Face masks prevent transmission of respiratory diseases: A meta-analysis of randomized controlled trials*

Hanna M. Ollila[1,2,3,], Markku Partinen[4,5], Jukka Koskela[1,2,6], Riikka Savolainen[7], Anna Rotkirch[8], and Liisa T. Laine[9,10]

[1]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland

[2]Broad Institute of MIT and Harvard, Cambridge, MA, USA

[3]Stanford University School of Medicine, Palo Alto, CA, USA

[4]Helsinki Sleep Clinic, Vitalmed Research Center

[5]Department of Clinical Neurosciences, Clinicum, University of Helsinki, Helsinki, Finland

[1]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland

[2]Broad Institute of MIT and Harvard, Cambridge, MA, USA

[6]Helsinki University and Helsinki University Hospital, Clinic of Gastroenterology Helsinki, Finland

[7]Newcastle University Business School, Newcastle-upon-Tyne, United Kingdom

[8]Population Research Institute, Väestöliitto – The Family Federation of Finland

[9]University of Pennsylvania, The Wharton School, Philadelphia, PA, USA

[10]Department of Medical Ethics and Health Policy, The Perelman School of Medicine, Philadelphia, PA, USA

*This version: August 6, 2020. We thank Tuomas Aivelo, Thomas Brand, Matti Heino, Risto Kuosa, Henri Lempikoski, Panu Poutvaara, Antti Ripatti, Harri Tohmo, Paula Bergman and Viljami Virolainen for their feedback and comments. Laine gratefully acknowledges funding from the National Institute on Aging of the National Institutes of Health under Award Number P30AG043073. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Corresponding author: Hanna M. Ollila, FIMM-EMBL Group leader, Institute for Molecular Medicine Finland, University of Helsinki, Finland; hanna.m.ollila@helsinki.fi and Liisa T. Laine, Department of Medical Ethics and Health Policy, The Perelman School of Medicine, Philadelphia, PA, USA; lainel@wharton.upenn.edu

Summary

Background. Coronavirus Disease 2019 (COVID-19) is caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) and spreads through droplet-mediated transmission on contaminated surfaces and in air. Mounting scientific evidence from observational studies suggests that face masks for the general public may reduce the spread of infections. However, results from evaluations based on randomized control trials (RCT) have been presented as inconclusive. Thus, concerns related to the effects and safety of non-surgical face masks in non-clinical settings remain. This controversy calls for a meta-analysis of RCTs conducted in non-hospital and non-household settings, and which also adjusts for non-compliance in RCTs, the time-lag in benefits of masking, and possible adverse effects.

Methods. We performed a meta-analysis of RCTs of non-surgical face masks in preventing viral respiratory infections in non-hospital and non-household settings at cumulative and maximum follow-up as primary endpoints. The search for RCTs yielded five studies published before May 29th, 2020. We pooled estimates from the studies and performed a random-effects meta-analysis and a mixed-effects meta-regression across the studies, accounting for covariates in compliance vs. non-compliance in treatment.

Results. Face masks decreased infections across all the studies at maximum follow-up ($p < 0.05$, $RR = 0.608$ $[0.387 - 0.956]$), and particularly in the studies without a non-compliance bias. We found significant between-study heterogeneity in the studies with a bias ($I^2 = 71.2\%, p < 0.01$). We also performed an adjusted meta-regression to account for heterogeneity. There is a statistically significant protective effect of masking ($p < 0.001, \beta = 0.0214, SE = 0.0062$). No severe adverse effects were reported.

Interpretation. Our meta-analysis using existing randomized control studies on non-hospital and non-household settings found support that face masks prevent respiratory infections among the general public. No serious health harms to the face mask users were detected. Recommendations and clear communication on the benefits of face masks should be provided to limit spread of respiratory infections, including COVID-19.

Funding. National Institute on Aging of the National Institutes of Health under Award Number P30AG043073.

1 Introduction

COVID-19 is caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) and spreads through droplet-mediated transmission on contaminated surfaces and in air. COVID-19 has caught the medical community and policymakers off guard through both the pace and initial unpredictability of viral transmission. Minimizing the exposure time of the uninfected to viral particles in droplets and aerosols is central to limit the spread of the disease. Moreover, the role of superspreading events seems to be crucial for transmission of COVID-19[1], and preventing a single superspreading event can have a large impact on the spread of the virus (Figure 1).[1] Also, viral particles may linger in the air, even after the infected individual has left the shared space [5, 6]. Limiting the time and the magnitude of a release of droplets and aerosols by those infected is thus emerging as the key factor in reducing COVID-19 transmission.

A combination of different non-pharmaceutical interventions (NPIs), including testing and tracing, good hand hygiene, maintaining physical distance, and the use of cloth masks and coverings (masking), is currently seen as the primary intervention to limit COVID-19 infections. Among these NPIs, face masks have caused a scientific and political debate, leading to conflicting and confusing messages to the general public. Doubts and controversy about the safety and efficacy of face masks worn by the public has been substantial. Even though masks are currently mandated or recommended in several countries [7], a some differences remain (Figure 2). In addition, concerns over the potential harms have been voiced, such as a possibility that face masks may spread COVID-19 or that masks could potentially create a false sense of security [8].

We examine whether there is an independent protective, or predisposing, effect (in addition to the pure source control) of the use of face masks on respiratory infections by conducting a meta-analysis focusing solely on face mask RCTs in community (non-hospital and non-household) settings. Indeed, the gap in the field is that only a few RCTs in a community setting exist. Currently, the largest meta-analysis contains three such publications [9]. A larger set of publications has been included in one white paper [10]. However, the previous literature does not contain a formal meta-analysis of the RCTs in a community setting that are currently available.

[1] A visualization of possible infection pathways shows that when 80% of individuals use any face covering, including cloth masks, the infection rates could be halved [2, 3, 4].

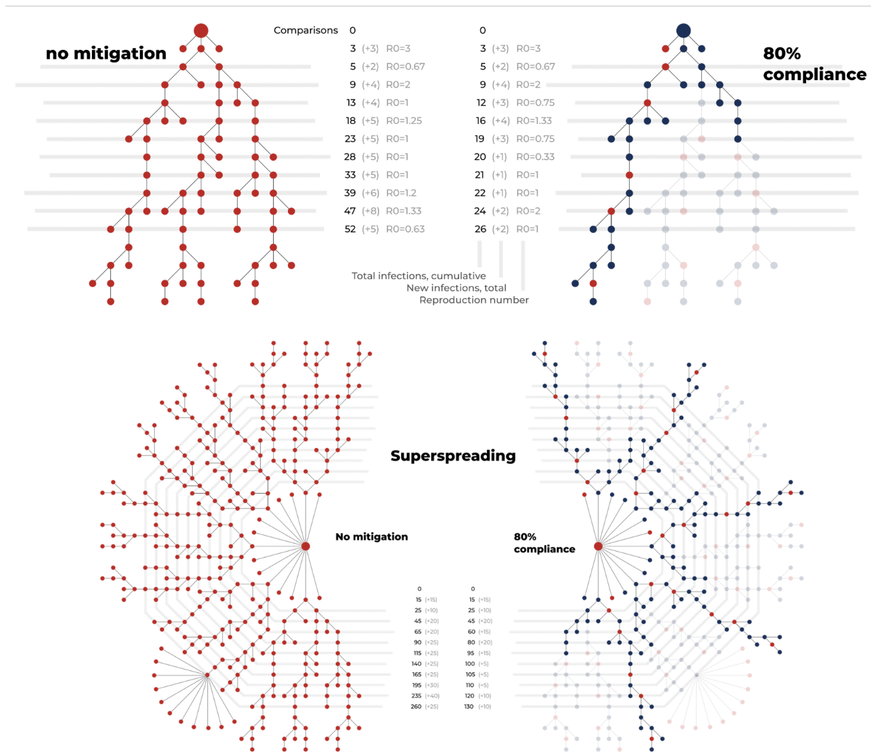

FIGURE 1: Illustration of possible effects of face mask use in preventing infections.

Notes: Typical chain of infections without masking (top). When 80 percent compliance in mask use (blue) is achieved, half of the infections may be prevented. Visualization of cluster of infections through one super spreading event below. Mitigation through 80 percent compliance in face mask use together with other protective measures can prevent individual infections and infections through super spreading events. Red shows infection without mask, blue indicates use of face mask and infection. Image by Riina Rupponen.

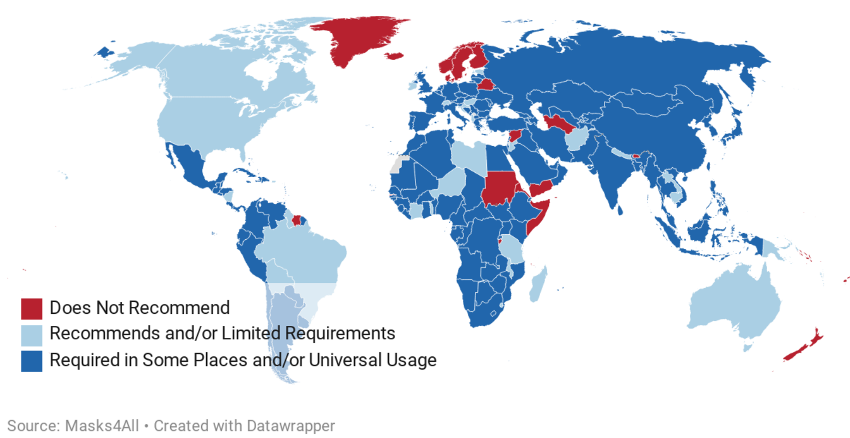

FIGURE 2: Countries that require or recommend public mask usage to help contain COVID-19 infections (date: July 30, 2020)

The current evidence of the effects of face masks stems from prediction models showing that universal masking in public can have a substantial impact on spreading and does not require the use of medical masks or 100% compliance [2, 3, 11]. A review assessing masks in source control (in contrast to protection) recommended their use in the general population [12]. Other empirical evidence indicates that countries and regions using masks have lower increases in COVID-19 infection rates and death rates [3]. Finally, a pooled meta-analysis of the spread of infectious viral diseases of up to 172 studies showed a consistent effect regarding the efficacy of face masks in preventing infections [4].

