. doi:[10.3390/v13081541](https://doi.org/10.3390/v13081541)

3. Centers for Disease Control. Burden of Norovirus Illness in the U.S. | CDC. Published April 5, 2021. Accessed December 10, 2021. <https://www.cdc.gov/norovirus/trends-outbreaks/burden-US.html>

4. Siebenga JJ, Beersma MFC, Vennema H, van Biezen P, Hartwig NJ, Koopmans M. High prevalence of prolonged norovirus shedding and illness among hospitalized patients: A model for in vivo molecular evolution. *J Infect Dis*. 2008;198(7):994-1001. doi:[10.1086/591627](https://doi.org/10.1086/591627)

5. Davis A, Cortez V, Grodzki M, et al. Infectious Norovirus Is Chronically Shed by Immunocompromised Pediatric Hosts. *Viruses*. 2020;12(6, 6):619. doi:[10.3390/v12060619](https://doi.org/10.3390/v12060619)

6. Atmar RL, Opekun AR, Gilger MA, et al. Norwalk virus shedding after experimental human infection. *Emerg Infect Dis*. 2008;14(10):1553-1557. doi:[10.3201/eid1410.080117](https://doi.org/10.3201/eid1410.080117)

7. Costantini VP, Cooper EM, Hardaker HL, et al. Epidemiologic, Virologic, and Host Genetic Factors of Norovirus Outbreaks in Long-term Care Facilities. *Clinical Infectious Diseases*. 2016;62(1):1-10. doi:[10.1093/cid/civ747](https://doi.org/10.1093/cid/civ747)

8. Patel MM, Hall AJ, Vinjé J, Parashar UD. Noroviruses: A comprehensive review. *Journal of Clinical Virology*. 2009;44(1):1-8. doi:[10.1016/j.jcv.2008.10.009](https://doi.org/10.1016/j.jcv.2008.10.009)

9. Teunis PF, Moe CL, Liu P, et al. Norwalk virus: How infectious is it? *J Med Virol*. 2008;80(8):1468-1476. doi:[10.1002/jmv.21237](https://doi.org/10.1002/jmv.21237)

10. GAYTHORPE KAM, TROTTER CL, LOPMAN B, STEELE M, CONLAN AJK. Norovirus transmission dynamics: A modelling review. *Epidemiol Infect*. 2018;146(2):147-158. doi:[10.1017/S0950268817002692](https://doi.org/10.1017/S0950268817002692)

11. de Graaf M, van Beek J, Koopmans MPG. Human norovirus transmission and evolution in a changing world. *Nat Rev Microbiol*. 2016;14(7, 7):421-433. doi:[10.1038/nrmicro.2016.48](https://doi.org/10.1038/nrmicro.2016.48)

12. Donaldson EF, Lindesmith LC, LoBue AD, Baric RS. Viral shape-shifting: Norovirus evasion of the human immune system. *Nat Rev Microbiol*. 2010;8(3, 3):231-241. doi:[10.1038/nrmicro2296](https://doi.org/10.1038/nrmicro2296)

13. Lindesmith L, Moe C, Marionneau S, et al. Human susceptibility and resistance to Norwalk virus infection. *Nat Med*. 2003;9(5):548-553. doi:[10.1038/nm860](https://doi.org/10.1038/nm860)

14. Marionneau S, Ruvoën N, Le Moullac–Vaidye B, et al. Norwalk virus binds to histo-blood group antigens present on gastroduodenal epithelial cells of secretor individuals. *Gastroenterology*. 2002;122(7):1967-1977. doi:[10.1053/gast.2002.33661](https://doi.org/10.1053/gast.2002.33661)

15. Lindesmith LC, Brewer-Jensen PD, Mallory ML, et al. Virus–Host Interactions Between Nonsecretors and Human Norovirus. *Cellular and Molecular Gastroenterology and Hepatology*. 2020;10(2):245-267. doi:[10.1016/j.jcmgh.2020.03.006](https://doi.org/10.1016/j.jcmgh.2020.03.006)

16. Teunis PFM, Le Guyader FS, Liu P, Ollivier J, Moe CL. Noroviruses are highly infectious but there is strong variation in host susceptibility and virus pathogenicity. *Epidemics*. 2020;32:100401. doi:[10.1016/j.epidem.2020.100401](https://doi.org/10.1016/j.epidem.2020.100401)

17. Atmar RL, Opekun AR, Gilger MA, et al. Determination of the 50% human infectious dose for Norwalk virus. *J Infect Dis*. 2014;209(7):1016-1022. doi:[10.1093/infdis/jit620](https://doi.org/10.1093/infdis/jit620)

18. Munn Z, Barker TH, Moola S, et al. Methodological quality of case series studies: An introduction to the JBI critical appraisal tool. *JBI Evidence Synthesis*. 2020;18(10):2127-2133. doi:[10.11124/JBISRIR-D-19-00099](https://doi.org/10.11124/JBISRIR-D-19-00099)

19. Stijnen T, Hamza TH, Özdemir P. Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. *Statistics in Medicine*. 2010;29(29):3046-3067. doi:[10.1002/sim.4040](https://doi.org/10.1002/sim.4040)

20. Harrer M, Cuijpers P, Furukawa TA, Ebert DD. *Doing Meta-Analysis in R: A Hands-On Guide*. Chapmann & Hall/CRC Press; 2021. Accessed December 2, 2021. <https://bookdown.org/MathiasHarrer/Doing_Meta_Analysis_in_R/>

21. Knapp G, Hartung J. Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine*. 2003;22(17):2693-2710. doi:[10.1002/sim.1482](https://doi.org/10.1002/sim.1482)

22. Sidik K, Jonkman JN. A simple confidence interval for meta-analysis. *Statistics in Medicine*. 2002;21(21):3153-3159. doi:[10.1002/sim.1262](https://doi.org/10.1002/sim.1262)

23. IntHout J, Ioannidis JP, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC Medical Research Methodology*. 2014;14(1):25. doi:[10.1186/1471-2288-14-25](https://doi.org/10.1186/1471-2288-14-25)

24. Langan D, Higgins JPT, Jackson D, et al. A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. *Research Synthesis Methods*. 2019;10(1):83-98. doi:[10.1002/jrsm.1316](https://doi.org/10.1002/jrsm.1316)

25. Schwarzer G, Chemaitelly H, Abu-Raddad LJ, Rücker G. Seriously misleading results using inverse of Freeman-Tukey double arcsine transformation in meta-analysis of single proportions. *Res Synth Methods*. 2019;10(3):476-483. doi:[10.1002/jrsm.1348](https://doi.org/10.1002/jrsm.1348)

26. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. *Introduction to Meta-Analysis*. John Wiley & Sons; 2009.

