frontiers
in Psychology

METHODS
published: 11 May 2021
doi: 10.3389/fpsyg.2021.614236

A Similarity-Weighted Informative Prior Distribution for Bayesian Multiple Regression Models

*Christoph König**

Department of Educational Psychology, Institute of Psychology, Goethe University Frankfurt, Frankfurt, Germany

Specifying accurate informative prior distributions is a question of carefully selecting studies that comprise the body of comparable background knowledge. Psychological research, however, consists of studies that are being conducted under different circumstances, with different samples and varying instruments. Thus, results of previous studies are heterogeneous, and not all available results can and should contribute equally to an informative prior distribution. This implies a necessary weighting of background information based on the similarity of the previous studies to the focal study at hand. Current approaches to account for heterogeneity by weighting informative prior distributions, such as the power prior and the meta-analytic predictive prior are either not easily accessible or incomplete. To complicate matters further, in the context of Bayesian multiple regression models there are no methods available for quantifying the similarity of a given body of background knowledge to the focal study at hand. Consequently, the purpose of this study is threefold. We first present a novel method to combine the aforementioned sources of heterogeneity in the similarity measure ω. This method is based on a combination of a propensity-score approach to assess the similarity of samples with random- and mixed-effects meta-analytic models to quantify the heterogeneity in outcomes and study characteristics. Second, we show how to use the similarity measure ω as a weight for informative prior distributions for the substantial parameters (regression coefficients) in Bayesian multiple regression models. Third, we investigate the performance and the behavior of the similarity-weighted informative prior distribution in a comprehensive simulation study, where it is compared to the normalized power prior and the meta-analytic predictive prior. The similarity measure ω and the similarity-weighted informative prior distribution as the primary results of this study provide applied researchers with means to specify accurate informative prior distributions.

Keywords: informative prior distributions, prior information, heterogeneity, similarity, Bayesian multiple regression, comparability

OPEN ACCESS

Edited by:
Kwanghee Jung,
Texas Tech University, United States

Reviewed by:
Navid Feroze,
University of Azad Jammu and Kashmir, Pakistan
Kyungtae Kim,
Tennessee Department of Education, United States

***Correspondence:**
Christoph König
koenig@psych.uni-frankfurt.de

Specialty section:
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

Received: *05 October 2020*
Accepted: *13 April 2021*
Published: *11 May 2021*

Citation:
König C (2021) A Similarity-Weighted Informative Prior Distribution for Bayesian Multiple Regression Models. Front. Psychol. 12:614236. doi: 10.3389/fpsyg.2021.614236

INTRODUCTION

Informative prior distributions are a crucial element of Bayesian statistics, and play a pivotal role for scientific disciplines that aim at constructing a cumulative knowledge base. Informative prior distributions are background knowledge quantified and introduced in a Bayesian analysis. Their use allows studies to build upon each other, hence to update the knowledge base of a scientific discipline continuously. This is also a central tenet of the new statistics (Cumming, 2014). Despite the increase of Bayesian statistics in various scientific disciplines over the last years, the use of informative prior distributions is still relatively rare (for instance in Psychology, see van de Schoot et al., 2017; for Educational Science see König and van de Schoot, 2018). Thus, the potential of Bayesian statistics for cumulative science is not fully realized yet.

Goldstein (2006) states that the tentative use of informative prior distributions is due to their frequently criticized subjective nature. Vanpaemel (2011) adds the lack of methods to formalize background knowledge as another reason. From an applied viewpoint, this is more severe: if the background knowledge is inaccurate, which is the case if the prior mean does not equal the population mean, parameter estimates may be biased (McNeish, 2016; Finch and Miller, 2019). Specifying accurate informative prior distributions is a question of carefully selecting studies that comprise the body of comparable background knowledge. Psychological research, however, consists of studies that are being conducted under different circumstances, with different samples and varying instruments. Thus, results of previous studies include different sources of heterogeneity, and not all available results can and should contribute equally to an informative prior distribution (Zhang et al., 2017). This implies a necessary weighting of background information based on the similarity of the previous studies to the focal study at hand. Current approaches to account for heterogeneity by weighting informative prior distributions are either not easily accessible or incomplete. For example, the power prior weighs the likelihood of the data and requires complicated intermediate steps in order to use the quantified heterogeneity properly (Ibrahim et al., 2015; Carvalho and Ibrahim, 2020). The meta-analytic predictive prior (Neuenschwander et al., 2010) is more intuitive by weighting the informative prior distribution directly, but uses heterogeneity in outcomes only. To complicate matters further, to date there are no methods available for investigating and quantifying the similarity of a given body of background knowledge to the focal study at hand. Specifying accurate informative prior distributions, however, requires an approach that quantifies all sources of heterogeneity in a body of background knowledge into a measure of similarity, and using this measure to weight the associated informative prior distribution in a direct and intuitive way.