However, the epidemiologically most robust evidence on the effects of face mask use among the general public – that from randomized control trials – has been consistently noted as still lacking, also among scholars favoring mask use [13, 14]. Several studies also note that non-compliance may have skewed and confused RCT results, yet no study has statistically corrected for non-compliance in a meta-analysis (ibid). We fill this gap in the literature.

2 Search strategy and selection criteria

Data collection process. Only a few analyses of RCTs and face masks have been conducted in order to assess the ability of masks to protect the wearer from the infection. Barasheed et al. (2011)[15] and Brainard et al. (2020)[9] included three RCTs of face masks in non-hospital settings. They found very weak support for face mask efficacy, and stressed the problem of non-compliance in the few existing RCTs.

Commissioned report done by the Ministry of Social Affairs and Health in Finland and the consulting group Summaryx Oy included a systematic review (MSAH 2020)[10] in spring 2020 and detected in total five RCTs in non-hospital and non-household settings which were published prior to May 29th, 2020. Three of them were used in Baynard et al. (2020), namely, Aiello et al. (2010)[16], Aiello et al. (2012)[17], and Alfelali et al. (2019)[18]. In addition, there were two more RCTs: Barasheed et al. (2014)[15] and Abdin et al.(2005)[19]. None of the five studies addressed heterogeneity and non-compliance of RCTs fully.

For our analysis, we consulted both Baynard et al. (2020) and MSAH (2020), but also performed the same searches independently and detected the same five RCTs. [15, 16, 17, 18, 19].

Search criteria. Our supplement Section A provides the flow chart of the original study inclusion and search strategy outlined here [10]. To summarize, the selection criteria included face masks that were

1. FFP1, cloth mask or surgical mask. Exclusion; FFP2 or FFP3.
2. Population prone to respiratory infection. Exclusion; health care workers
3. Comparison group; no face mask. Exclusion; different type of face mask
4. Outcome; Relative risk for infection, safety and efficacy, slowing of infection at the population level
5. Sample size and follow up did not have exclusion criteria
6. RCT with either individual or group level (clustered). Exclusion criteria; cohort study, case control study, study without controls
7. Setting. Community setting. Exclusion criteria; home, healthcare environment.
8. Publication format. Whole text available, preprints included. Exclusion criteria; only abstract available
9. Language of original publication; Finnish, English, Spanish, Danish, Swedish, Norwegian, German or French. Exclusion criteria; other language

Data items. We extracted relative risks (RR), number of individuals per RCT group, and baseline study characteristics including type of study; pilgrim, college dorm, household, use of masks in control group, study time points, outcome measures: symptoms for infection, laboratory confirmed infection, compliance in cases, and adverse effects.

Included studies. We included RCT studies based on the search criteria above as reported by MSAH (2020)[10].

Study endpoints. The primary endpoint was the relative risk of infection across the studies using a meta-regression. In addition, we computed fixed effects meta-analysis and random effects meta-analysis using The DerSimonian and Laird (DL) and restricted maximum likelihood (REML). Sub-analyses were calculated in the cohorts with a similar follow-up time, a similar population and no substantial heterogeneity.

Statistical analysis. Analyses were performed using R version 3.5.0 (2018-04-23) and the packages *meta*, *metafor*, *data.table*, *dplyr*, and visualized with the *metaviz* and *forest* packages. The key data fields for each study are shown in Table 1 and in our supplement.

Heterogeneity. We computed between-study heterogeneity as estimated with I^2 and τ using Random effects meta-analysis and Mixed-model Meta-analysis as implemented in the R functions *metagen* using REML and DL analysis.

Significance and summary estimates. Summarized effects were computed using Random effects meta-analysis, and meta-regression as implemented in rma.uni to account for those parameters that induced significant heterogeneity between the studies. For the comparable studies with similar settings and no evidence of heterogeneity we used Fixed effects meta-analysis as implemented in the *metagen* package in R.

Sensitivity analyses. We calculated the effects by using a leave-one-out analysis whereby the summary effects were computed for all the studies leaving one RCT study out at a time. This analysis was done to ensure that, with fewer than ten studies in the meta-analysis, the effects were not significantly driven by any single study. Our supplement includes various additional robustness checks, such as sensitivity analysis through subgroups, leave-one-out, adjusted effects, estimation of between study heterogeneity, and comparison of per study weights under different models, Fixed effects, Random effects (DL and REML), meta-regression, and with different platforms, R and STATA. We're currently working on finalizing the robustness check using Bayesian approach.

Estimation and accounting for the biases in the study setting

We discovered the following discrepancies and treated them in the analysis as follows.

1) Use of face masks in controls and accounting for the bias caused by non-compliance. One challenge in RCT studies is the possibility that controls opt in the treatment if it is freely available. Hence, randomization of cases and controls into face mask RCT groups may fail since they can be purchased or made at home. Indeed, in three studies out of five analyzed here, face masks were also used by the control group.

The amount of face mask use by controls ranged from 12% to 53% (Table 1). Especially in the studies in which the use of face masks is high in controls (up to 50%), a comparison between face mask users and controls can be problematic because of a high uncertainty in the estimates. We took this into account in the meta-regressions either by controlling for the percentage of the face mask use in controls or by dichotomizing the use of face masks.

2) Non-compliance of the face mask use in the intervention group. Compliance in the face mask group varied substantially, being lowest in the Alfelali study (27% of the intervention group used the face mask daily) [18], and highest in Barasheed (76%)[15].

3) Examining potential selective reporting within the studies. We analyzed the cumulative risk and the maximum follow-up study endpoints and computed the meta-analysis estimates separately for the maximum follow-up and the cumulative estimates in the studies.

All the studies provided cumulative risk estimates, or numbers to compute them. In addition, two studies had longer follow-up periods of up to six weeks from the trial start date. We included these study endpoints because the incubation period of respiratory infections ranges from 1 to 10 days on average depending on the infection. This means that any protective effects of face masks appear with a time lag. Therefore the estimates from the first few days when protective measures are used do not account for infections that have been contracted prior to wearing a mask.

4) Blindedness. All RCTs were unblinded due to the nature of face mask use.

3 Results

We computed a meta-analysis across five RCT studies using a log transformed relative risk to estimate the effect of face masks on respiratory infections (Table 1). While the Fixed effects meta-analysis was statistically significant both using cumulative risk estimates and using maximum follow-up time points ($p < 0.05$, $RR = 0.91$ $[0.84 - 0.99]$, $p < 0.001$, $RR = 0.81$ $[0.73 - 0.90]$ computing I^2 with the Random effects model meta-analysis showed significant statistical heterogeneity between the studies ($I^2 = 71.2\%$ $[27.0\%; 88.6\%]$, $\tau^2 = 0.0253[0.0045; 0.3801]$, heterogeneity $p < 0.005$). Furthermore, accounting for heterogeneity using the Random effects model meta-analysis showed a statistically significant association with the maximum follow-up endpoint ($p < 0.05, RR = 0.608[0.387 - 0.956]$), whereas the cumulative risk did not ($p = 0.1$) (Figure 3).

Table 1: Characteristics of included studies

Publication	Population	Endpoints	RR [CI95%]	N per comparison	Controls use masks (%)
Barasheed 2014	Pilgrims	Mask vs. no mask	0.576 [0.332-1.007]	11/25/28/25	12
Abdin 2005	Pilgrims	Mask and education vs no mask	0.97 [0.79-1.2]	129/381/126/359	33.6
Alfelali preprint	Pilgrims	Mask vs. no mask	1.079 [0.935-1.244]	354/2845/322/2817	53
Aiello 2012	Students	Mask and hand hygiene vs. no mask	0.78 [0.59-1.05]	31/318/51/319	0
Aiello 2010	Students	Mask and hand hygiene vs. no mask	0.88 [0.75-1.03]	92/224/177/310	0

Notes: RR refers to the relative risk as reported in the original publication or, when RR was not reported, computed for the purposes of the analysis using the reported probabilities from the original study. CI95 % refers to the 95 % confidence interval. N per comparison is Mask infected/Mask uninfected/Control infected/Control uninfected. Values for cumulative endpoint are shown here.

Variance can be induced to the measurement for example through differences in a) study setting b) ethnicity, sex and underlying demographic factors and c) validity of the randomization, such as compliance in treatment or blindness. When examining population characteristics within the selected studies, we were thus able to attribute study the majority of heterogeneity to at least one of the above-mentioned differences. The predominant differences were the type of study population and non-compliance in the control group (outlined in Table 1). Two studies were conducted in a college student community setting in the United States, and the college students were mainly White ($> 80\%$) with a subset of Asian (16%) students [16, 17]. These studies had adjusted for the differences within the study population and for underlying baseline characteristics. By contrast, three studies were conducted within a pilgrim population during a Hajj pilgrimage with predominantly Saudi Arabian, Middle Eastern or Asian populations [15, 18, 19], and only Alfelali had adjusted for baseline measures [18].

The largest differences within the RCTs were due to non-compliance within the control groups in the pilgrim studies: the controls used face masks between 12 and 53% of the time. Obviously, if half of the group who is supposed not to be using face masks is using face masks, such biases will affect the estimates substantially and reduce the power to reliably estimate any effect.