27. Borenstein M, Higgins JPT. Meta-Analysis and Subgroups. *Prev Sci*. 2013;14(2):134-143. doi:[10.1007/s11121-013-0377-7](https://doi.org/10.1007/s11121-013-0377-7)

28. Baujat B, Mahé C, Pignon JP, Hill C. A graphical method for exploring heterogeneity in meta-analyses: Application to a meta-analysis of 65 trials. *Statistics in Medicine*. 2002;21(18):2641-2652. doi:[10.1002/sim.1221](https://doi.org/10.1002/sim.1221)

29. Viechtbauer W, Cheung MW-L. Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*. 2010;1(2):112-125. doi:[10.1002/jrsm.11](https://doi.org/10.1002/jrsm.11)

30. Olkin I, Dahabreh IJ, Trikalinos TA. GOSH – a graphical display of study heterogeneity. *Research Synthesis Methods*. 2012;3(3):214-223. doi:[10.1002/jrsm.1053](https://doi.org/10.1002/jrsm.1053)

31. Hartigan JA, Wong MA. Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society Series C (Applied Statistics)*. 1979;28(1):100-108. doi:[10.2307/2346830](https://doi.org/10.2307/2346830)

32. Schubert E, Sander J, Ester M, Kriegel HP, Xu X. DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN. *ACM Trans Database Syst*. 2017;42(3):19:1-19:21. doi:[10.1145/3068335](https://doi.org/10.1145/3068335)

33. Fraley C, Raftery AE. Model-Based Clustering, Discriminant Analysis, and Density Estimation. *Journal of the American Statistical Association*. 2002;97(458):611-631. doi:[10.1198/016214502760047131](https://doi.org/10.1198/016214502760047131)

34. Sterne JAC, Sutton AJ, Ioannidis JPA, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*. 2011;343:d4002. doi:[10.1136/bmj.d4002](https://doi.org/10.1136/bmj.d4002)

35. Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of Clinical Epidemiology*. 2008;61(10):991-996. doi:[10.1016/j.jclinepi.2007.11.010](https://doi.org/10.1016/j.jclinepi.2007.11.010)

36. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ*. 1997;315(7109):629-634. doi:[10.1136/bmj.315.7109.629](https://doi.org/10.1136/bmj.315.7109.629)

37. The EndNote Team. *EndNote*. Clarivate; 2013.

38. Center for History and New Media. *Zotero*. George Mason University; 2021.

39. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2021. <https://www.R-project.org/>

40. Schwarzer G. *Meta: General Package for Meta-Analysis*.; 2021. [https://github.com/guido-s/meta/
https://www.springer.com/gp/book/9783319214153](https://github.com/guido-s/meta/ https://www.springer.com/gp/book/9783319214153)

41. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: A practical tutorial. *Evidence-Based Mental Health*. 2019;(22):153-160.

42. Viechtbauer W. *Metafor: Meta-Analysis Package for r*.; 2021. <https://CRAN.R-project.org/package=metafor>

43. Viechtbauer W. Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*. 2010;36(3):1-48. <https://doi.org/10.18637/jss.v036.i03>

44. Cuijpers P, Furukawa T, Ebert DD. *Dmetar: Companion r Package for the Guide Doing Meta-Analysis in r*.; 2021. <https://dmetar.protectlab.org>

45. McGuinness L. *Robvis: Visualize the Results of Risk-of-Bias (ROB) Assessments*.; 2019. <https://github.com/mcguinlu/robvis>

46. Haddaway N, McGuinness L, Pritchard C. *Prisma2020: Make Interactive PRISMA Flow Diagrams*.; 2021. <https://CRAN.R-project.org/package=PRISMA2020>

47. Haddaway NR, Pritchard CC, McGuinness LA. *Prisma2020: R Package and ShinyApp for Producing PRISMA 2020 Compliant Flow Diagrams (Version 0.0.2)*.; 2021. doi:[10.5281/zenodo.5082518](https://doi.org/10.5281/zenodo.5082518)

48. Haddaway NR, Page MJ, Pritchard CC, McGuinness LA. PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. Published online July 15, 2021:2021.07.14.21260492. doi:[10.1101/2021.07.14.21260492](https://doi.org/10.1101/2021.07.14.21260492)

49. Gohel D. *Flextable: Functions for Tabular Reporting*.; 2021. <https://CRAN.R-project.org/package=flextable>

50. Allaire J, Xie Y, McPherson J, et al. *Rmarkdown: Dynamic Documents for r*.; 2021. <https://CRAN.R-project.org/package=rmarkdown>

51. Xie Y, Allaire JJ, Grolemund G. *R Markdown: The Definitive Guide*. Chapman; Hall/CRC; 2018. <https://bookdown.org/yihui/rmarkdown>

52. Xie Y, Dervieux C, Riederer E. *R Markdown Cookbook*. Chapman; Hall/CRC; 2020. <https://bookdown.org/yihui/rmarkdown-cookbook>

53. Xie Y. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*.; 2021. <https://yihui.org/knitr/>

54. Xie Y. Knitr: A comprehensive tool for reproducible research in R. In: Stodden V, Leisch F, Peng RD, eds. *Implementing Reproducible Computational Research*. Chapman; Hall/CRC; 2014. <http://www.crcpress.com/product/isbn/9781466561595>

55. Xie Y. *Dynamic Documents with R and Knitr*. 2nd ed. Chapman; Hall/CRC; 2015. <https://yihui.org/knitr/>

56. Xie Y. *Bookdown: Authoring Books and Technical Documents with r Markdown*.; 2021. <https://CRAN.R-project.org/package=bookdown>

57. Xie Y. *Bookdown: Authoring Books and Technical Documents with R Markdown*. Chapman; Hall/CRC; 2016. <https://bookdown.org/yihui/bookdown>

58. Müller K. *Here: A Simpler Way to Find Your Files*.; 2020. <https://CRAN.R-project.org/package=here>

59. Firke S. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*.; 2021. <https://github.com/sfirke/janitor>

60. Ooms J. *Magick: Advanced Graphics and Image-Processing in r*.; 2021. <https://CRAN.R-project.org/package=magick>

61. Ooms J. *Rsvg: Render SVG Images into PDF, PNG, PostScript, or Bitmap Arrays*.; 2021. <https://github.com/jeroen/rsvg#readme>

62. Iannone R. *DiagrammeRsvg: Export DiagrammeR Graphviz Graphs as SVG*.; 2016. <https://github.com/rich-iannone/DiagrammeRsvg>

