Online Appendix: A Many-Analysts Approach to the Relation Between Religiosity and Well-being

The MARP Team*†

In this online appendix, we list reported effect sizes from the individual teams (1) whose effect sizes could not be converted into standardized β coefficients or (2) who reported multiple effect sizes for different independent and/or dependent variables. This document is structured as follows: Table 1 and 2 contain effect sizes for teams who reported one effect size per subscale of the well-being measure. The standardized β coefficients for the subscales are also visualized in Figure 1. For other teams, we provide the effect sizes in separate paragraphs. Details on the analyses can be found on the individual team pages, which can be accessed via osf.io/vy8z7.

Team 006. The team reported a negative association between religious service attendance and anxiety: -0.17, $[-0.23, -0.11]$, i.e., a positive association between service attendance and well-being operationalized as the absence of negative affect. In addition, anxiety is positively related to belief in God: 0.03, [0.00, 0.06]. Note that research question 2 was not investigated.

Team 014. Team 14 conducted a multiverse analysis and reported the percentages of models in which the relation between religion and well-being was significant. For research question 1, this was 100% of the models. For research question 2, 0% of the significant models included significant religion-by-cultural norms interactions.

Team 018. For research question 1, the team used Cohen's d as the effect size metric, that is, Cohen's $d = .161$, $[.122, .200]$. For research question 2, partial η^2 was

---

*All authors with their affiliations appear at the end of the main paper.
† Corresponding author: Suzanne Hoogeveen. E-mail: suzanne.j.hoogeveen@gmail.com.

Table 1

Research Question 1: Reported Effect sizes and 95% Confidence/Credible Intervals for Teams That Reported One Effect Size per Subscale.

	Well-being subscale				
Team	General	Psychological	Social	Physical	Effect size type
004	–	.145 [.125, .164]	.087 [.067, .108]	−.011 [−.031, .009]	β
010	–	.100 [.081, .118]	.057 [.038, .076]	.001 [−.018, .020]	β
023	–	.156 [.137, .176]	.081 [.06, .101]	.017 [−.004, .038]	β
035	–	.162 [.144, .180]	.083 [.065, .102]	.012 [−.006, .031]	β
065	–	.165 [.147, .183]	.103 [.085, .121]	−.017 [−.035, .003]	β
070	–	.144 [.115, .174]	–	.01 [−.01, .03]	β
071	–	.101 [.064, .138]	.025 [−.012, .062]	.004 [−.029, .036]	β
072	–	.185 [.155, .195]	.11 [.09, .13]	.01 [−.01, .03]	β
074	–	.14 [.11, .17]	.08 [.05, .11]	.02 [.00, .04]	β
086	–	.162 [.145, .180]	.104 [.086, .122]	−.009 [−.028, .009]	β
109	–	.199 [.170, .228]	.112 [.082, .141]	.058 [.029, .087]	β
123	.061 [.045, .078]	.151 [.134, .168]	.076 [.058, .094]	.024 [.006, .042]	β
129	–	.146 [.126, .165]	.074 [.053, .095]	.016 [−.005, .036]	β
143	–	.191 [.148, .234]	.123 [.072, .174]	.042 [.010, .073]	β
158	–	.162 [.144, .179]	.099 [.080, .119]	−.010 [−.029, .008]	β
164	.172 [.131, .213]	.329 [.288, .369]	.237 [.196, .278]	.053 [.012, .094]	β
168	–	.169 [.129, .217]	.093 [.054, .143]	.041 [.005, .082]	β
179	.105 [.066, .144]	.177 [.132, .222]	.108 [.059, .158]	.082 [.040, .124]	β
181	–	.068 [.042, .095]	–	.008 [−.010, .026]	β
058	–	.305 [.348, .392]	.212 [.255, .299]	−.005 [−.038, .082]	Cohen's d
076	–	.164 [.142, .185]	.081 [.059, .102]	−.001 [−.022, .021]	Cohen's d
141	–	.345 [.310, .452]	.235 [.197, .279]	.043 [.003, .084]	Cohen's d

calculated from the reported F-value, that is $\eta_p^2 = .001$ [.000, .002].