To formally adjust for non-compliance, we computed a meta-regression adjusting for the percentages of face mask use in controls. This accounted for a great proportion of the cumulative heterogeneity (p heterogeneity remaining $= 0.08$) and showed a statistically significant association

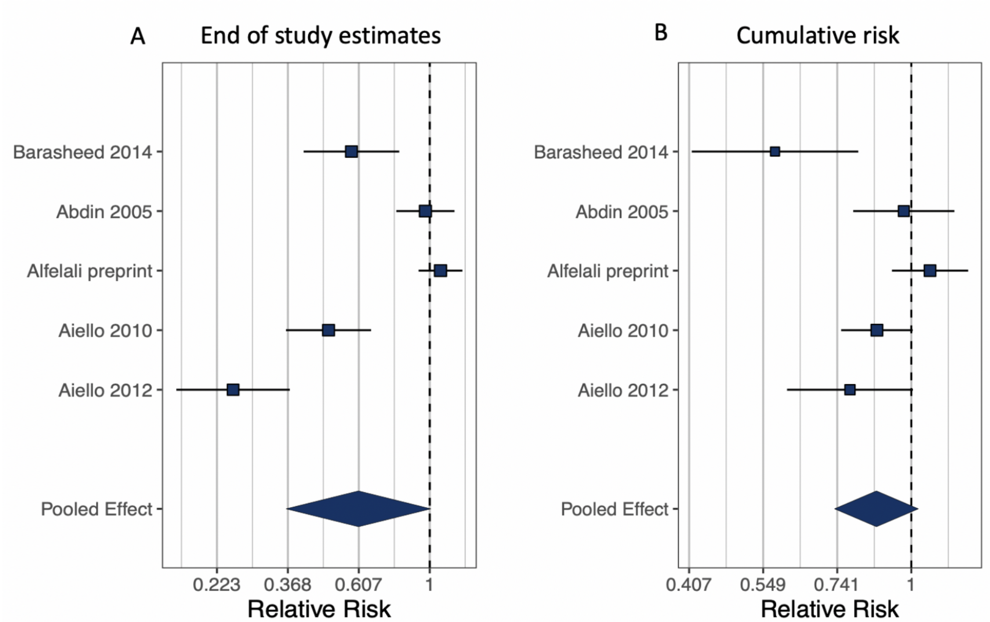

FIGURE 3: Random-effect meta-analysis of the five RCT studies.

Notes: There is statistically significant heterogeneity between studies.A statistically significant association was detected with Random effects meta-analysis a) at the maximum follow-up endpoints ($p < 0.05$), b) but not with cumulative estimates $p = 0.1$).

at the cumulative endpoints under the Fixed effects model ($p < 0.01$, $\beta = -0.0051$, $SE = 0.0019$).

While the maximum follow-up was statistically significant ($p < 0.0001$, $\beta = -0.0186$, $SE = 0.0026$), there is also statistically significant heterogeneity at the maximum follow-up endpoint ($p < 0.01$). We therefore computed a Random effects meta-regression using REML estimates, detecting similarly a statistically significant effect from face masks ($p < 0.001$, $\beta = -0.0214$, $SE = 0.0062$).

No study reported any severe adverse outcomes in the group using face masks. On the contrary, overall, masking did not increase the rate of infections but rather reduced laboratory-confirmed infections. Self-reported negative outcomes were assessed in the preprint by Alfelali, in which 26% of those that used face masks in Alfelali study reported a difficulty in breathing. Discomfort was assessed in the studies by Aiello and Alfelali, and the most commonly reported issue was discomfort (22%); in addition, 3% reported feeling hot, sweating, a bad smell, or blurred vision with eyeglasses [15, 16, 17, 18, 19].

To estimate the effect of face masks across similar study settings, we computed a conservative stratified analysis separately for those papers in which only the treatment group used face masks and for those in which also the controls used them. This stratified analysis found a strong protective effect for face masks in the studies where controls did not use masks ($p = 0.0344$; $\beta = -0.1653$ $[-0.3184; -0.0122]$ and $p = 0.0024$; $\beta = -0.8343$ $[-1.3741; -0.2946]$), using cumulative risk or maximum follow-up, respectively. The effects were strongest when combined with other protective measures and at the maximum follow-up, with the maximal time for the protective effect to develop (Figure 4).

Finally, we performed sensitivity analyses using a leave-one-out model under meta-regression. The adjusted effect sizes at the maximum follow-up were systematically distributed across the comparisons, suggesting that no single study was responsible for the protective effect, and the adjusted meta-regression estimates were statistically significant irrespective of which study was left out (Figure 5). These sensitivity analyses support the protective effect of face masks in preventing respiratory infections.

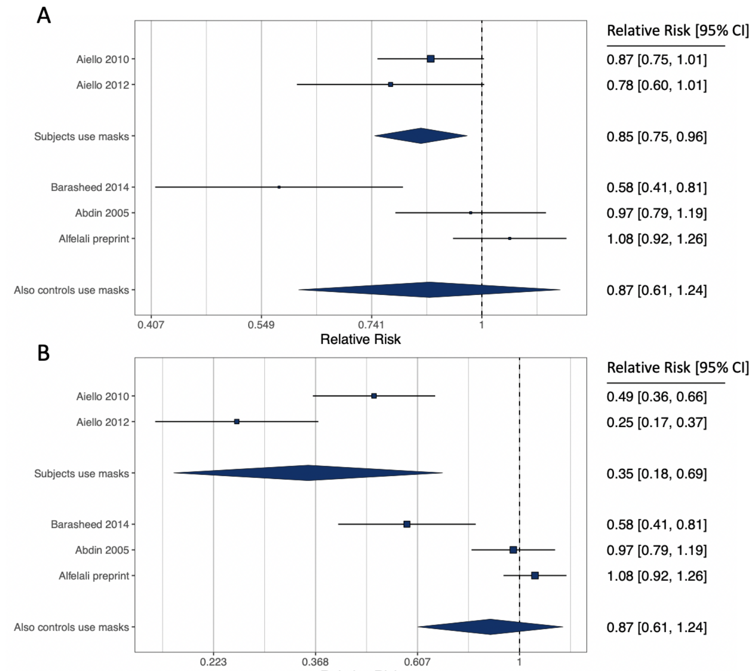

Figure 4: Stratified analysis.

Notes: Stratified analysis of the studies where face masks are used in the study population show a statistically significant protective effect. a) Cumulative time point b) maximum follow-up time point.

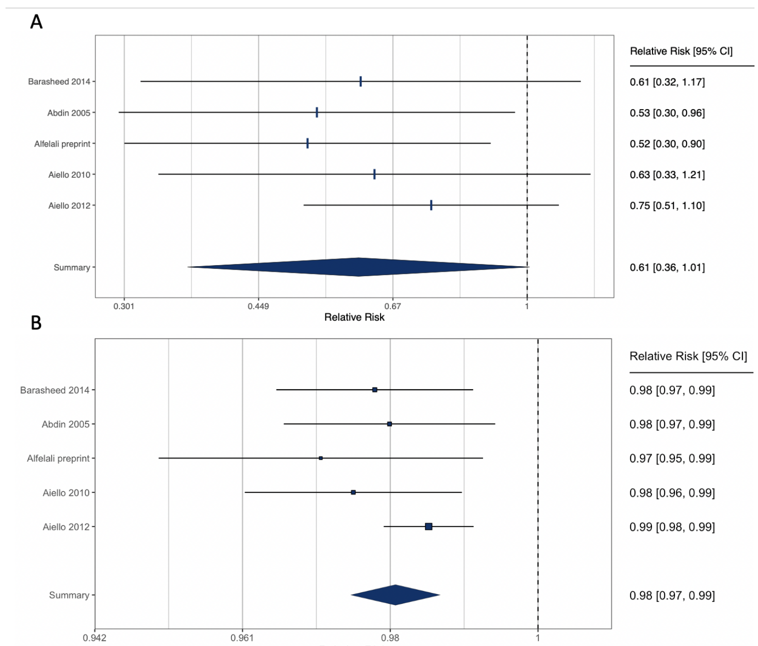

Figure 5: Leave one out -analyses.

Notes: Leave one out -analyses under a) unadjusted random effects raw values show directionality of mean under $RR < 1$, and b) after accounting for use in face masks in controls adjusted meta-regressions show a robust, protective effect of face masks independently of which RCT study is omitted, suggesting that no single study drives the raw estimates to significance ($p < 0.05$ for all models).

4 Discussion

4.1 Implications

While there is growing evidence on if the universal masking in public settings would be a way to prevent respiratory diseases, only a few randomized-controlled trials exist and these separately have been interpreted to provide weak or no evidence of the effects of of face masks. We provide a meta-analysis of the five RCTs on the mask use in non-hospital and non-household settings and find a robust evidence of the effect of face masks on preventing respiratory infections. Four of the analyzed studies evaluated the use of masks on respiratory infections directly, and in one the primary outcome was compliance with mask use. We found a statistically significant protective effect of face masks preventing respiratory infections with pooled effect across these studies. Our sensitivity and sub-group analyses support the claim that masks have a concrete impact on curtailing infections in the general population.

Three out of the five original studies concluded that face masks can have an impact in managing the spread of respiratory infections, while two of them concluded that there was no clear evidence for the use of face masks [15, 16, 17, 18, 19]. Our study with a formal meta-analysis and a meta-regression reaches the conclusion that masks can have an impact on managing spread of respiratory infections such as COVID-19 also in an RCT setting.

While medical masks have been the gold standard to prevent doctor-to-patient or patient-to-doctor -transmitted infections, the current discussion on the use of face masks has focused on whether masks also protect the wearer, whether individuals who are not formally trained to use masks are able to wear them safely, and whether masks may represent an infection risk due to contamination effects or by causing respiratory problems.

Only a few RCT studies have assessed whether masks can prevent respiratory infections of the general public in non-hospital settings. Most studies have been conducted using case control settings or with masks with a strong filtering capacity. Furthermore, meta-analyses summarizing these pooled effects are numerous, including by far the largest meta-analysis of face masks [4]. Our study is aligned with these findings, similarly showing a strong protective effect of face masks. The estimated number needed to mask to prevent one infection has ranged from 3 (N95 masks) to 6 (face masks), and the number is even higher when infection risk is low to start with [20].

While masks have been shown to be effective in themselves, their impact and therefore efficacy is largest when combined with other protective measures [4]. Also in our study, the effect of masks was further accentuated when combined with complementary measures such as hand hygiene [16, 17]. Furthermore, other complementary measures for disease control, such as physical distance, have a large impact on the spread and the number of particles in the air and therefore also add to the effect of face masks.

Approximating the effect of masks on population health can be addressed by specific measures such as exploring the number needed to treat (NNT), e.g. how many individuals need to wear a mask to prevent one person from becoming infected. NNT ranged from 4.5 to 145 at the cumulative comparison. However, currently all the studied RCTs have been performed prior to the emergence of COVID-19.