63. Wickham H, François R, Henry L, Müller K. *Dplyr: A Grammar of Data Manipulation*.; 2021. <https://CRAN.R-project.org/package=dplyr>

64. Agus SG, Dolin R, Wyatt RG, Tousimis AJ, Northrup RS. Acute infectious nonbacterial gastroenteritis: Intestinal histopathology. Histologic and enzymatic alterations during illness produced by the Norwalk agent in man. *Ann Intern Med*. 1973;79(1):18-25. doi:[10.7326/0003-4819-79-1-18](https://doi.org/10.7326/0003-4819-79-1-18)

65. Atmar RL, Bernstein DI, Harro CD, et al. Norovirus vaccine against experimental human Norwalk Virus illness. *N Engl J Med*. 2011;365(23):2178-2187. doi:[10.1056/NEJMoa1101245](https://doi.org/10.1056/NEJMoa1101245)

66. Bernstein DI, Atmar RL, Lyon GM, et al. Norovirus vaccine against experimental human GII.4 virus illness: A challenge study in healthy adults. *J Infect Dis*. 2015;211(6):870-878. doi:[10.1093/infdis/jiu497](https://doi.org/10.1093/infdis/jiu497)

67. Dolin R, Blacklow NR, DuPont H, et al. Transmission of acute infectious nonbacterial gastroenteritis to volunteers by oral administration of stool filtrates. *J Infect Dis*. 1971;123(3):307-312. doi:[10.1093/infdis/123.3.307](https://doi.org/10.1093/infdis/123.3.307)

68. Frenck R, Bernstein DI, Xia M, et al. Predicting susceptibility to norovirus GII.4 by use of a challenge model involving humans. *J Infect Dis*. 2012;206(9):1386-1393. doi:[10.1093/infdis/jis514](https://doi.org/10.1093/infdis/jis514)

69. Gordon IG, Patterson PR, Whitney E. Immunity in volunteers recovered from non-bacterial gastroenteritis. *J Clin Invest*. 1956;35(2):200-205. doi:[10.1172/JCI103264](https://doi.org/10.1172/JCI103264)

70. Graham DY, Jiang X, Tanaka T, Opekun AR, Madore HP, Estes MK. Norwalk virus infection of volunteers: New insights based on improved assays. *J Infect Dis*. 1994;170(1):34-43. doi:[10.1093/infdis/170.1.34](https://doi.org/10.1093/infdis/170.1.34)

71. Keswick BH, Satterwhite TK, Johnson PC, et al. Inactivation of Norwalk virus in drinking water by chlorine. *Appl Environ Microbiol*. 1985;50(2):261-264. doi:[10.1128/aem.50.2.261-264.1985](https://doi.org/10.1128/aem.50.2.261-264.1985)

72. Leon JS, Kingsley DH, Montes JS, et al. Randomized, double-blinded clinical trial for human norovirus inactivation in oysters by high hydrostatic pressure processing. *Appl Environ Microbiol*. 2011;77(15):5476-5482. doi:[10.1128/AEM.02801-10](https://doi.org/10.1128/AEM.02801-10)

73. Lindesmith L, Moe C, Lependu J, Frelinger JA, Treanor J, Baric RS. Cellular and humoral immunity following Snow Mountain virus challenge. *J Virol*. 2005;79(5):2900-2909. doi:[10.1128/JVI.79.5.2900-2909.2005](https://doi.org/10.1128/JVI.79.5.2900-2909.2005)

74. Madore HP, Treanor JJ, Buja R, Dolin R. Antigenic relatedness among the Norwalk-like agents by serum antibody rises. *J Med Virol*. 1990;32(2):96-101. doi:[10.1002/jmv.1890320206](https://doi.org/10.1002/jmv.1890320206)

75. Mateo R, Lindesmith LC, Garg SJ, et al. Production and clinical evaluation of norwalk GI.1 virus lot 001-09NV in norovirus vaccine development. *J Infect Dis*. 2020;221(6):919-926. doi:[10.1093/infdis/jiz540](https://doi.org/10.1093/infdis/jiz540)

76. Okhuysen PC, Jiang X, Ye L, Johnson PC, Estes MK. Viral shedding and fecal IgA response after Norwalk virus infection. *J Infect Dis*. 1995;171(3):566-569. doi:[10.1093/infdis/171.3.566](https://doi.org/10.1093/infdis/171.3.566)

77. Parker SP, Cubitt WD. Measurement of IgA responses following Norwalk virus infection and other human caliciviruses using a recombinant Norwalk virus protein EIA. *Epidemiol Infect*. 1994;113(1):143-151. doi:[10.1017/s0950268800051566](https://doi.org/10.1017/s0950268800051566)

78. Seitz SR, Leon JS, Schwab KJ, et al. Norovirus infectivity in humans and persistence in water. *Appl Environ Microbiol*. 2011;77(19):6884-6888. doi:[10.1128/AEM.05806-11](https://doi.org/10.1128/AEM.05806-11)

79. Treanor JJ, Madore HP, Dolin R. Development of an enzyme immunoassay for the Hawaii agent of viral gastroenteritis. *J Virol Methods*. 1988;22(2-3):207-214. doi:[10.1016/0166-0934(88)90103-6](https://doi.org/10.1016/0166-0934(88)90103-6)

80. Ajami NJ, Barry MA, Carrillo B, et al. Antibody responses to norovirus genogroup GI.1 and GII.4 proteases in volunteers administered Norwalk virus. *Clin Vaccine Immunol*. 2012;19(12):1980-1983. doi:[10.1128/CVI.00411-12](https://doi.org/10.1128/CVI.00411-12)

81. Atmar RL, Bernstein DI, Lyon GM, et al. Serological correlates of protection against a GII.4 norovirus. *Clin Vaccine Immunol*. 2015;22(8):923-929. doi:[10.1128/CVI.00196-15](https://doi.org/10.1128/CVI.00196-15)

82. Ball JM, Graham DY, Opekun AR, Gilger MA, Guerrero RA, Estes MK. Recombinant Norwalk virus-like particles given orally to volunteers: Phase I study. *Gastroenterology*. 1999;117(1):40-48. doi:[10.1016/s0016-5085(99)70548-2](https://doi.org/10.1016/s0016-5085(99)70548-2)

83. Brinker JP, Blacklow NR, Estes MK, Moe CL, Schwab KJ, Herrmann JE. Detection of Norwalk virus and other genogroup 1 human caliciviruses by a monoclonal antibody, recombinant-antigen-based immunoglobulin M capture enzyme immunoassay. *J Clin Microbiol*. 1998;36(4):1064-1069. doi:[10.1128/JCM.36.4.1064-1069.1998](https://doi.org/10.1128/JCM.36.4.1064-1069.1998)