Consequently, the purpose of this study is threefold. We first present a novel method to combine the aforementioned sources of heterogeneity in the similarity measure ω. This method is based on a combination of a propensity-score approach to assess the similarity of samples with random- and mixed-effects meta-analytic models to quantify the heterogeneity in outcomes and study characteristics (e.g., Tipton, 2014; Cheung, 2015). Second, we show how to use the novel similarity measure ω as a weight for informative prior distributions for the substantial parameters (regression coefficients) in Bayesian multiple regression models. Third, we investigate the performance and the behavior of the similarity–weighted informative prior distribution in a comprehensive simulation study, where it is compared to the normalized power prior (Carvalho and Ibrahim, 2020) and the meta-analytic predictive prior (Weber et al., 2019). The similarity measure ω and the similarity-weighted informative prior distribution as the primary results of this study provide applied researchers with means to specify accurate informative prior distributions.

The structure of this paper is as follows. First, the conceptual background of similarity is illustrated. Next, it is shown how these sources of heterogeneity can be quantified and combined in the similarity measure ω. Based on this, the similarity-weighted informative prior distribution is described. The design and results of the simulation investigating the performance and behavior of this distribution is presented next, followed by a discussion of how the similarity measure ω and the similarity-weighted informative prior distribution contribute to building confidence in and to systemizing the use of informative prior distributions in Psychological research. Please note that, in order to keep the manuscript as accessible as possible, mathematical details are kept at a minimum.

CONCEPTUAL BACKGROUND

The Concept of Similarity

When specifying informative prior distributions, researchers are confronted with a body of background knowledge comprised of conceptual replications of studies (Schmidt, 2009). Conceptual replications focus on the general theoretical process, without copying the methods of previously conducted studies (Makel et al., 2012). Thus, the studies differ in samples, variables, and other characteristics. Without assessing their similarity to the focal study at hand, using studies for informative prior distributions might imply an unwarranted generalization; excluding studies might be too restrictive and imply that no background knowledge is available, when in truth there is. Hence, an adequate similarity measure should take into account all relevant sources of heterogeneity in research results. Consequently, the conceptual framework of the similarity measure ω follows Shadish et al. (2002), who build upon Cronbach (1982), and distinguishes between units and treatments (*UT*), outcomes (*O*), and settings (*S*) of the studies as sources for heterogeneity. More specifically, we conceptualize *UT* as samples and predictor variables, *O* as outcome variables or effect sizes, and *S* as study characteristics commonly investigated as moderators in mixed-effects meta-analytic models. Thus, we define similarity as the variability in research results due to the three sources of heterogeneity. This differentiation takes into account that heterogeneity in outcomes is not sufficient for an adequate assessment of similarity (Lin et al., 2017). The quantification of the three sources of heterogeneity is addressed next.

Quantifying Sources of Heterogeneity

For a similarity measure to work adequately, it is pivotal that the different sources of heterogeneity can be quantified accurately with state-of-the-art methods. More specifically, the similarity measure ω is based on three components:

(a) the modified generalizability index $\overline{B}$ that is based on Tipton (2014), (b) the between-study heterogeneity τ^2 resulting from (Bayesian) random-effects meta-analytic models, and (c) δ_{τ^2}, the difference between the residual variance τ^2_{res} of (Bayesian) mixed-effects meta-analytic models and τ^2 (for an overview see, for instance, Jak, 2015). Each individual measure quantifies important aspects of the comparability of research results.

Quantifying Similarity in Predictors and Samples With $\overline{B}$

The first component of the similarity measure ω is the modified generalizability index $\overline{B}$. In its original form, the generalizability index B is a propensity score-based measure of distributional similarity between a sample and a population (Tipton and Olsen, 2018). We modified it so that it describes the similarity between the samples of the focal study and a previously conducted study that is part of the body of available background knowledge. The generalizability index and its modified version takes values between zero and one, which indicate no and perfect similarity of the two samples, respectively. It is based on $s(\mathbf{X})$, a theoretical sampling propensity score defined as $s(\mathbf{X}) = \Pr(Z = 1|\mathbf{X})$, and describes the probability Z of an individual being in the sample of the focal study (vs. being in the sample of the previously conducted study) based on a set of covariates $\mathbf{X}$ (Tipton, 2014). The sampling propensity score can be estimated by a logistic regression model $\log[s(X)/1\text{-}s(X)] = \alpha_0+\alpha_m+X_m$, where $m = 1, , m$ is the number of covariates. Adapting Tipton (2014), for a set of covariates $\mathbf{X}$ and sampling propensity score $s(\mathbf{X})$, the modified generalizability index is then defined as $\beta = \int \sqrt{f_f(s)f_p(s)ds}$, where $f_f(s)$ and $f_p(s)$ are the distributions of sampling propensity scores in the sample of the focal and previously conducted study, respectively. An estimator of β is provided by a discrete version of the generalizability index $B = \sum_h \sqrt{w_{fh}w_{ph}}$, where h is the number of bins and w_{fh} and w_{ph} are the proportions of the focal and previously conducted study samples, respectively (Tipton, 2014). In case of multiple previously conducted studies, the modified version of the generalizability index B is calculated for each comparison of the samples of the focal and previously conducted studies. It is the average of the individual indices $\overline{B} = \frac{1}{k}\sum_k B_k$, with k being the number of previously conducted studies. We implemented this procedure as a kernel density estimation with a Gaussian kernel and a non-parametric bandwidth selector (Moss and Tveten, 2019), so that the number of bins does not have to be chosen a priori.