Obviously, these NNTs are only approximations since the type of infection and the reproduction number R is differs between different viral infections. Similarly, if there are no active infections, NNT will be large: no infections can be prevented as no infections are present in the population.

With these limitations in mind, we calculated that for respiratory infections, NNT might range up to $NNT = 150$. To put this into context, let us presume that in a metropolitan area with a population of one million residents, and 30% of them use face masks. With NNT=150, this might prevent 2,000 infections. Such effect size is comparable to the NNT of pharmaceuticals. Importantly, for example, NNT for statin, one of the most widely prescribed drug, in primary prevention of major vascular events at low levels of a CVD risk (5-10% within 5 years) ranges from 67 to 170 and is of a similar scale [21].

4.2 Potential limitations

The limitations of the individual RCT studies concern the compliance with treatment. Overall non-compliance for face mask use in the individuals with face masks was approximately 20% across studies. Surprisingly, we saw non-compliance also in the controls. For example, in one study, over half of the controls, who were not supposed to wear masks, were wearing them at some point [18]. Unsurprisingly, the largest protective effect with face masks was seen in the studies where compliance in both treatment and control groups was high, and was also seen in the sensitivity analysis. This was especially evident in the college student cohorts but also in the study of pilgrims [15], where 76% of the treatment group used masks but only 12% of individuals in the control arm used masks ($p < 0.05$). We stress that even relatively heterogeneous studies, or studies where RCT has been biased through non-compliance, can be useful when pooled using carefully selected statistical methods, which account for study heterogeneity.

One possible concern is that masking could cause risk compensation so that mask wearers become more careless with other mitigation measures, making masks counterproductive.[22] However, the current evidence does not support this concern [23, 24]. Moreover, using data from the U.S., [25] found that mobility decreased in most settings when the policy of masks for employees started in a state. Mask mandates in Germany were associated first with a decreased mobility and with no change in the longer-run mobility [26]. Finally, a paper using the store-level data from Germany documents that the mask mandate was not associated with a change in the distance keeping around the experimenter [24].

4.3 Avenues for future research

Conducting an RCT during a global pandemic to study of the overall effects of face mask use for the general public does not seem very ethical. Thus, quasi-experimental and observational studies should be used to assess the overall impact of face masks on the COVID-19 infections.

After the pandemic, further face mask RCTs could be conducted, especially focusing on the factors associated with heterogeneity in compliance. This is especially important, given that COVID-19 has demonstrated that the epidemic impact can differ substantially by demographic group. Finally, with elderly being especially vulnerable to respiratory infections, the net impact of face masks should be evaluated in the elderly, given that masks can increase dyspnoea in elderly. We show that those studies where hand hygiene was assessed together with mask use, effects with multiplicative protective measures were the strongest. Our results strongly support the WHO guidelines that recommend the use of face masks together with physical distancing and hand hygiene as primary measures for controlling the spread of the COVID-19 virus.

5 Conclusions

Our meta-analysis using RCTs in non-hospital and non-household settings provides support for the public health policy of face mask use to limit the spread of COVID-19 and other infectious diseases. Our analysis suggests that face masks protect both the wearer and the people around them and are particularly useful when combined with other non-pharmaceutical interventions, such as physical distance and hand hygiene.

Recommendations and clear communication about the benefits of face masks should be given by policy makers to limit the number of infections and ultimately deaths and to control COVID-19 disease clusters, thus providing time for vaccine development.

Conflict of interest

The authors declare no conflict of interest.

References

[1] Miller, D., Martin, M., Harel, N. et al. Full genome viral sequences inform patterns of SARS-CoV-2 spread into and within Israel. *medRxiv*, 2020. doi: 10.1101/2020.05.21.20104521.

[2] Kai, D., Goldstein, G-P., Morgunov, A. et al. Universal masking is urgent in the COVID-19 pandemic: SEIR and agent based models, empirical validation, policy recommendations, 2020.

[3] Stutt, R., Retkute, R. Bradley, M. et al. A modelling framework to assess the likely effectiveness of facemasks in combination with lock-down in managing the COVID-19 pandemic. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 476(2238):20200376, 2020. doi: 10.1098/rspa.2020.0376.

[4] Chu, D., Akl, E., Duda, S. et al. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. *Lancet*, 2020. doi: 10.1016/S0140-6736(20)31142-9.

[5] Buonanno, G., Morawska, L. and Stabile, L. Quantitative assessment of the risk of airborne transmission of SARS-CoV-2 infection: prospective and retrospective applications. *medRxiv*, 2020. doi: 10.1101/2020.06.01.20118984.

[6] Miller, S., Nazaroff, W., Jimenez, J. et al. Transmission of SARS-CoV-2 by inhalation of respiratory aerosol in the skagit valley chorale superspreading event. *medRxiv*, 2020. doi: 10.1101/2020.06.15.20132027.

[7] Royal Society. Face masks and coverings for the general public: Behavioural knowledge, effectiveness of cloth coverings and public messaging. 2020.

[8] Yan, Y., Bayham, J, Fenichel, E. and Richter, A. Do face masks create a false sense of security? A COVID-19 dilemma. *medRxiv*, 2020. doi: 10.1101/2020.05.23.20111302.

[9] Brainard, J., Jones, N., Lake, I. et al. Facemasks and similar barriers to prevent respiratory illness such as COVID-19: A rapid systematic review. *medRxiv*, 2020. doi: 10.1101/2020.04.01.20049528.

[10] Ministry of Social Affairs and Health Finland (MSAH) and Summaryx Oy. Selvitys väestön kasvosuojusten käytöstä COVID-19-epidemian leviämisen ehkäisyssä. `http://urn.fi/URN:ISBN:978-952-00-5421-2`, 2020. In Finnish.

[11] Hoertel, N., Blachier, M., Blanco, C. et al. A stochastic agent-based model of the SARS-CoV-2 epidemic in france. *Nat Med*, 2020. doi: 10.1038/s41591-020-1001-6.

[12] Howard J., Li, Z., Tufekci, Z. et al. Face masks against COVID-19: An evidence review. *Preprints*, 2020. doi: 10.20944/preprints202004.0203.v1.

[13] Abaluck J., Chevalier, J., Christakis, N. et al. The case for universal cloth mask adoption and policies to increase supply of medical masks for health workers. *SSRN*, 2020. doi: 10.2139/ssrn.3567438.

[14] Miyazawa, D. and Kaneko, G. Face mask wearing rate predicts country's COVID-19 death rates: with supplementary state-by-state data in the united states. *medRxiv*, 2020. doi: 10.1101/2020.06.22.20137745.

[15] Barasheed, O., Almasri, N., Badahdah, A-M. et al. Pilot randomised controlled trial to test effectiveness of facemasks in preventing influenza-like illness transmission among australian hajj pilgrims in 2011. *Infectious Disorders - Drug Targets*, 14:110–116, 03 2014. doi: 10.2174/1871526514666141021112855.

[16] Aiello, A., Murray, G., Perez, V. et al. Mask use, hand hygiene, and seasonal influenza-like illness among young adults: A randomized intervention trial. *The Journal of Infectious Diseases*, 201(4):491–498, 02 2010. doi: 10.1086/650396.

[17] Aiello, A., Perez, V., Coulborn, R. et al. Facemasks, hand hygiene, and influenza among young adults: A randomized intervention trial. *PLOS ONE*, 7(1):1–8, 01 2012. doi: 10.1371/journal.pone.0029744.

[18] Alfelali, M., Haworth, E., Barasheed, O. et al. Facemask versus no facemask in preventing viral respiratory infections during hajj: A cluster randomised open label trial. *SSRN Electronic Journal*, 01 2019. doi: 10.2139/ssrn.3349234.

[19] Abdin, EZ., Choudhry AJ., and Al-Naji A. A effect of use of face mask on hajj-related respiratory infection among hajjis from riyhad. a health promotion intervention study. *Saudi Epidemiology Bulletin*, pages 27–28, 01 2005.

[20] Jefferson, T., Foxlee, R., Del Mar, C., et al. Physical interventions to interrupt or reduce the spread of respiratory viruses: systematic review. *BMJ*, 336:77–80, 2020. doi: 10.1136/bmj.39393.510347.BE.

[21] Taylor, F., Huffman, MD., Macedo, AF., et al. Statins for the primary prevention of cardiovascular disease. *Cochrane Database Syst Rev.*, 2013. doi: 10.1002/14651858.CD004816.pub5.

[22] Lazzarino, A. I., Steptoe, A., Hamer, M., and Michie, S. Covid-19: Important potential side effects of wearing face masks that we should bear in mind. *BMJ*, 369, 2020. doi: 10.1136/bmj.m2003.

[23] Rubin G. J. Mantzari, E. and T. M. Marteau. Is risk compensation threatening public health in the covid-19 pandemic? *BMJ*, 370, 2020. doi: 10.1136/bmj.m2913.

[24] Balleyer A. H. Cerutti N. Friedrichsen J. Seres, G. and M. Süer. Face mask use and physical distancing before and after mandatory masking: Evidence from public waiting lines. *SSRN Electronic Journal*, 06 2020.

[25] Kasahara H. Chernozhukov, V. and P. Schrimpf. Causal impact of masks, policies, behavior on early covid-19 pandemic in the u.s. *medRxiv*, 2020. doi: 10.1101/2020.05.27.20115139.

[26] Dunaiski M. Kovacs, R. and J. Tukiainen. Compulsory face mask policies do not affect community mobility in germany. *SSRN Electronic Journal*, 06 2020. doi: 10.2139/ssrn.3620070.