84. Brinker JP, Blacklow NR, Jiang X, Estes MK, Moe CL, Herrmann JE. Immunoglobulin M antibody test to detect genogroup II Norwalk-like virus infection. *J Clin Microbiol*. 1999;37(9):2983-2986. doi:[10.1128/JCM.37.9.2983-2986.1999](https://doi.org/10.1128/JCM.37.9.2983-2986.1999)

85. Czako R, Atmar RL, Opekun AR, Gilger MA, Graham DY, Estes MK. Serum hemagglutination inhibition activity correlates with protection from gastroenteritis in persons infected with Norwalk virus. *Clin Vaccine Immunol*. 2012;19(2):284-287. doi:[10.1128/CVI.05592-11](https://doi.org/10.1128/CVI.05592-11)

86. Czako R, Atmar RL, Opekun AR, Gilger MA, Graham DY, Estes MK. Experimental human infection with Norwalk virus elicits a surrogate neutralizing antibody response with cross-genogroup activity. *Clin Vaccine Immunol*. 2015;22(2):221-228. doi:[10.1128/CVI.00516-14](https://doi.org/10.1128/CVI.00516-14)

87. Dai YC, Zhang XF, Xia M, et al. Antigenic relatedness of norovirus GII.4 variants determined by human challenge sera. *PLoS One*. 2015;10(4):e0124945. doi:[10.1371/journal.pone.0124945](https://doi.org/10.1371/journal.pone.0124945)

88. Erdman DD, Gary GW, Anderson LJ. Serum immunoglobulin A response to Norwalk virus infection. *J Clin Microbiol*. 1989;27(6):1417-1418. doi:[10.1128/jcm.27.6.1417-1418.1989](https://doi.org/10.1128/jcm.27.6.1417-1418.1989)

89. Erdman DD, Gary GW, Anderson LJ. Development and evaluation of an IgM capture enzyme immunoassay for diagnosis of recent Norwalk virus infection. *J Virol Methods*. 1989;24(1-2):57-66. doi:[10.1016/0166-0934(89)90007-4](https://doi.org/10.1016/0166-0934(89)90007-4)

90. Gary GW, Anderson LJ, Keswick BH, et al. Norwalk virus antigen and antibody response in an adult volunteer study. *J Clin Microbiol*. 1987;25(10):2001-2003. doi:[10.1128/jcm.25.10.2001-2003.1987](https://doi.org/10.1128/jcm.25.10.2001-2003.1987)

91. Gray JJ, Cunliffe C, Ball J, Graham DY, Desselberger U, Estes MK. Detection of immunoglobulin M (IgM), IgA, and IgG Norwalk virus-specific antibodies by indirect enzyme-linked immunosorbent assay with baculovirus-expressed Norwalk virus capsid antigen in adult volunteers challenged with Norwalk virus. *J Clin Microbiol*. 1994;32(12):3059-3063. doi:[10.1128/jcm.32.12.3059-3063.1994](https://doi.org/10.1128/jcm.32.12.3059-3063.1994)

92. Griffin SM, Converse RR, Leon JS, et al. Application of salivary antibody immunoassays for the detection of incident infections with Norwalk virus in a group of volunteers. *J Immunol Methods*. 2015;424:53-63. doi:[10.1016/j.jim.2015.05.001](https://doi.org/10.1016/j.jim.2015.05.001)

93. Grohmann GS, Murphy AM, Christopher PJ, Auty E, Greenberg HB. Norwalk virus gastroenteritis in volunteers consuming depurated oysters. *Aust J Exp Biol Med Sci*. 1981;59:219-228. doi:[10.1038/icb.1981.17](https://doi.org/10.1038/icb.1981.17)

94. Harrington PR, Lindesmith L, Yount B, Moe CL, Baric RS. Binding of Norwalk virus-like particles to ABH histo-blood group antigens is blocked by antisera from infected human volunteers or experimentally vaccinated mice. *J Virol*. 2002;76(23):12335-12343. doi:[10.1128/jvi.76.23.12335-12343.2002](https://doi.org/10.1128/jvi.76.23.12335-12343.2002)

95. Hesse S, Neill FH, Estes MK, et al. Serological responses to a norovirus nonstructural fusion protein after vaccination and infection. *Clin Vaccine Immunol*. 2016;23(2):181-183. doi:[10.1128/CVI.00595-15](https://doi.org/10.1128/CVI.00595-15)

96. Hutson AM, Atmar RL, Graham DY, Estes MK. Norwalk virus infection and disease is associated with ABO histo-blood group type. *J Infect Dis*. 2002;185(9):1335-1337. doi:[10.1086/339883](https://doi.org/10.1086/339883)

97. Hutson AM, Airaud F, LePendu J, Estes MK, Atmar RL. Norwalk virus infection associates with secretor status genotyped from sera. *J Med Virol*. 2005;77(1):116-120. doi:[10.1002/jmv.20423](https://doi.org/10.1002/jmv.20423)

98. Jiang X, Wang J, Graham DY, Estes MK. Detection of Norwalk virus in stool by polymerase chain reaction. *J Clin Microbiol*. 1992;30(10):2529-2534. doi:[10.1128/jcm.30.10.2529-2534.1992](https://doi.org/10.1128/jcm.30.10.2529-2534.1992)

99. Jiang X, Wang J, Estes MK. Characterization of SRSVs using RT-PCR and a new antigen ELISA. *Arch Virol*. 1995;140(2):363-374. doi:[10.1007/BF01309870](https://doi.org/10.1007/BF01309870)

100. Kavanagh O, Estes MK, Reeck A, et al. Serological responses to experimental Norwalk virus infection measured using a quantitative duplex time-resolved fluorescence immunoassay. *Clin Vaccine Immunol*. 2011;18(7):1187-1190. doi:[10.1128/CVI.00039-11](https://doi.org/10.1128/CVI.00039-11)

101. Kirby AE, Shi J, Montes J, Lichtenstein M, Moe CL. Disease course and viral shedding in experimental Norwalk virus and Snow Mountain virus infection. *J Med Virol*. 2014;86(12):2055-2064. doi:[10.1002/jmv.23905](https://doi.org/10.1002/jmv.23905)

102. Kirby AE, Streby A, Moe CL. Vomiting as a symptom and transmission risk in norovirus illness: Evidence from human challenge studies. *PLoS One*. 2016;11(4):e0143759. doi:[10.1371/journal.pone.0143759](https://doi.org/10.1371/journal.pone.0143759)