Team 021. In the main report, *religious behavior* was used as the independent variable for research question 1 ($\beta = .068$ [.042, .095]). In addition, the team reported an effect size for *religious beliefs* as the independent variable: $\beta = .020$ [−.006, .046]. For research question 2, the team used the interaction between religious behavior and cultural norms as the independent variable ($\beta = 0.032$ [0.004, 0.060]). In addition, they reported an effect size for the interaction between religious beliefs and cultural norms: $\beta = .011$ [−.016, .038].

Team 030. For research question 1, the team reported η^2 as their effect size metric, that is, $\eta^2 = .013$ [.009, .017]. The reported β for research question 2 is shown in the main article text.

Table 2

Research Question 2. Reported Effect Sizes and 95% Confidence/Credible Intervals for Teams That Reported One Effect Size per Subscale.

	Well-being subscale				
Team	General	Psychological	Social	Physical	Effect size type
004	–	.046 [.029, .062]	.040 [.023, .058]	.009[−.008, .027]	β
010	–	.032 [.013, .051]	.020 [.000, .039]	−.007 [−.025, .013]	β
023	–	.046 [.004, .089]	.021 [−.021, .063]	.034 [−.005, .073]	β
035	–	.057 [.039, .076]	.023[.004, .042]	.036 [.017, .055]	β
058	–	.058 [.026, .090]	.045 [.012, .078]	.014 [−.020, .048]	β
065	–	.094 [.058, .130]	.095[.057, .132]	.020 [−.017, .057]	β
070	–	.035 [.017, .053]	–	.02 [.01, .04]	β
071	–	−.006 [−.032, .02]	−.011 [−.031, .009]	−.016 [−.042, .009]	β
072	–	.017 [.009, .025]	.02 [.01, .03]	.007 [−.001, 015]	β
074	–	.02 [.01, .04]	.02 [.00, .03]	.01 [−.01, .02]	β
076	–	.038 [.020, .056]	.011 [−.009, .028]	.000 [−.019, .018]	β
086	–	.049 [.031, .068]	.041 [.022, .060]	.008[−.011, .027]	β
109	–	.051 [.031, .071]	.027 [.007, .048]	.035 [.014, .056]	β
123	−.003 [−.018, .012]	.020 [.005, .035]	.006 [−.01, .022]	.007[−.009, .023]	β
129	–	.052 [.034, .069]	.043 [.025, .061]	.036 [.018, .054]	β
143	–	.040 [.019, .061]	.025 [.003, .046]	.026 [.005, .046]	β
158	–	.047 [.029, .065]	.051 [.031, .071]	.008 [−.011, .027]	β
164	−.029 [−.088, .030]	−.021 [−.079, .037]	.000 [−.059, .059]	−.030 [−.090, .029]	β
168	–	.070 [.033, .115]	.052 [.018, .100]	.065 [.035, .108]	β
179	−.001 [−.018, .017]	.010 [−.007, .027]	−.002 [−.019, .016]	.015 [−.002, .032]	β
181	–	.031 [.017, .044]	–	.011 [−.014, .037]	β

Team 033. For research question 1, the team calculated Cohen's d based on the t-values for the *religious* vs. *atheist* contrast, that is, Cohen's $d = 0.059$ [0.02, 0.097]. For research question 2, the team calculated Cohen's d based on the t-value for the the *religious* vs. *atheist* contrast and the interaction with cultural norms of religious lifestyle and with cultural norms of religious beliefs (i.e., the three-way interaction): Cohen's $d = 0.051$ [0.013, 0.089].

Team 038. For research question 1, the team reported η^2 as the effect size metric, that is, $\eta^2 = .012$ [.008, .017]. For research question 2, a network analysis was performed which suggested that the perceived cultural norms of religion do not significantly affect the relationship between religiosity and well-being. That is, the global connectivity of the