Quantifying Heterogeneity in Outcomes With τ^2

The second component of the similarity measure ω is the between-study heterogeneity τ^2, which is a measure for the variance in effect sizes, such as standardized mean differences, log-odds ratios, and more recently, partial and semi-partial correlations as effect sizes for regression coefficients (Aloe and Thompson, 2013). It is the variance component of random-effects meta-analytic models, which assume that the population effect sizes are not equal across the studies. Several studies show that this assumption is usually correct: the typical between-study heterogeneity in outcomes ranges from 0.13 to 0.24 (van Erp et al., 2017; Stanley et al., 2018; Kenny and Judd, 2019). Random-effects meta-analytic models allow individual studies to have their own effect (e.g., Cheung, 2015). Let y_k be the effect found in study k. The study-specific model is then $y_k = \overline{\beta} + u_k + \varepsilon_k$ where $\overline{\beta}$ is the average effect size, u_k are deviations from the average effect size, ε_k is the study-specific error term and $Var(\varepsilon_k)$ is the known sampling variance. The variance of these deviations $Var(u_k)$ is the between-study heterogeneity τ^2 indicating the variability of the effect sizes across the studies included in the meta-analysis. The between-study heterogeneity is strictly positive $\tau^2 > 0$. When τ^2 increases, consensus in the average effect decreases. This lack of consensus in the average effect, the uncertainty quantified by τ^2, should be represented in a weight of an informative prior distribution. However, only the meta-analytic predictive prior distribution uses τ^2 as weight. Both the average effect and the between-study heterogeneity τ^2 can be estimated by Maximum Likelihood, Restricted Maximum Likelihood and Bayesian estimation methods (for overviews, see Veroniki et al., 2016; Williams et al., 2018). For situations with a small number of studies, and the known problems of ML and REML estimators regarding τ^2 in these cases, we implemented a hierarchical Bayesian random-effects meta-analytic model to estimate $\tau^{\ 2}$ accurately.

Quantifying Heterogeneity in Study Characteristics with δ_{τ^2}

The third component of the similarity measure ω is δ_{τ^2}, the difference between the residual variance τ^2_{res} in the effect sizes, estimated by a (Bayesian) mixed-effects meta-analytic model, and their estimated between-study heterogeneity τ^2. Mixed-effects meta-analytic models extend random-effects meta-analytic models by introducing study characteristics as potential moderators of the effects. The study-specific model is then $y_k = \beta x_k + u_k + \varepsilon_k$, where $\mathbf{x}_k$ is a vector of predictors including a constant of one (Cheung, 2015). Under the mixed-effects meta-analytic model, the variance of the deviations $Var(u_k)$ is the residual variance τ^2_{res} in the effect sizes after controlling for study characteristics as moderators. If $\tau^2_{res} < \tau^2$, the study characteristics explain variance in the effect sizes. This implies that the effect sizes not only vary across studies, but also across specific study characteristics. For example, it is possible that effects found in the 1980s differ systematically from effects found in the 2010s. Thus, there is additional uncertainty in the average effect that is quantified by δ_{τ^2}. If $\tau^2_{res} \geq \tau^2$, the study characteristics do not explain any variance in the effect sizes, and δ_{τ^2} is truncated to zero. Hence, $\delta_{\tau^2} > 0$ if $\tau^2_{res} < \tau^2$, and 0 otherwise. Similar to the random-effects meta-analytic models, for situations with a small number of studies we implemented a hierarchical Bayesian mixed-effects meta-analytic model to estimate τ^2_{res} and, subsequently, calculate δ_{τ^2} accurately.

The Similarity Measure ω

The similarity measure ω integrates the three components into a single index. It is conceptually similar to the variance component of a Bayesian hierarchical model (comparable to the a_0-parameter of the power prior; Ibrahim et al., 2015; Neuenschwander et al., 2009). Thus, its use as weight for informative prior distributions places certain demands on the measure, both mathematically and conceptually. First, similar to the a_0-parameter of the power prior (Ibrahim et al., 2015), the similarity measure ω needs to take values between zero and one, $\omega \in [0, 1]$. This avoids any potential overweighting of the quantified background knowledge, compared to the information contained in the data of the focal study. Moreover, the similarity measure $\omega \rightarrow 1$ as the comparability of the previously conducted studies in the body of background knowledge and the focal study increases. On the one hand, when $\omega = 0$ the previously conducted studies and the focal study are not comparable, and no information contained in the informative prior distribution is used. On the other hand, when $\omega = 1$, the focal study is a direct replication of the previously conducted studies in the body of background knowledge, and the information contained in the prior distribution is used fully. Second, the similarity measure ω needs to adequately reflect the inverse relation between B, and τ^2 and δ_{τ^2}. While an increasing B indicates an increased comparability, increasing τ^2 and δ_{τ^2} indicate a decreasing comparability. Thus, the similarity measure needs to align the conceptual meaning of the three indices to reflect the comparability of the focal study with the study in the body of background knowledge adequately. Third, the similarity measure ω needs to be flexible in specification and discriminate strongly across the range of plausible values especially for τ^2 and δ_{τ^2}, which we know to typically range between 0.13 and 0.24 (van Erp et al., 2017; Stanley et al., 2018; Kenny and Judd, 2019). This aims at conservative estimates of ω, again to avoid the informative prior distribution overwhelming the likelihood of the data of the focal study. Considering all these requirements, the similarity measure ω can be expressed formally as,

$$\omega = \left(\frac{1}{1 + \exp\left[10 * \left(\sqrt{\tau^2 + \ \delta_{\tau^2}} - 0.24\right)\right]} \right) * \overline{B} \tag{1}$$