Supplement

A Flow chart

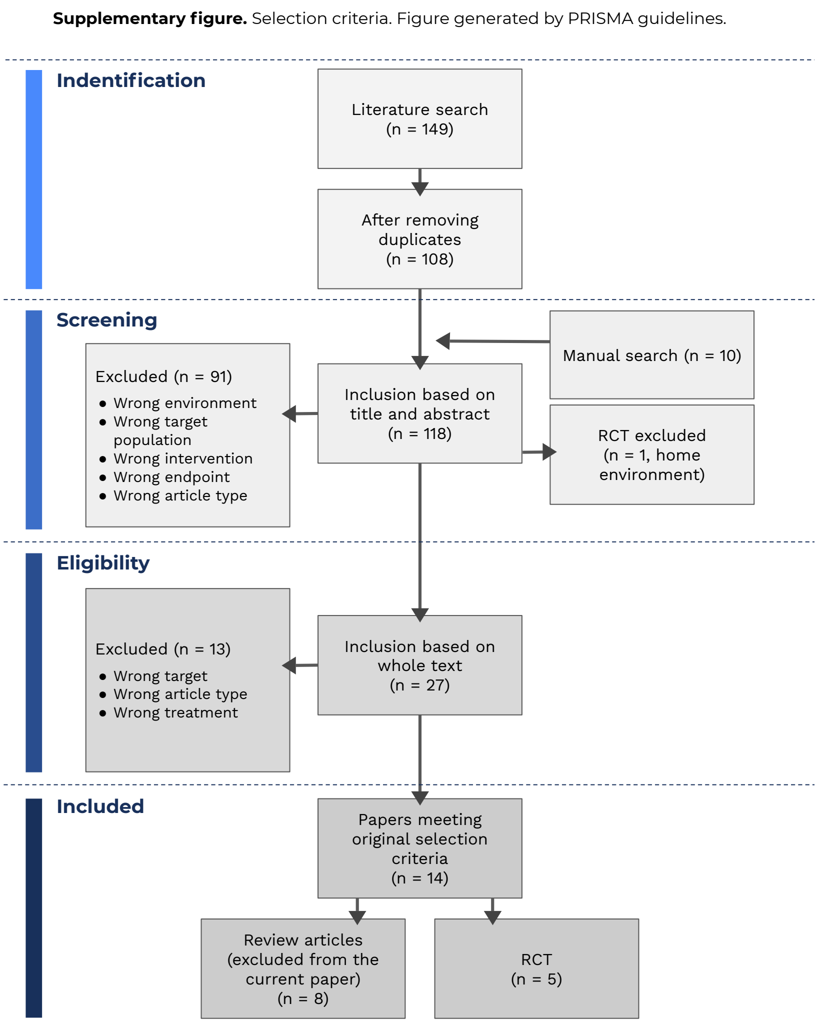

Figure A1: Flow chart.

Notes: Original search strategy has been described in [10]

B Sensitivity Analysis: Tables

Table A1: Characteristics of included studies

Publication	N total	Median age	Study group	Comparison	Endpoints	Gender	Mask use (%)	Mask use in controls	Mask use in cases (%)
Barasheed 2014	164	48	Pilgrim	Mask vs. control	At the end	51.7% male	12%	yes	76%
Abdin 2005	995	35	Pilgrim	Mask vs. control	At the end	57% male	33.60%	yes	81.30%
Alfelali preprint	7687	37	Pilgrim	Mask vs. control	At the end	43% male	53%	yes	25+48% (21% did not)
Aiello 2010	1297	19	Student	Mask and hand hygiene vs. control	Cumulative Week 6	51% male cases, 18% male controls	0 0	no no	3.1h per day
				Mask vs. control	Cumulative Week 6	39% male cases, 18% male controls	0 0	no no	3.9h per day
Aiello 2012	1188	19	Student	Mask and hand hygiene vs. control	Cumulative Week 6	48% male cases, 44% male controls	0 0	no no	5.1h per day
				Mask vs. control	Cumulative Week 6	42% male cases, 44% male controls	0 0	no no	5h per day

Table A2: Number needed to treat (NNT)

Publication	Population	Comparison	NNT
Barasheed 2014	Pilgrims	Mask vs control	4.489
Abdin 2005	Pilgrims	Mask and education vs control	145.929
Alfelali preprint	Pilgrims	Mask vs control	NA
Aiello 2010	Students	Mask vs control	11
Aiello 2010	Students	Mask and hygiene vs. control	12
Aiello 2012	Students	Mask vs control	NA
Aiello 2012	Students	Mask and hygiene vs. control	20

Only computed if RR < 1

Table A3: Relative risk per group

Publication	Population	Comparison	RR	RRlower	RRupper	logRR	seRR	Adjustment**
Abdin 2005	Pilgrims	Mask and education vs control	0.970	0.790	1.200	-0.0132	0.0010	Raw value
Alfelali preprint	Pilgrims	Mask vs control	1.079	0.935	1.244	0.0330	0.0008	adjusted
Aiello 2010	Students	Mask vs control	0.900	0.770	1.050	-0.0458	0.0007	adjusted
Aiello 2012	Students	Mask vs control	1.100	0.880	1.380	0.0414	0.0013	adjusted
Aiello 2012	Students	Mask vs control	1.020	0.460	2.250	0.0086	0.0046	adjusted
Aiello 2010	Students	Mask vs control	0.580	0.340	1.000	-0.2366	0.0017	adjusted
Barasheed 2014	Pilgrims	Mask vs control	0.576	0.332	1.007	-0.2396	0.0017	Raw value
Aiello 2010	Students	Mask vs control	0.920	0.790	1.060	-0.0362	0.0007	Raw value
Aiello 2012	Students	Mask vs control	1.080	0.860	1.340	0.0334	0.0012	Raw value
Aiello 2010	Students	Mask vs control	0.570	0.340	0.970	-0.2441	0.0016	Raw value
Aiello 2012	Students	Mask vs control	1.140	0.540	2.250	0.0569	0.0044	Raw value
Aiello 2010	Students	Mask and hygiene vs. control	0.870	0.730	1.020	-0.0605	0.0007	adjusted
Aiello 2012	Students	Mask and hygiene vs. control	0.780	0.570	1.080	-0.1079	0.0013	adjusted
Aiello 2012	Students	Mask and hygiene vs. control	0.250	0.070	0.870	-0.6021	0.0020	adjusted
Aiello 2010	Students	Mask and hygiene vs. control	0.490	0.270	0.870	-0.3098	0.0015	adjusted
Aiello 2012	Students	Mask and hygiene vs. control	0.780	0.590	1.050	-0.1079	0.0012	Raw value
Aiello 2010	Students	Mask and hygiene vs. control	0.880	0.750	1.030	-0.0555	0.0007	Raw value
Aiello 2012	Students	Mask and hygiene vs. control	0.300	0.090	0.980	-0.5229	0.0023	Raw value
Aiello 2010	Students	Mask and hygiene vs. control	0.510	0.300	0.900	-0.2924	0.0015	Raw value

***Value from adjusted vs. unadjusted analysses in the original study*

Table A4: Model results

	Comparison	Model	Estimate	Ci lower	Ci upper	Z	p-value
Cumulative endpoint	Mask with Hand Hygiene	FE	-0.0934	-0.1787	-0.0081	-2.15	0.0318
	Mask with Hand Hygiene	REML	-0.1458	-0.3292	0.0377	-1.56	0.1194
	Mask with Hand Hygiene	DL	-0.1413	-0.3107	0.0281	-1.63	0.1022
	Mask only	REML	-0.0442	-0.1282	0.0399	-1.03	0.3032
	Mask only	FE	-0.0442	-0.1282	0.0399	-1.03	0.3032
	Mask only	DL	-0.07	-0.2319	0.092	-0.85	0.3972
	All data points	DL	-0.0995	-0.2203	0.0212	-1.62	0.1063
	All data points	FE	-0.0816	-0.1516	-0.0117	-2.29	0.0221
	All data points	REML	-0.1009	-0.2262	0.0245	-1.58	0.1147
Maximum endpoint	Mask with Hand Hygiene	FE	-0.2067	-0.311	-0.1024	-3.88	0.0001
	Mask with Hand Hygiene	REML	-0.5023	-1.0114	0.0069	-1.93	0.0532
	Mask with Hand Hygiene	DL	-0.4972	-0.9496	-0.0449	-2.15	0.0312
	Mask only	FE	-0.0868	-0.1953	0.0217	-1.57	0.1167
	Mask only	REML	-0.2129	-0.5079	0.0822	-1.41	0.1573
	Mask only	DL	-0.2112	-0.4929	0.0704	-1.47	0.1416
	All data points	FE	-0.2343	-0.3331	-0.1354	-4.64	<0.0001
	All data points	REML	-0.458	-0.8438	-0.0721	-2.33	0.02
	All data points	DL	-0.4578	-0.8282	-0.0874	-2.42	0.0154

Table A5: Meta-regression Random effects

Endpoint	p-value	Heterogeneity p-value
Cumulative masks only adjusted for mask use in controls (yes vs. no)	0.6127	0.0049
Cumulative masks only adjusted for mask use in controls %	0.573	0.015
Cumulative masks and hand hygiene adjusted for mask use in controls (yes vs. no)	0.7885	0.0091
Cumulative masks and hand hygiene adjusted for mask use in controls %	0.0781	0.0882
Cumulative all groups, adjusted for Aiello et al 2010 and 2012 comparison, and mask use in controls (%)	0.9112	0.0055
Cumulative all groups, adjusted for Aiello et al 2010 and 2012 comparison	0.8179	0.0047
Cumulative all groups, mask use in controls (%)	0.573	0.015
Maximum exposure masks only adjusted for mask use in controls (yes vs. no)	0.45	0.0063
Maximum exposure masks only adjusted for mask use in controls %	p < .0001	0.3388
Maximum exposure masks and hand hygiene adjusted for mask use in controls (yes vs. no)	0.0106	0.0004
Maximum exposure masks and hand hygiene adjusted for mask use in controls %	0.0006	0.0078
Maximum exposure all groups, adjusted for Aiello et al 2010 and 2012 comparison, and mask use in controls (%)	0.0002	0.3267
Maximum exposure all groups, adjusted for Aiello et al 2010 and 2012 comparison	0.5279	0.0003
Maximum exposure all groups, mask use in controls (%)	0.0001	0.3388

*Random effects p-values are shown

C Sensitivity Analysis: Random effects

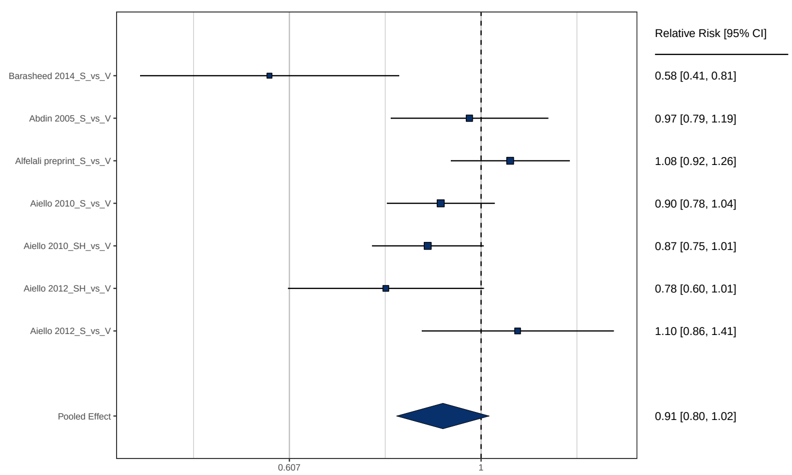

Figure A2: Random effects: Cumulative endpoint, combined groups.