103. Kirby AE, Kienast Y, Aldeco M, et al. Snow Mountain Virus recovery by synthetic human histo-blood group antigens is heavily influenced by matrix effects. *Sci Rep*. 2020;10(1):4661. doi:[10.1038/s41598-020-60639-6](https://doi.org/10.1038/s41598-020-60639-6)

104. Lindesmith LC, Donaldson E, Leon J, et al. Heterotypic humoral and cellular immune responses following Norwalk virus infection. *J Virol*. 2010;84(4):1800-1815. doi:[10.1128/JVI.02179-09](https://doi.org/10.1128/JVI.02179-09)

105. Lindesmith LC, Beltramello M, Swanstrom J, et al. Serum immunoglobulin a cross-strain blockade of human noroviruses. *Open Forum Infect Dis*. 2015;2(3):ofv084. doi:[10.1093/ofid/ofv084](https://doi.org/10.1093/ofid/ofv084)

106. Lindesmith LC, McDaniel JR, Changela A, et al. Sera antibody repertoire analyses reveal mechanisms of broad and pandemic strain neutralizing responses after human norovirus vaccination. *Immunity*. 2019;50(6):1530-1541 e8. doi:[10.1016/j.immuni.2019.05.007](https://doi.org/10.1016/j.immuni.2019.05.007)

107. Liu P, Escudero B, Jaykus LA, et al. Laboratory evidence of norwalk virus contamination on the hands of infected individuals. *Appl Environ Microbiol*. 2013;79(24):7875-7881. doi:[10.1128/AEM.02576-13](https://doi.org/10.1128/AEM.02576-13)

108. Liu P, Rahman M, Leon J, Moe C. Less severe clinical symptoms of Norwalk virus 8fIIb inoculum compared to its precursor 8fIIa from human challenge studies. *J Med Virol*. 2021;93(6):3557-3563. doi:[10.1002/jmv.26578](https://doi.org/10.1002/jmv.26578)

109. LoBue AD, Lindesmith L, Yount B, et al. Multivalent norovirus vaccines induce strong mucosal and systemic blocking antibodies against multiple strains. *Vaccine*. 2006;24(24):5220-5234. doi:[10.1016/j.vaccine.2006.03.080](https://doi.org/10.1016/j.vaccine.2006.03.080)

110. Messner MJ, Berger P, Nappier SP. Fractional poisson–a simple dose-response model for human norovirus. *Risk Anal*. 2014;34(10):1820-1829. doi:[10.1111/risa.12207](https://doi.org/10.1111/risa.12207)

111. Moe CL, Sair A, Lindesmith L, Estes MK, Jaykus LA. Diagnosis of norwalk virus infection by indirect enzyme immunoassay detection of salivary antibodies to recombinant norwalk virus antigen. *Clin Diagn Lab Immunol*. 2004;11(6):1028-1034. doi:[10.1128/CDLI.11.6.1028-1034.2004](https://doi.org/10.1128/CDLI.11.6.1028-1034.2004)

112. Newman KL, Marsh Z, Kirby AE, Moe CL, Leon JS. Immunocompetent adults from human norovirus challenge studies do not exhibit norovirus viremia. *J Virol*. 2015;89(13):6968-6969. doi:[10.1128/JVI.00392-15](https://doi.org/10.1128/JVI.00392-15)

113. Newman KL, Moe CL, Kirby AE, Flanders WD, Parkos CA, Leon JS. Human norovirus infection and the acute serum cytokine response. *Clin Exp Immunol*. 2015;182(2):195-203. doi:[10.1111/cei.12681](https://doi.org/10.1111/cei.12681)

114. Newman KL, Moe CL, Kirby AE, Flanders WD, Parkos CA, Leon JS. Norovirus in symptomatic and asymptomatic individuals: Cytokines and viral shedding. *Clin Exp Immunol*. 2016;184(3):347-357. doi:[10.1111/cei.12772](https://doi.org/10.1111/cei.12772)

115. Parker S, Cubitt D, Jiang JX, Estes M. Efficacy of a recombinant Norwalk virus protein enzyme immunoassay for the diagnosis of infections with Norwalk virus and other human "candidate" caliciviruses. *J Med Virol*. 1993;41(3):179-184. doi:[10.1002/jmv.1890410302](https://doi.org/10.1002/jmv.1890410302)

116. Patin NV, Pena-Gonzalez A, Hatt JK, Moe C, Kirby A, Konstantinidis KT. The role of the gut microbiome in resisting norovirus infection as revealed by a human challenge study. *mBio*. 2020;11(6). doi:[10.1128/mBio.02634-20](https://doi.org/10.1128/mBio.02634-20)

117. Ramani S, Neill FH, Opekun AR, et al. Mucosal and cellular immune responses to norwalk virus. *J Infect Dis*. 2015;212(3):397-405. doi:[10.1093/infdis/jiv053](https://doi.org/10.1093/infdis/jiv053)

118. Ramani S, Neill FH, Ferreira J, et al. B-cell responses to intramuscular administration of a bivalent virus-like particle human norovirus vaccine. *Clin Vaccine Immunol*. 2017;24(5). doi:[10.1128/CVI.00571-16](https://doi.org/10.1128/CVI.00571-16)

119. Reeck A, Kavanagh O, Estes MK, et al. Serological correlate of protection against norovirus-induced gastroenteritis. *J Infect Dis*. 2010;202(8):1212-1218. doi:[10.1086/656364](https://doi.org/10.1086/656364)

120. Swanstrom J, Lindesmith LC, Donaldson EF, Yount B, Baric RS. Characterization of blockade antibody responses in GII.2.1976 Snow Mountain virus-infected subjects. *J Virol*. 2014;88(2):829-837. doi:[10.1128/JVI.02793-13](https://doi.org/10.1128/JVI.02793-13)

121. Treanor JJ, Jiang X, Madore HP, Estes MK. Subclass-specific serum antibody responses to recombinant Norwalk virus capsid antigen (rNV) in adults infected with Norwalk, Snow Mountain, or Hawaii virus. *J Clin Microbiol*. 1993;31(6):1630-1634. doi:[10.1128/jcm.31.6.1630-1634.1993](https://doi.org/10.1128/jcm.31.6.1630-1634.1993)

122. Williams AM, Ladva CN, Leon JS, et al. Changes in micronutrient and inflammation serum biomarker concentrations after a norovirus human challenge. *Am J Clin Nutr*. 2019;110(6):1456-1464. doi:[10.1093/ajcn/nqz201](https://doi.org/10.1093/ajcn/nqz201)

123. Wyatt RG, Dolin R, Blacklow NR, et al. Comparison of three agents of acute infectious nonbacterial gastroenteritis by cross-challenge in volunteers. *J Infect Dis*. 1974;129(6):709-714. doi:[10.1093/infdis/129.6.709](https://doi.org/10.1093/infdis/129.6.709)

# Other information

The review has not been registered and no formal protocol was prepared before the review process began. The authors wish to declare that they have no competing interests.