Thus, the similarity measure ω essentially is a logistic function of τ^2 and δ_{τ^2} with maximum value $L = 1$, midpoint $\omega_0 = 0.24$ and slope $s = 10$, weighted by $\overline{B} = \frac{1}{K}\sum_k B_k$, where $k = 1...K$ is the number of previously conducted studies. The parameters of this weighted logistic function are chosen so that the resulting values of the similarity measure ω adequately reflects the characteristics of Psychological research: the midpoint is carefully chosen following van Erp et al. (2017), and the slope is chosen to discriminate adequately across the typical range of between-study heterogeneity (Stanley et al., 2018; Kenny and Judd, 2019). We assume an additive relationship between τ^2 and δ_{τ^2}. Taken together, the behavior of the similarity measure is as required: $\omega \rightarrow 1$ as τ^2 and δ_{τ^2} decrease and $\overline{B}$ increases. Applying equation (1) to a situation of a Bayesian multiple regression model with three predictors and ten previously conducted studies yields three parameter-specific similarity measures, which can be used to weigh an informative prior distribution.

Applying ω – The Similarity-Weighted Informative Prior Distribution

The similarity measure ω can now be used to weight an informative prior distribution and integrate it, without any necessary intermediary calculations, in a usual Bayesian analysis. Contrary to the power prior of Ibrahim et al. (2015), who weight the likelihood of the previously conducted studies, in this case it involves raising the informative prior distribution to the power ω, $p\,(\theta \mid D) \propto p\,(D \mid \theta)\ \pi\,(\theta)^{\omega}$ where $p\,(\theta \mid D)$ is the posterior distribution of a parameter θ, $p\,(D \mid \theta)$ is the likelihood of the data, and $\pi\,(\beta\theta)^{\omega}$ is the similarity-weighted informative prior distribution. Because this prior distribution utilizes data from previously conducted studies, it belongs to the class of evidence-based informative prior distributions (Kaplan, 2014). We illustrate the use of the similarity measure ω as weight for an informative prior distribution with an example of a simple Bayesian multiple regression with three predictors. Let $\mathbf{y}$ be a n × 1-vector of outcomes, and $\mathbf{X}$ a n × p predictor matrix, where n is the sample size of the focal study and $p = 3$ the number of predictors. Then,

$$y \sim N(\beta_0 + \mathbf{X}\boldsymbol{\beta},\ \sigma^2) \tag{2}$$

is the likelihood of the Bayesian multiple regression model, with β_0 being the intercept, $\boldsymbol{\beta}$ a p × 1-vector of regression coefficients, and σ^2 being the error variance. The prior specification is as follows:

$$\beta_0 \sim N(0,\ 10) \tag{3}$$

$$\boldsymbol{\beta} \sim N(\mu_p, SE_p^2)^{\omega_p} \tag{4}$$

$$\sigma^2 \sim half - Cauchy(0,\ 2.5) \tag{5}$$

Both β_0 and σ^2 receive weakly informative prior distributions, and the hyperparameters of the informative prior distributions (means and standard deviations) for the regression coefficients β_p are the average effects μ_p and their standard errors SE_p^2 estimated by multiple univariate or a single multivariate random-effects meta-analysis (Cheung, 2015; Smid et al., 2020). They are weighted by the parameter-specific similarity measures ω_p. Generally speaking, as $\omega \rightarrow 0$ the peak around the mean of the informative prior distribution flattens, and the distribution becomes broader. A broader prior distribution carries less information about the parameter of interest; hence, the broader the distribution the lesser its informativeness.

SIMULATION

We conducted a comprehensive simulation to assess the behavior of the similarity measure ω and to investigate the performance of the similarity-weighted informative prior distribution. R-code, functions, and data of the simulation are available at https://doi.org/10.17605/OSF.IO/8AEF4.

Design

The design consisted of the following, systematically varied factors. First, the number of previously conducted studies that are part of the available body of background knowledge ($K = 3, 5, 10$). Second, the sample sizes of the previously conducted studies, indicated by the difference between the average sample sizes of these studies and the sample size of the focal study (smaller and larger $\triangle_N = -100, 100$). Third, the similarity of the predictors, indicated by the differences in means of the respective distributions (i.e., their overlap) between the previously conducted studies and the focal study (from large overlap to no overlap $\triangle_\mu = 0.25, 0.5, 1, 2, 3$). Fourth, the between-study heterogeneity in the effect sizes, thus the (lack of) consensus in the background knowledge (small to large $\tau^2 = 0.025, 0.05, 0.10, 0.15, 0.20, 0.35, 0.5$). Moreover, we simulated one moderator variable that explained 10% of the between-study heterogeneity in the effect sizes. Thus, the simulated amount of variance in outcomes and study characteristics is $\tau^2 + \delta_{\tau^2} = 0.0275, 0.055, 0.110, 0.165, 0.275, 0.385, 0.550$. In total, the design of the simulation consisted of 210 conditions.