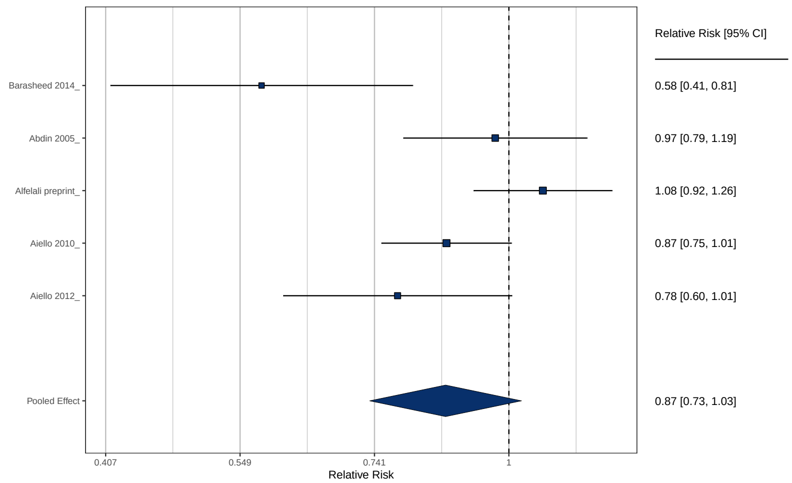

Figure A3: Random effects: Cumulative endpoint, mask and hand hygiene.

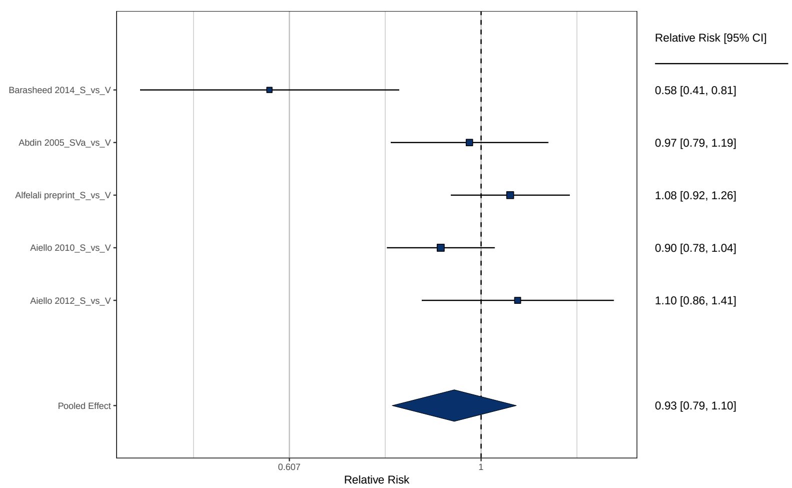

Figure A4: Random effects: Cumulative endpoint, masks only.

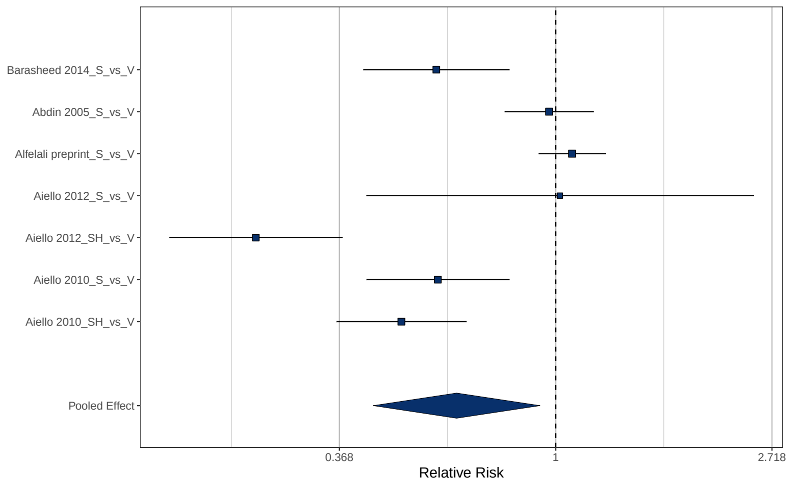

Figure A5: Random effects: Maximum endpoint, combined groups.

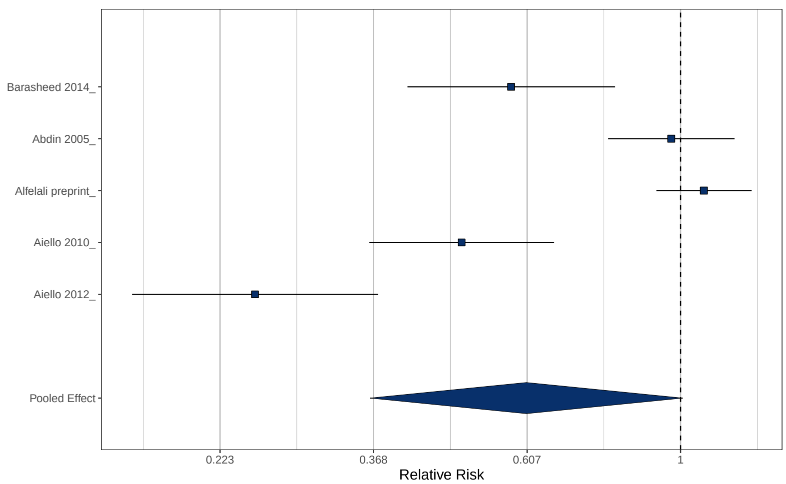

Figure A6: Random effects: Maximum endpoint, mask and hand hygiene.

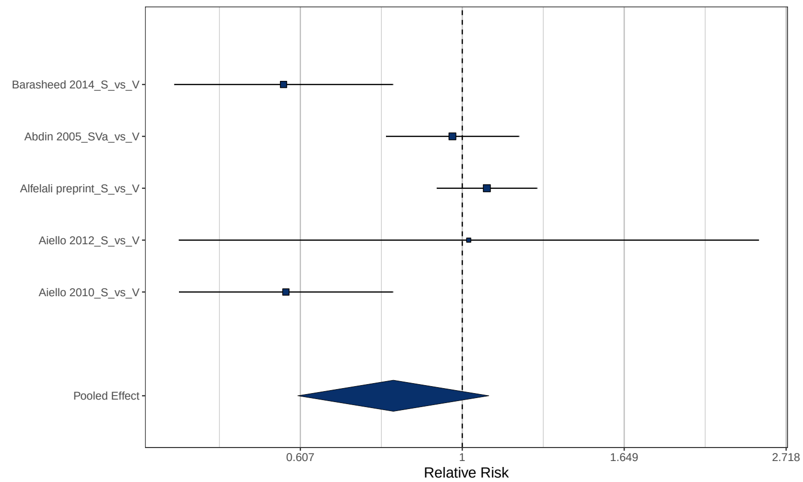

Figure A7: Random effects: Maximum endpoint, masks only.

D Sensitivity: Stratified analysis

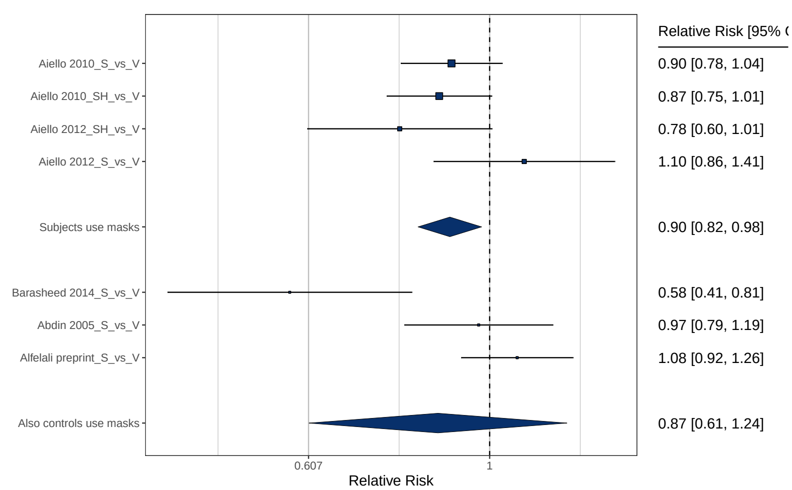

Figure A8: Stratified analysis: Cumulative endpoint, combined groups.