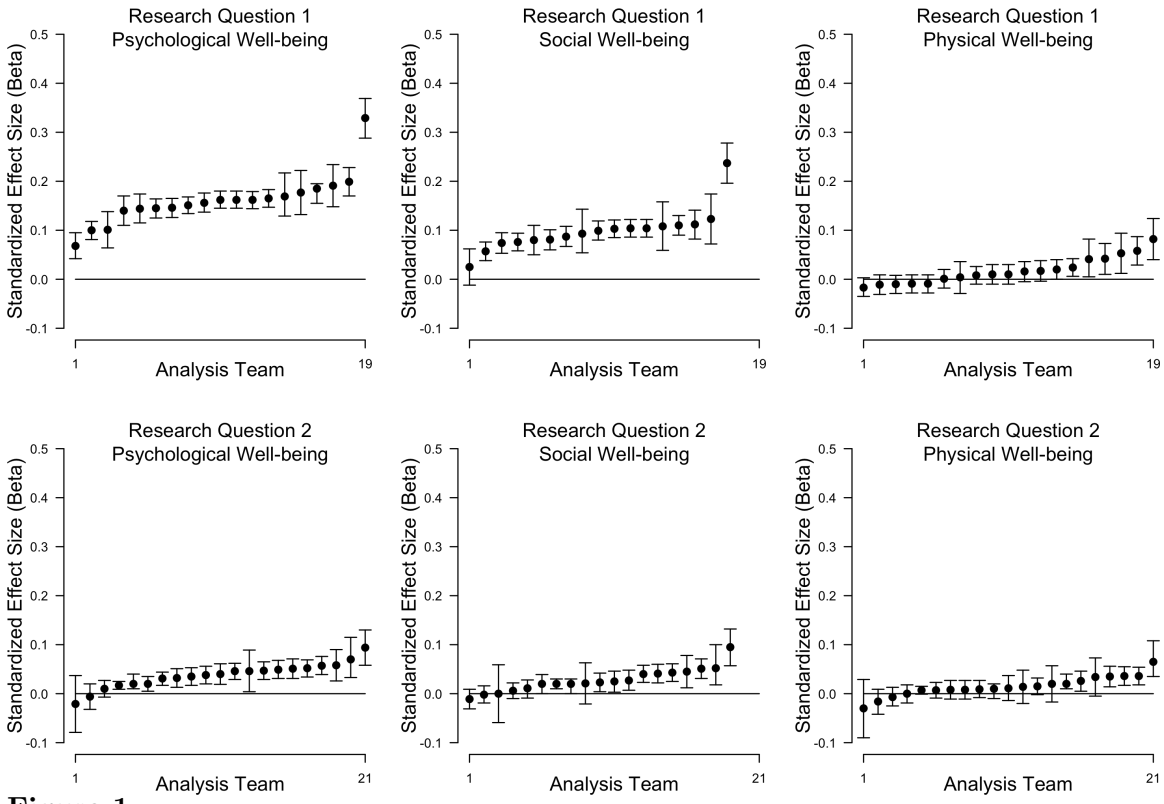

Figure 1

Standardized beta coefficients for the effect of religiosity on self-reported well-being (research question 1) and the moderating effect of cultural norms of religion on the association between religiosity and well-being (research question 2) for each of the three different well-being subscales as reported by some of the teams. The betas are ordered from smallest to largest and displayed with 95% confidence or credible intervals.

variables in the group with lower levels of perceived cultural norms was not significantly different ($S = .28,\ p = .32$) than its higher counterpart.

Team 047. For research question 2, in the main report, the effect size for the interaction between religiosity and *cultural norms of a religious lifestyle* was reported, that is, $\beta = 0.038$ [0.020, 0.056]. In addition, the team reported the effect size for the interaction between religiosity and *cultural norms of belief in God*: $\beta = 0.030$ [0.012, 0.049].

Team 082. In the main report, for research question 1, the effect size for the relation between religiosity and *satisfaction with life domains* was reported (i.e., $\beta =$ 0.148 [0.129, 0.167]). In addition, the team reported the effect size for *life enjoyment*

as the dependent variable: $\beta = .135$ [.116, .154]. For research question 2, s*atisfaction with life domains* was again used as the dependent variable in the main report (i.e., $\beta = 0.051$ [0.033, 0.069]). In addition, for *life enjoyment* as the dependent variable the reported effect size was $\beta = .024$ [.005, .042].

Team 087. The team reported non-standardized gamma-weights as the effect sizes metric. For research question 1, the effect size for within-country differences in religiosity was .002 [−.052, .054], $p = .95$. The weight for the between-country differences in religiosity was .178 [−.387, .742], $p = .52$. Individual differences within a country in religiosity did not interact at a statistically significant level with individual differences within a country in perceptions of country-wide religiosity to account for well-being; within-country differences in religiosity: .125 [−.049, .299], $p = .16$ No evidence that differences between-countries in religiosity interacted at a statistically significant level with differences between countries in perceptions of country-wide religiosity to account for well-being: −0.050[−.383, .283], $p = .76$.