Data Generation and Analysis

We applied the following procedure to generate the datasets in each condition. First, we simulated the dataset of the focal study, according to the multiple regression model in equation (2), with fixed sample size $N_F = 200$, true regression coefficients $\beta_F = (0.5, 0.25, -0.5)$ and a normally distributed error $\sigma_F^2 \sim N(0, 1)$. Predictors in $\mathbf{X}_F$ were drawn from standard normal distributions. Next, we constructed the database of previously conducted studies, also according to the multiple regression model in equation (2) with normally distributed error $\sigma_D^2 \quad N(0, 1)$. As a first step, the sample size for the k-th ($k = 1 \ldots K$) study of the database was drawn from a normal distribution $N(N_{P_i}, 25)$, where $N_{P_i} = N_F + \triangle_N$. In the second step, for the k-th study of the database a vector of regression coefficients β_k was drawn from a multivariate normal distribution with mean vector $\mu_{\beta_k} = (0.4, 0.0, 0.3)$, i.e., their meta-analytic means, and variance τ^2. Compared to β_F, the mean coefficients in μ_{β_k} represent certainty, disagreement, and contradiction in the size of the effect. Predictors in $\mathbf{X}_k$ were drawn from normal distributions $N(\mu_{N_P}, 1)$, where $\mu_{N_P} = \triangle_{\mu_P}$. This procedure was repeated one hundred times in each condition, resulting in 21,000 datasets (i.e., the simulated dataset of the focal study and the databases of the previously conducted studies).

Each dataset was analyzed with a Bayesian multiple regression model with (a) non-informative priors for the regression coefficients (pooled analysis), (b) the normalized power prior (NPP), (c) the meta-analytic predictive prior (MAP), and (d) the similarity-weighted informative prior distribution (SWIP). For the non-informative model, the datasets of the focal and previously conducted studies were pooled into a single dataset. The NPP was implemented as a standard normal-inverse gamma model as described in Carvalho and Ibrahim (2020). For both the MAP and SWIP a Bayesian random-effects meta-analysis was run with the generated database of previously conducted studies to calculate the meta-analytic mean effect, its standard error, and the between-study heterogeneity τ^2. The meta-analytic mean effect and its standard error were used as hyperparameters of the MAP and SWIP. The meta-analysis was based on Fisher's r-to-z transformed partial correlation coefficients using the metafor-package (Viechtbauer, 2010). This follows Aloe and Thompson (2013) who introduced partial or semi-partial correlations as adequate effect sizes for regression coefficients. The specification of the MAP model and its robustification procedure followed the standard implementation of the RBesT-package outlined in Weber et al. (2019). Prior to the SWIP analysis, the modified generalizability index $\overline{B}$ for the previously conducted studies and the similarity measure ω was calculated as in equation (1). The similarity measure ω was then introduced as parameter-specific weight for the informative prior distributions for the regression coefficients as in equation (4). All models were specified with Stan and its R interface *RStan* (Stan Development Team, 2020). Four chains each of length 2,000 with 1,000 burn-in cycles were set up. Different random starting values were supplied to each chain. Convergence was assessed using the Gelman-Rubin R-statistic (Gelman and Rubin, 1992), where $R < 1.02$ indicated convergence. All solutions converged.

Evaluation Criteria

To assess the behavior of the similarity measure ω we focused on its relation to $\tau^2 + \delta_{\tau^2}$ and $\triangle_\mu$, and its relation to the shrinkage in the parameter estimates. Therefore, we estimated linear models. Shrinkage was defined as the difference between the focal-study estimates (the true values β_F) and the estimates obtained by the similarity-weighted informative prior distribution. Moreover, comparing the performance of the different prior distributions involved, for each condition, averaging the parameter estimates and their standard errors over replications, $\overline{\beta} = \frac{1}{R}\sum_R \beta$ and $\overline{SE_\beta} = \sqrt{\frac{1}{R}\sum_R SE_\beta^2}$, respectively. The similarity measure behaves as expected if it decreases as $\tau^2 + \delta_{\tau^2}$ and $\triangle_\mu$ increase. Moreover, shrinkage should increase as the similarity increases. Good performance of the different informative prior distributions is indicated by increasing shrinkage of the parameter estimates toward their meta-analytic means, as well as decreasing standard errors of the parameter estimates, depending on the degree of similarity.