All data sheets, along with cleaning and analysis code, are available on [GitHub](link), and the code at time of writing is archived on Zenodo: [doi url](link).

# R session information

**R version 4.1.1 (2021-08-10)**

**Platform:** x86\_64-w64-mingw32/x64 (64-bit)

**locale:** *LC\_COLLATE=English\_United States.1252*, *LC\_CTYPE=English\_United States.1252*, *LC\_MONETARY=English\_United States.1252*, *LC\_NUMERIC=C* and *LC\_TIME=English\_United States.1252*

**attached base packages:** *stats*, *graphics*, *grDevices*, *utils*, *datasets*, *methods* and *base*

**other attached packages:** *dplyr(v.1.0.7)*, *DiagrammeRsvg(v.0.1)*, *rsvg(v.2.1.2)*, *magick(v.2.7.3)*, *janitor(v.2.1.0)*, *here(v.1.0.1)*, *dmetar(v.0.0.9000)*, *metafor(v.3.0-2)*, *Matrix(v.1.3-4)*, *meta(v.5.0-0)*, *robvis(v.0.3.0)*, *PRISMA2020(v.0.0.3)*, *flextable(v.0.6.9)*, *knitr(v.1.36)*, *bookdown(v.0.24)* and *rmarkdown(v.2.11)*

**loaded via a namespace (and not attached):** *nlme(v.3.1-153)*, *lubridate(v.1.7.10)*, *rprojroot(v.2.0.2)*, *prabclus(v.2.3-2)*, *tools(v.4.1.1)*, *utf8(v.1.2.2)*, *R6(v.2.5.1)*, *DBI(v.1.1.1)*, *colorspace(v.2.0-2)*, *nnet(v.7.3-16)*, *netmeta(v.2.0-1)*, *tidyselect(v.1.1.1)*, *gridExtra(v.2.3)*, *curl(v.4.3.2)*, *compiler(v.4.1.1)*, *xml2(v.1.3.2)*, *officer(v.0.4.0)*, *diptest(v.0.76-0)*, *scales(v.1.1.1)*, *DEoptimR(v.1.0-9)*, *robustbase(v.0.93-9)*, *systemfonts(v.1.0.2)*, *stringr(v.1.4.0)*, *digest(v.0.6.28)*, *minqa(v.1.2.4)*, *base64enc(v.0.1-3)*, *pkgconfig(v.2.0.3)*, *htmltools(v.0.5.2)*, *MuMIn(v.1.43.17)*, *lme4(v.1.1-27.1)*, *highr(v.0.9)*, *fastmap(v.1.1.0)*, *rlang(v.0.4.11)*, *generics(v.0.1.0)*, *jsonlite(v.1.7.2)*, *mclust(v.5.4.7)*, *zip(v.2.2.0)*, *magrittr(v.2.0.1)*, *modeltools(v.0.2-23)*, *Rcpp(v.1.0.7)*, *munsell(v.0.5.0)*, *fansi(v.0.5.0)*, *abind(v.1.4-5)*, *gdtools(v.0.2.3)*, *lifecycle(v.1.0.1)*, *stringi(v.1.7.4)*, *yaml(v.2.2.1)*, *CompQuadForm(v.1.4.3)*, *snakecase(v.0.11.0)*, *mathjaxr(v.1.4-0)*, *MASS(v.7.3-54)*, *flexmix(v.2.3-17)*, *grid(v.4.1.1)*, *parallel(v.4.1.1)*, *ggrepel(v.0.9.1)*, *crayon(v.1.4.2)*, *lattice(v.0.20-45)*, *splines(v.4.1.1)*, *pander(v.0.6.4)*, *poibin(v.1.5)*, *pillar(v.1.6.4)*, *uuid(v.0.1-4)*, *boot(v.1.3-28)*, *fpc(v.2.2-9)*, *stats4(v.4.1.1)*, *magic(v.1.5-9)*, *glue(v.1.4.2)*, *evaluate(v.0.14)*, *V8(v.3.4.2)*, *data.table(v.1.14.2)*, *vctrs(v.0.3.8)*, *png(v.0.1-7)*, *nloptr(v.1.2.2.2)*, *gtable(v.0.3.0)*, *purrr(v.0.3.4)*, *kernlab(v.0.9-29)*, *assertthat(v.0.2.1)*, *ggplot2(v.3.3.5)*, *xfun(v.0.26)*, *class(v.7.3-19)*, *tibble(v.3.1.4)*, *cluster(v.2.1.2)* and *ellipsis(v.0.3.2)*

# Study quality rubric

1. (Clear Inclusion Criteria) Were there clear criteria for inclusion in the case series?
   * Yes: Inclusion criteria are stated clearly in the study.
   * No: Study was sampled by convenience or without clear exclusion criteria.
   * Unclear: Criteria for inclusion are not described.
2. (Reliable Condition Measure) Was the condition measured in a standard, reliable way for all participants included in the case series?
   * Yes: The same method was used to assess norovirus infection for all patients.
   * No: Participants were assessed for infection in different ways.
   * Unclear: Method of assessing infection was not reported.
3. (Valid Condition Methods) Were valid methods used for identification of the condition for all participants included in the case series?
   * Yes: Infection was assessed using any molecular method (e.g. serology, PCR, etc.).
   * No: Infection was assessed using clinical symptoms alone.
   * Unclear: Method of assessing infection is not reported.
4. (Same Cohort) Were all participants sampled from the same underlying population at the same time? (I.e. did the participants form a single cohort?)
   * Yes: All participants were sampled from the same general population, or if multiple cohorts were pooled together during the study, the results are stratified by cohort. All patients were infected with the same inoculum, or stratified by inoculum.
   * No: Participants were recruited heterogeneously, or pooled cohort results were not stratified. Or patients with different inocula were pooled.
   * Unclear: Impossible to tell from given information whether participants were pooled from various cohorts without stratification or not.
5. (Complete Demographics) Was there clear reporting of the demographics of the participants included in the study?
   * Yes: Age range of participants was reported and medical history of patients was mentioned.
   * No: Age range is not reported or medical history not mentioned.
6. (Complete Clinical) Was there clear reporting of clinical information of the participants?
   * Yes: Infection outcome is reported for the same number of patients who were recruited. Or, if numbers are different, the loss of participants is explained.
   * No: Number of recruits reported does not match results, or number of total recruits is not reported.
7. (Site Information) Was there clear reporting of the presenting site or clinic demography?
   * Yes: authors describe the target population and state where the trial was conducted.
   * No: authors do not describe the study population or do not state where the trial was conducted.
   * Unclear: authors briefly describe the study site without detail, or describe the study site only as “multiple centers” or equivalent.
8. Overall risk of bias
   * Low: Yes in at least six domains, No in zero domains.
   * Moderate: Yes in at least five domains, No in at most one domain.
   * High: Yes in fewer than five domains, or No in more than one domain.