Team 088. For research question 1, the team reported a partial η^2 as effect size metric: $\eta_p^2 = .007$ [.004, .011]. For research question 2, the team reported two effect sizes that were converted into standardized β's; in the main report, the item on *cultural norms of belief in God* was used for the interaction term ($\beta = .029$ [−.005, .063]). In addition, the effect size for the interaction with *cultural norms of a religious lifestyle* was reported: $\beta = .058$ [.031, .085].

Team 104. The team did not run their planned analyses as they concluded that measurement invariance was violated and therefore they could thus not get valid results.

Team 110. For research question 1, the team used ordinal logistic regression with religious status as the predictor of interest. Odds were converted to Cohen's d. The effect size for the comparison between *atheists* and *religious individuals* was $d = .249$ [.195, .302]. In addition, the comparison between *atheists* and *non-religious individuals* was reported: $d = .116$ [.065, .168].

Team 113. This team used machine learning to answer the research questions, which did not produce effect size measures. Instead, they concluded that in the random forest model, SES (Mean Decrease in Accuracy = 0.122) was more than four times as important in predicting overall well-being as any of the religiosity variables, which barely contributed in the prediction (Mean Decrease in Accuracy = 0.004 - 0.029). The Lasso regularized regression model confirmed this notion as SES was again the most important predictor ($\beta = .199$) and none of the religiosity items contributed at all ($\beta = 0$). No interaction effects between religiosity and cultural norms meaningfully contributed to the Lasso model, with three of them having a marginal influence ($\beta = .002 - .006$) and the rest having no influence at all ($\beta = 0$).

Team 117. For research question 1, the team provided the effect size for the comparison between *non-religious* and *religious individuals* in Cohen's d ($d = .193$ [.149, .218]). In addition, they reported the effect size for the comparison between *atheists* and *religious individuals*: $d = .237$ [.183, .292].

Team 136. In the main report for research question 2, the team randomly selected *cultural norms of a religious lifestyle* as the predictor for the interaction term; $\beta = .105$ [.065, .144]). In addition, they reported the effect size for the interaction with *cultural norms of belief in God*: $\beta = .028$ [−.003, .059].

Team 143. In the main report, effect sizes for overall well-being were reported. In addition to the effect sizes per subscale of the well-being measure (see Table 1 and 2), the team also reported an additional effect size for research question 2. That is, besides cultural norms of religiosity as predictor in the interaction, they also used the average religiosity value per country as a measure of cultural norms of religiosity: $\beta = .058$ [.017, .099].

Team 147. In the main report, the effect size for *religious beliefs* as the predictor of interest is reported ($\beta = .129$ [.102, .156]). In addition, the team provided an effect size for *religious lifestyle*: $\beta = .033$ [.006, .060] for research question 1. For research question 2, the effect size of the interaction term with *religious lifestyle* is $\beta = .018$ [.001, .035].

Team 162. For the main report, the team took their measure of religious centrality as the predictor of interest. In addition, they reported effect sizes for *religious practices* (β = .068 [.041, .095] and β = .010 [−.018, .038]), *self-identification* (β = −.024 [−.055, .007]), *belief in God* (β = .026 [−.008, .060]), *belief in an afterlife* (β = .049 [.000, .098]), and *spirituality* (β = .012 [−.019, .042]). For research question 2, effect sizes were reported for the interaction terms including these same religiosity measures: β = .077 [.026, .127]; β = .063 [−.006, .133]; β = .029 [−.028, .086]; β = .031 [−.033, .095]; β = .046 [−.024, .116]; β = .045 [−.009, .099].

Team 164. In the main report, the results for *general quality of life* as the dependent variable are given. In addition, for research question 1, the team reported effect sizes for *general satisfaction with health*, *psychological well-being*, *social well-being*, and *physical well-being* (see Table 1). For research question 2, *cultural norms of belief in God* was randomly chosen as the predictor for the interaction term in the main report. In addition to reporting effect sizes for the subscales of well-being (see Table 2), the team also reported results for the interaction with *cultural norms of a religious lifestyle*; for *general satisfaction with health*: β = .077 [.019, .135], for *psychological well-being*: β = .073 [−.079, .129], for *social well-being*: β = .084 [−.059, .141], and for *physical well-being*: β = .006 [−.052, .064].