RESULTS

Behavior of the Similarity Measure ω

Figure 1 illustrates the behavior of the similarity measure ω conditional on $\tau^2 + \delta_{\tau^2}$ for different levels of $\triangle_\mu$ combined for all three regression coefficients (left panel), and the behavior of the shrinkage of the estimates of the three regression coefficients, conditional on the similarity measure ω (right panel), across all simulation conditions. The similarity measure ω behaves as expected; as both $\tau^2 + \delta_{\tau^2}$ and $\triangle_\mu$ increase, i.e., the similarity between the focal and the previously conducted studies decreases, the similarity measure ω decreases as well. Moreover, we have a non-compensatory relation between the

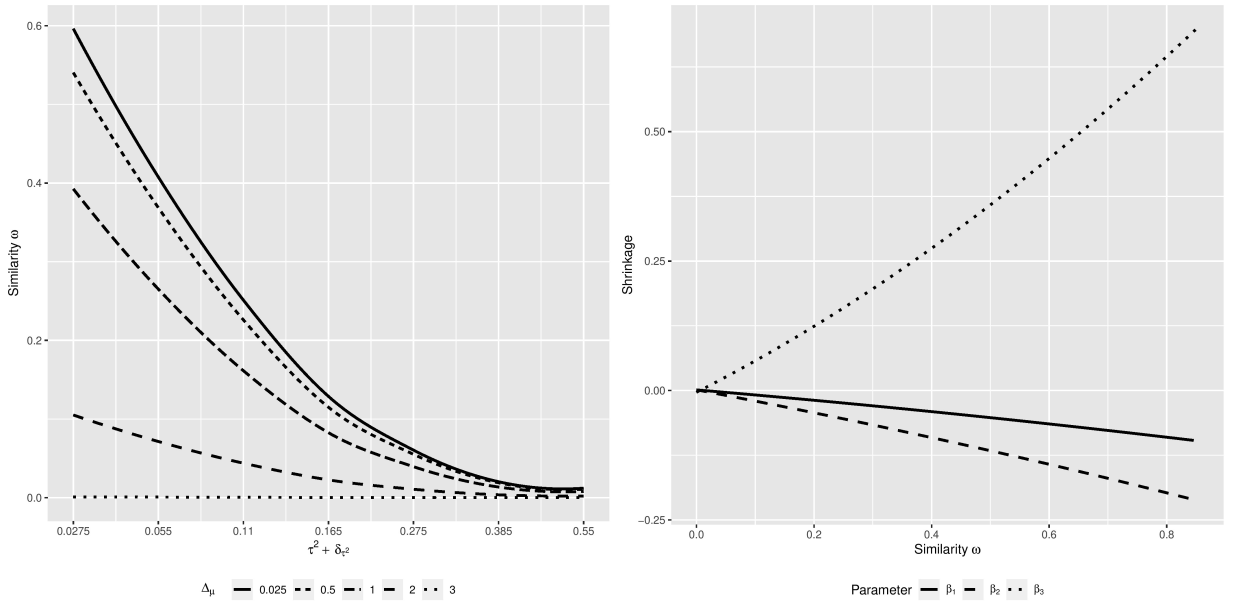

FIGURE 1 | Regression curves of the relation between the similarity measure ω and $\tau^2 + \delta_{\tau^2}$ for different levels of $\triangle_\mu$ (**left panel**), and the relation between the shrinkage and the similarity measure ω, for each β-parameter (**right panel**), based on estimates from 21,000 simulated datasets.

components of the similarity measure. High similarity in samples and predictors does not compensate for a lack of similarity regarding outcomes and study characteristics, and vice versa. The shrinkage of the parameter estimates behaves accordingly: as the focal and the previously conducted studies become more similar, indicated by an increasing similarity measure ω, the estimates of the regression coefficients shrink toward their meta-analytic means. If the focal and previously conducted studies are highly dissimilar, shrinkage is close to zero, and the estimates of the regression coefficients remain at estimates resulting from the focal study. Lastly, shrinkage is stronger when the meta-analytic means and the focal-study estimates of the regression coefficients are considerably apart (see β_3, compared to the other two parameters). This is, however, just an effect of the distance between the values of $\beta_3 = -0.5$ and its meta-analytic mean $\mu_{\beta_3} = 0.3$. With an increasing distance between a parameter estimate and it meta-analytic mean, the potential amount of shrinkage increases as well. Moreover, the different direction of the shrinkage in case of β_3 is due to the meta-analytic mean being larger than the focal-study estimate. In case of the other regression coefficients, their meta-analytic means are smaller than their focal-study estimates, thus the shrinkage is negative.

Performance of the Similarity-Weighted Informative Prior Distribution

Figures 2, **3** illustrate the behavior of the estimates of the three regression coefficients and their standard errors, respectively, obtained from the pooled Bayesian analysis, the NPP, the MAP, and the SWIP, conditional on the simulated factors. The estimated regression coefficients obtained with the SWIP lie consistently between their true values β_F and their true meta-analytic means μ_{β_k}. Shrinkage toward the true meta-analytic means is sensitive to changes in both $\tau^2 + \delta_{\tau^2}$ and $\triangle_\mu$. In contrast, the MAP consistently yields parameter estimates close to the true values β_F, except for β_3 when $\tau^2 + \delta_{\tau^2} < .10$. Thus, the MAP is largely insensitive to changes in both $\tau^2 + \delta_{\tau^2}$ and $\triangle_\mu$. Compared to the NPP, shrinkage of the parameter estimates of the SWIP is comparably sensitive to changes in both $\tau^2 + \delta_{\tau^2}$ and $\triangle_\mu$, but more conservative. For example, when $\triangle_\mu$ is large, the NPP sometimes yields overestimated parameters. Moreover, while the SWIP shrinks the parameters never beyond their estimates obtained with the pooled analysis, the NPP shrinks the parameter estimates in some cases beyond their meta-analytic means.