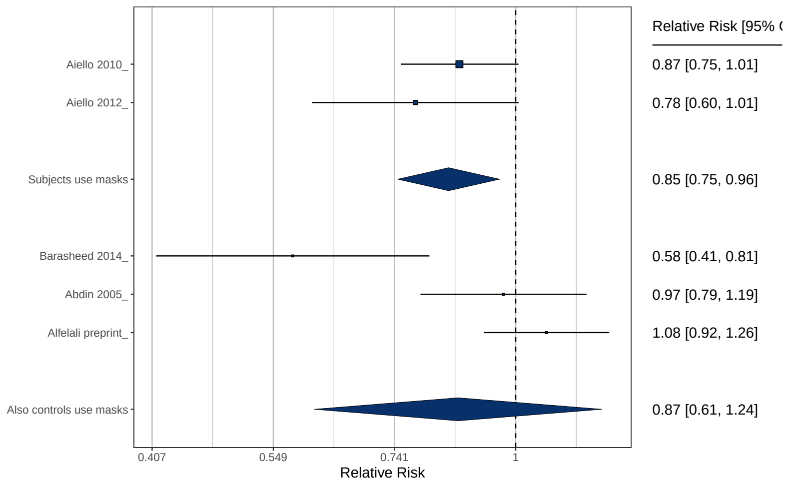

FIGURE A9: Stratified analysis: Cumulative endpoint, mask and hand hygiene.

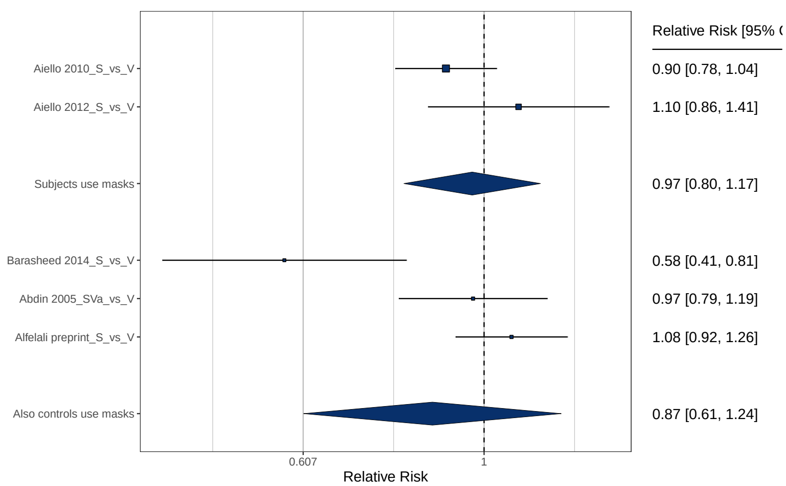

FIGURE A10: Stratified analysis: Cumulative endpoint, masks only.

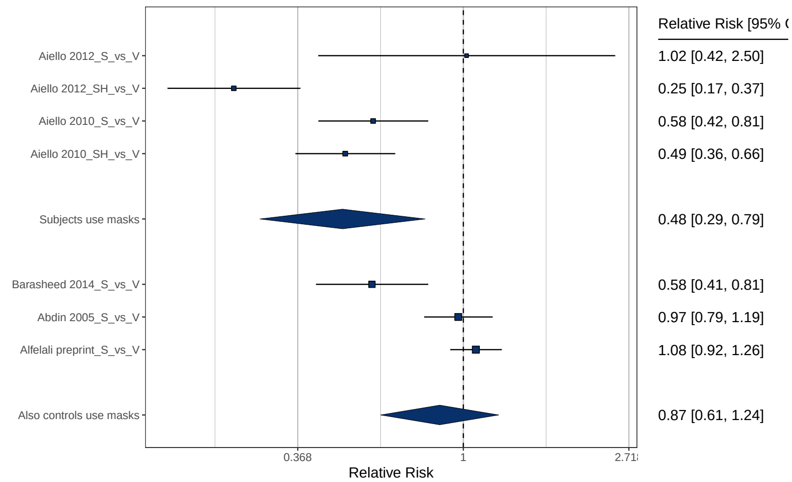

Figure A11: Stratified analysis: Maximum endpoint, combined groups.

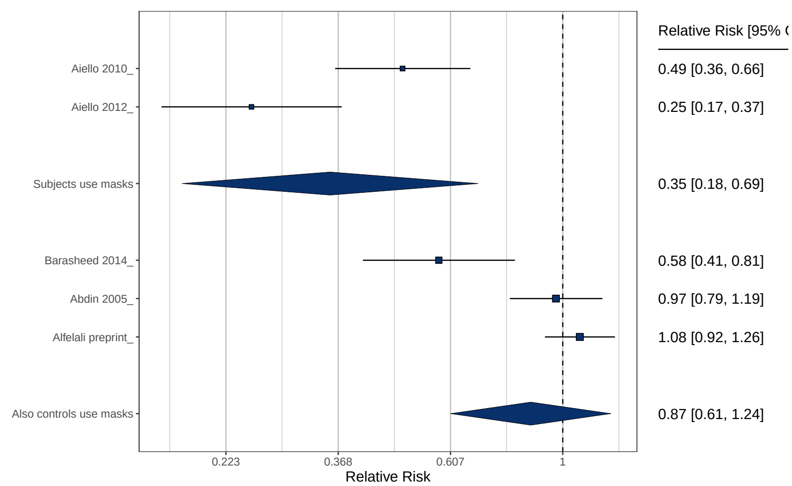

Figure A12: Stratified analysis: Maximum endpoint, mask and hand hygiene.

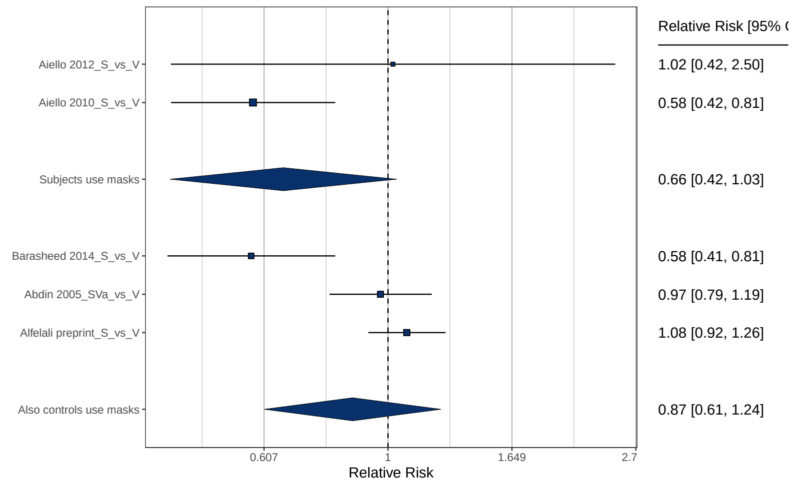

Figure A13: Stratified analysis: Maximum endpoint, masks only.

E Sensitivity analysis: Leave one out -analysis

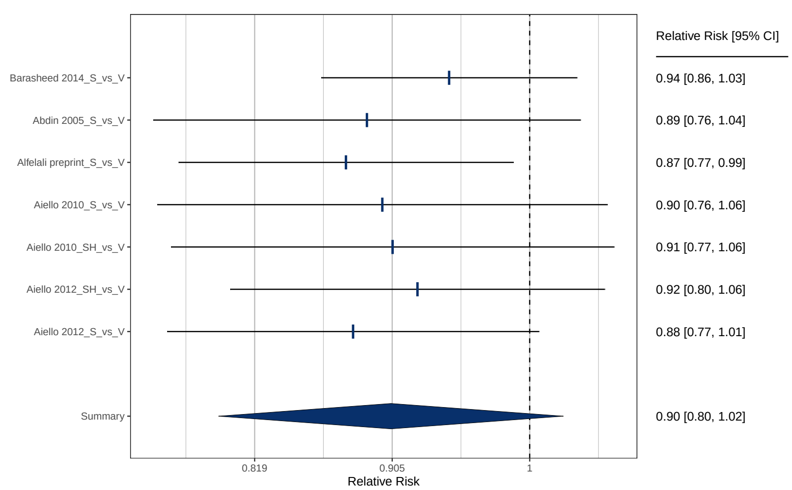

Figure A14: Leave one out -analysis: Cumulative endpoint, combined groups.

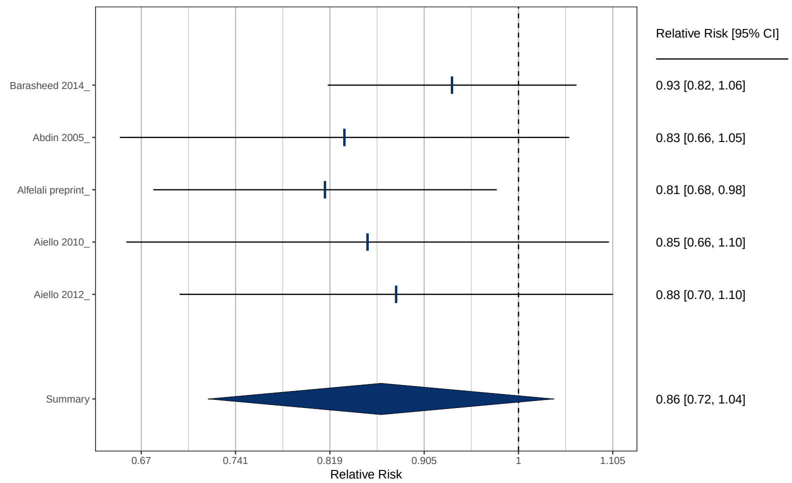

FIGURE A15: Leave one out -analysis: Cumulative endpoint, mask and hand hygiene.

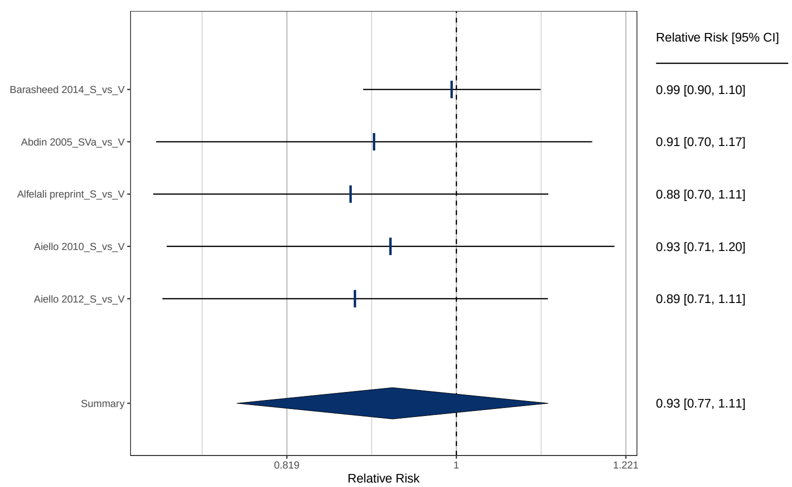

FIGURE A16: Leave one out -analysis: Cumulative endpoint, masks only.