Team 165. In the main report, results are given for *religious practices* as the predictor of interest. In addition, the team reported effect sizes for *religious beliefs*: β = .059 [.026, .092], as well as for the *religious* vs. *non-religious* self-identification comparison: β = .009 [−.043, .061] and the *religious* vs. *atheist* comparison: β = −.052 [−.128, .025]. For research question 2, these different predictors were also included as part of the interaction term with *cultural norms of religion*: β = .089 [.000, .177]; β = −.079 [−.177, .020]; $\beta-$ = .007 [−.103, .117].

Team 178. For research question 1, the team used a Spearman's correlation coefficient as effect size, that is, Spearman r = 0.106 [0.088, 0.125]. For research question 2, a Spearman partial correlation coefficient was reported, controlling for cultural norms of religiosity: Spearman r_p = 0.087 [0.067, 0.104].

Team 179. In the main report, the average of religiosity items $1-9$ was used as the independent variable and *overall well-being* as dependent variable for research question 1 ($\beta = .145$ [.096, 0.194]). The team conducted nine multilevel models, one for each indicator of well-being. In addition to effect sizes for the well-being subscales given in Table 1, they reported *medical well-being*: $\beta = -.070$ [$-.098$, $-.043$], *quality of life* $\beta = .098$ [.060, .136], *satisfaction with health* $\beta = .085$ [.050, .119], and *self-esteem* $\beta = .126$ [.089, .163].

For research question 2, the authors constructed six different models for each of the nine dependent variables. Here, we will report the results of the first model which evaluates *individual-level perceived cultural norms*. In the main report, the average of religiosity items $1-9$ was used as the independent variable and *overall well-being* as dependent variable for research question 2 ($\beta = .009$ [-0.008, 0.026]). Results for the well-being subscales are given in Table 2 and for the remaining indicators: *medical well-being* $\beta = -.035$ [$-.052$, $-.018$], *quality of life* $\beta = -.012$ [$-.029$, .005], *satisfaction with health* $\beta = .009$ [$-.009$, .027], and *self-esteem* $\beta = .013$ [$-.005$, .030].

Team 181. In the main report, the average of religiosity items $1-7$ was used as the independent variable and *overall well-being* as dependent variable for research question 1 ($\beta = .072$ [.036, .109]). In addition to the results reported in Table 1, the team used *quality of life* as the dependent variable ($\beta = 0.099$ [0.069, 0.129]). Note that in their models, the authors included the moderator *personal importance of religion* and set up three models. In addition, the team computed effect sizes for *religious attendance* as independent variable which will not be reported here.

For research question 2, the team set up an additive moderation model and a moderated moderation model in which *personal importance of religion* and *cultural importance of religion* were added as moderators. The effect sizes of the additive moderation model were for *overall well-being* (.016, [.005, .028]), *psychological well-being* ($\beta = .031$ [.017, .044]), *physical well-being* ($\beta = .011$ [$-.014$, .037]), and *quality of life* ($\beta = -.002$ [$-.017$, 0.013]). The effect sizes of the moderated moderation model were for *overall well-being* (.005 [$-.022, .023$]), *psychological well-being* ($\beta = .005$ [$-.022, .033$]),

physical well-being ($\beta = -.009$ [$-.035,\ .017$]), and *quality of life* ($\beta = .003$ [$-.028,\ .034$]).

Team 188. For research question 1, *religious beliefs* was used as the independent variable for ($\beta = .079$ [$.059,\ .100$]). For research question 2, the team used the interaction between religious beliefs and cultural norms as the independent variable. Importantly, the team divided cultural norms into three categories. Here, we report the interaction between religious beliefs and a high level of cultural norms, that is, $\beta = .007$ [$-.018,\ .031$]

Team 189. For research question 1, *religious status* was used as the independent variable for research question 1. The team reported log odds ratios as effect size metric which we converted to Cohen's d ($d = 0.18$ [$0.137,\ 0.223$]). For research question 2, a transformation to Cohen's d was not possible. We therefore report only the result of the model comparison, that is the comparison between the model that features the interaction to the model that does not feature the interaction, $\chi^2(8, 10136) = 13.94, p = 0.083$.