This general pattern is similar in case of the standard error of the parameter estimates. In case of the SWIP, the standard errors decrease as the similarity of the focal and previously conducted studies increases. More specifically, they converge to the standard errors of the pooled Bayesian analysis. This implies a similarity-dependent borrowing of information from the previously conducted studies that increases the precision of the parameter estimates of the focal study. This is true for all simulation conditions, although it is most distinct when the number of available studies is large ($K = 10$). In contrast, the standard errors of the estimates of the MAP do not converge; they largely remain at around 0.7. Thus, the MAP does not borrow information from the previously conducted studies. The standard errors of the estimates of the NPP tend to be smaller than the standard errors of the SWIP, especially when the number of previously conducted studies is large ($K = 10$). Thus, the NPP borrows more information. When the focal-study estimates and their meta-analytic means

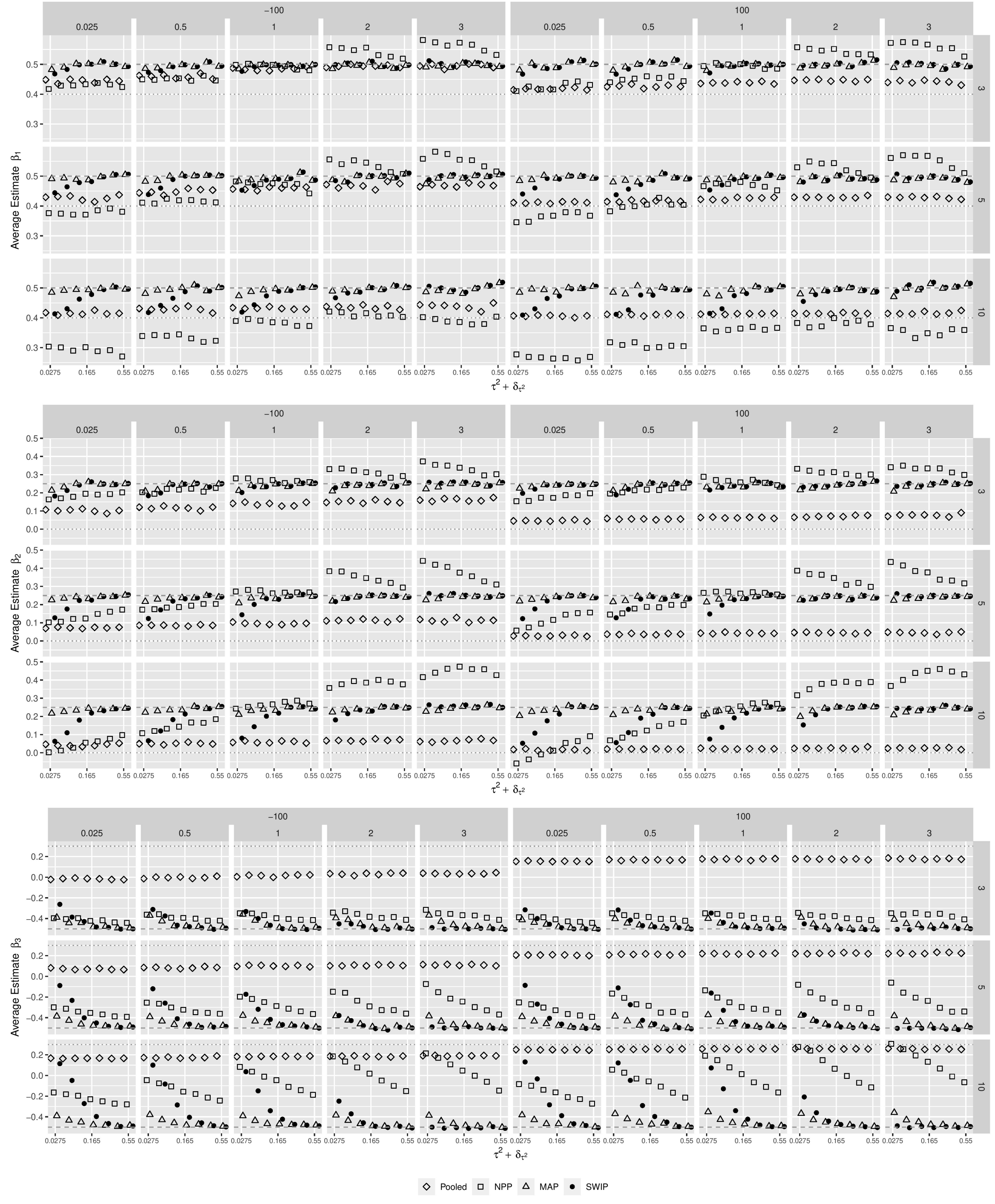

FIGURE 2 | The behavior of the parameter estimates across simulation conditions. The similarity of the focal and the previously conducted studies decreases from **left** to **right**. Pooled = pooled Bayesian analysis; NPP = normalized power prior; MAP = meta-analytic predictive prior; SWIP = similarity-weighted informative prior distribution. The dashed horizontal line represents the true value of the respective regression coefficient of the focal study. The dotted horizontal line represents the true (generating) meta-analytic mean of the respective regression coefficient.