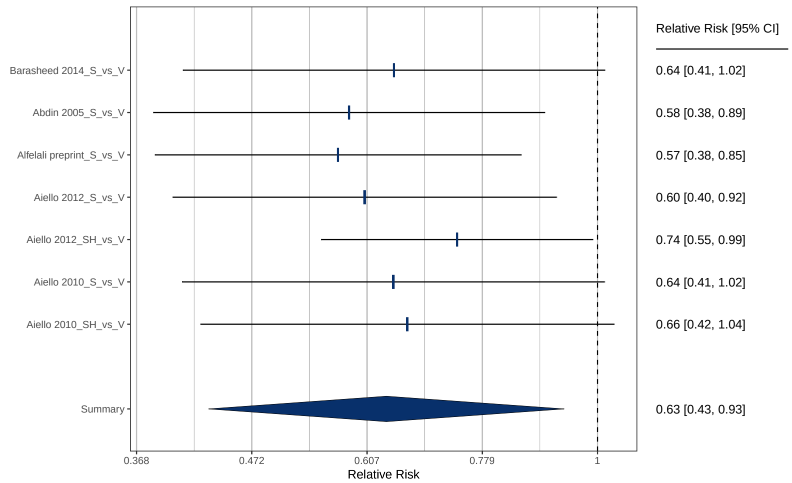

Figure A17: Leave one out -analysis: Maximum endpoint, combined groups.

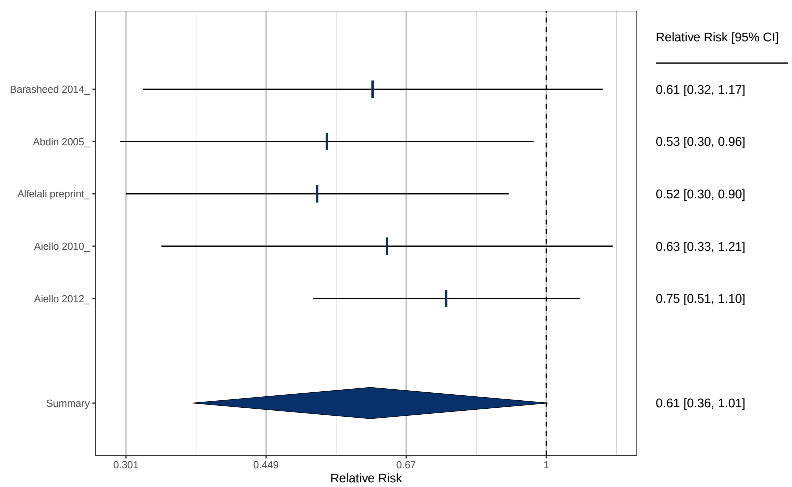

Figure A18: Leave one out -analysis: Maximum endpoint, mask and hand hygiene.

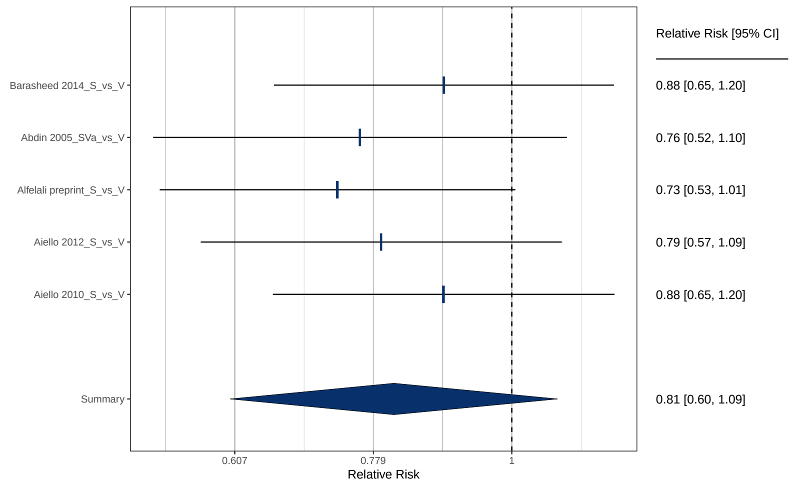

FIGURE A19: Leave one out -analysis: Maximum endpoint, masks only.

F Sensitivity: Meta regression leave one out

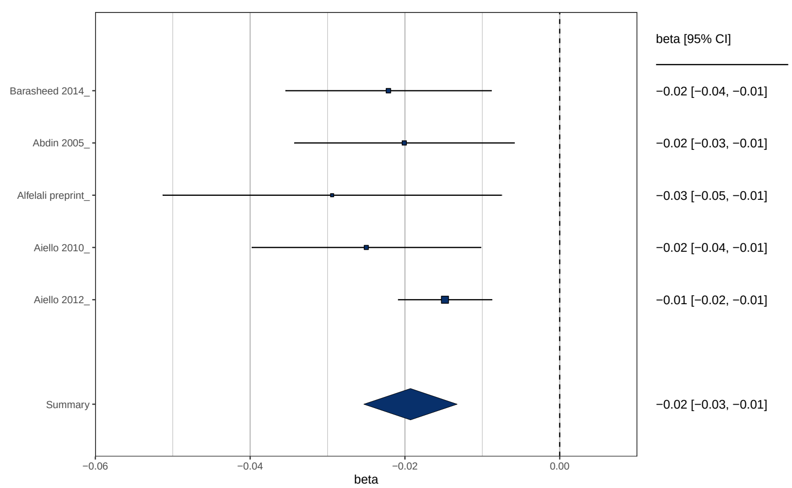

FIGURE A20: Meta regression leave one out: Maximum endpoint, masks and hand hygiene.

G Sensitivity analysis: Pooled analyses

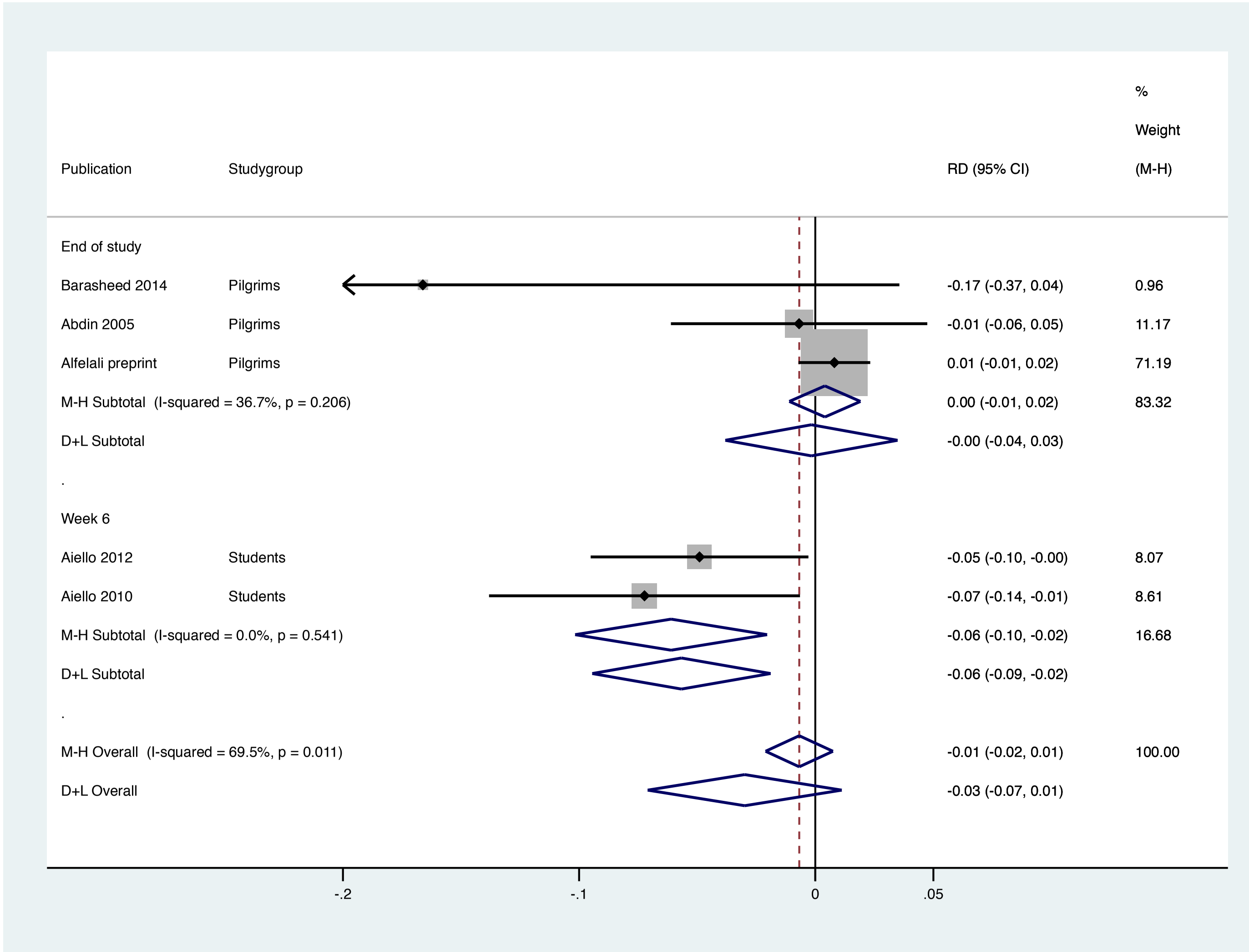

FIGURE A21: Global meta-analysis.

Notes: Figure summarizes our global meta-analysis. It shows that risk difference shows similar association as relative risk. There is statistically significant heterogeneity ($p = 0.01$). Pooled effect is not statistically significant. Subgroup analyses with Aiello 2010 and Aiello 2012 papers show a statistically significant association whereas other comparisons do not ($p > 0.05$).