# Supplementary figures

The diagnostic results for the leave-one-out analysis are shown in Figure 16, the results of the three clustering algorithms applied to the GOSH data are shown along with diagnostics in Figure 17, and the GOSH distributions based on inclusion and exclusion of the identified outliers by the GOSH clustering analysis are shown in Figure 18.

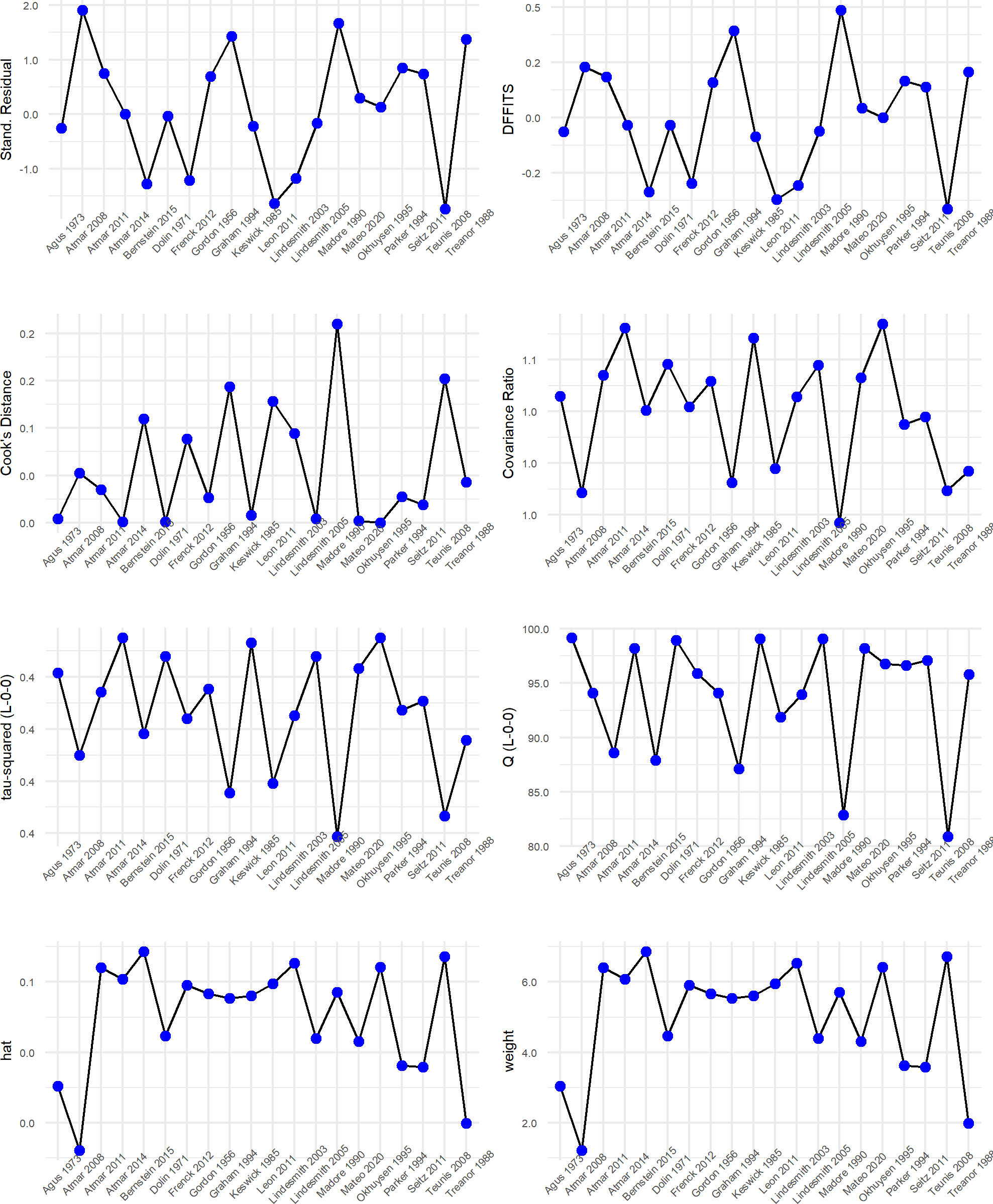


Figure 16: Diagnostics for leave-one-out influence analysis recommended by Viechtbauer. No studies appear to be critical outliers which lie far beyond the others in any metric, so the Baujat plot was used instead to identify outliers from the LOO analysis.

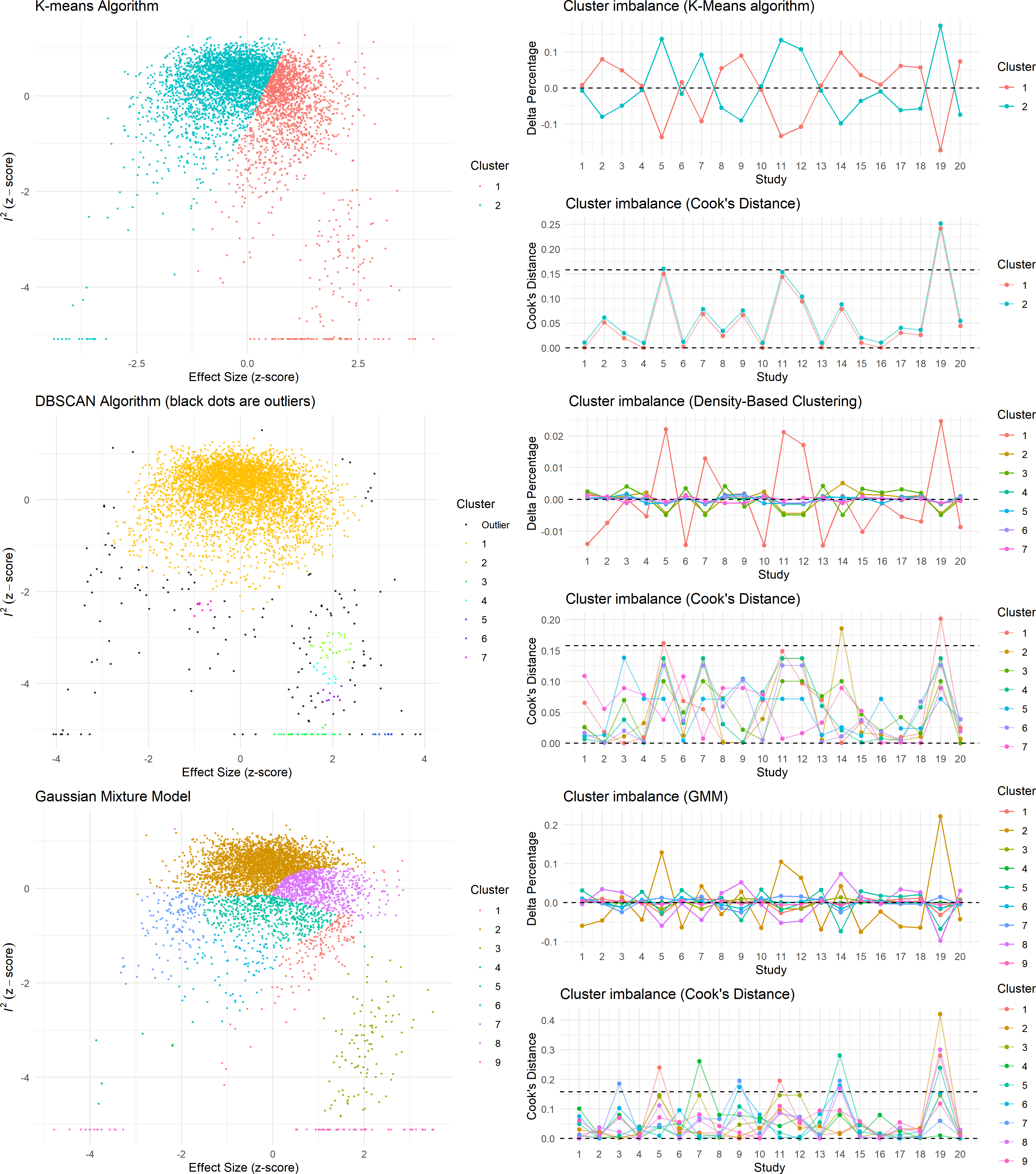


Figure 17: Clusters identified by the three different algorithms recommended by Harrer. Parameters for clustering algorithms were chosen based on Harrer’s recommendation to obtain a reasonable number of detected outliers.

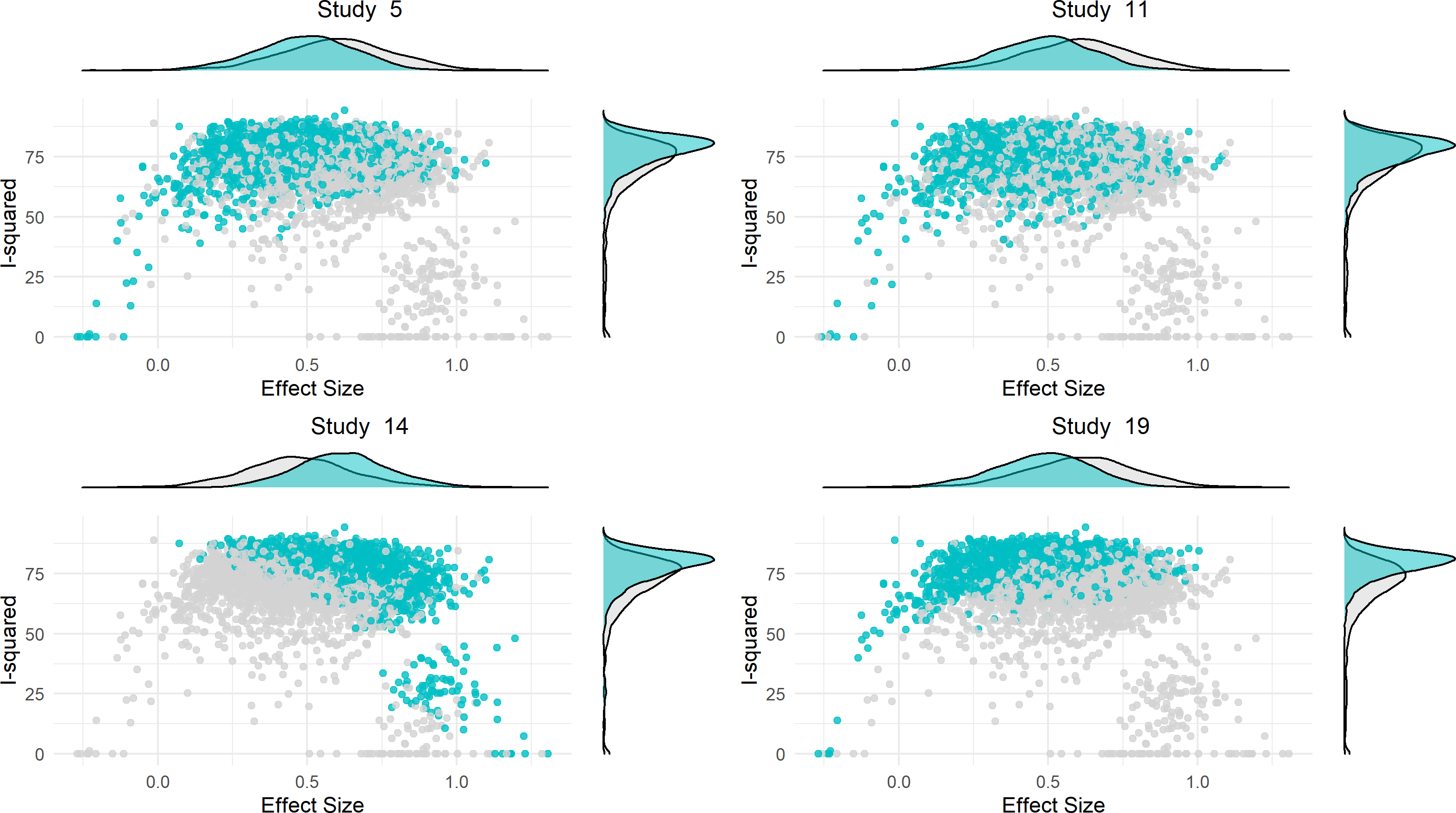


Figure 18: Differences in GOSH distributions for effect size and heterogeneity between random subsets which contain (blue) or do not contain (gray) a given study. Study 5 is Bernstein 2015, study 11 is Leon 2011, study 14 is Madore 1990, and study 19 is Teunis 2008. These four studies were identified as having noticeable differences in included vs. excluded distributions in the GOSH analysis based on cluster imbalance across the three algorithms. (I.e. based on which studies are over-represented in different clusters.)