contradict (in case of β_3), however, the standard errors of the estimates of the NPP tend to be larger, especially when the number of previously conducted studies is small and $\triangle_\mu$ is large.

Overall, the performance of the SWIP is more consistent and sensitive to changes in similarity between the focal and previously conducted studies, compared to both the NPP and MAP, while yielding conservative estimates. As the similarity

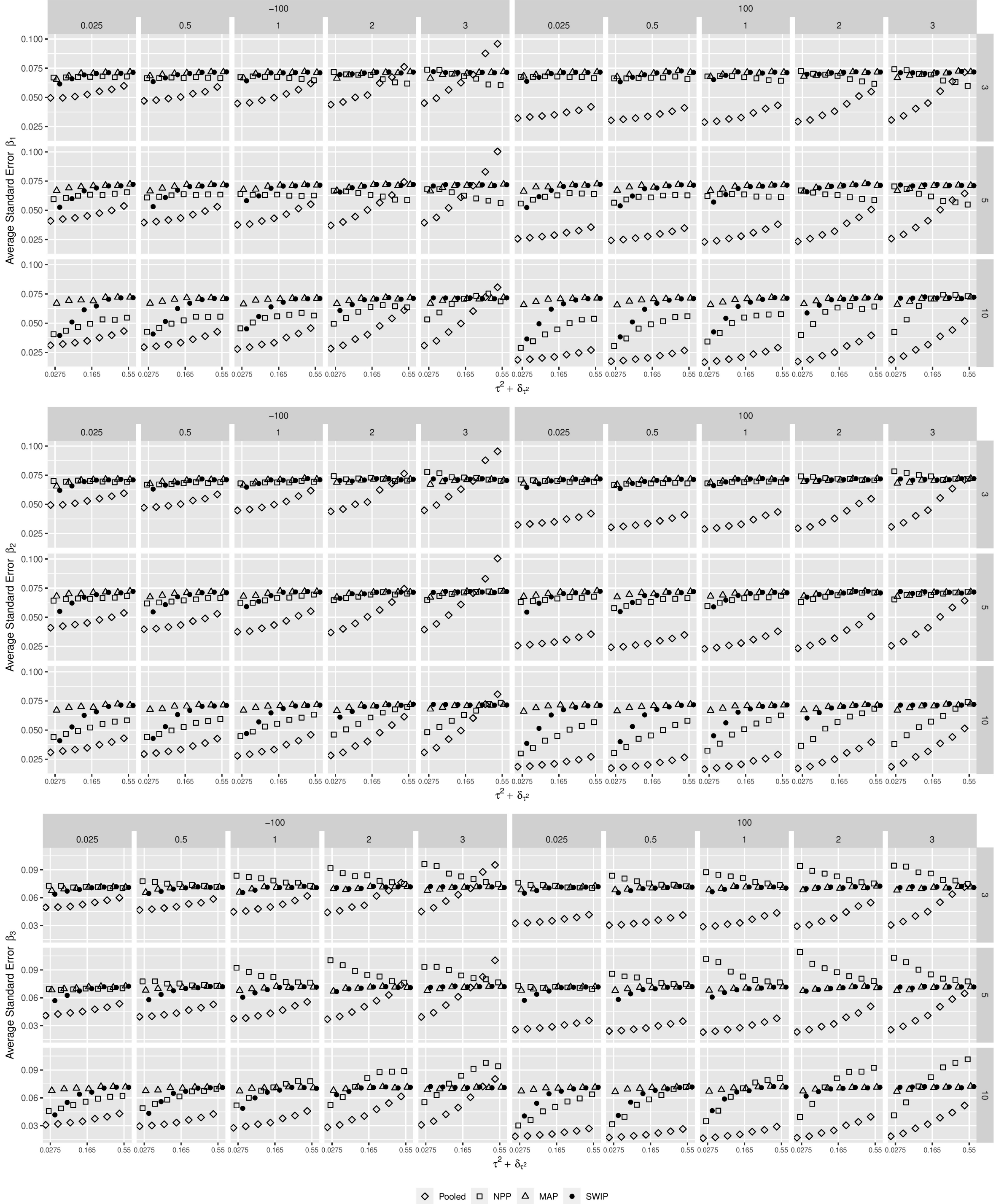

FIGURE 3 | The behavior of the standard errors of the parameter estimates across simulation conditions. The similarity of the focal and the previously conducted studies decreases from **left** to **right**. Pooled = pooled Bayesian analysis; NPP = normalized power prior; MAP = meta-analytic predictive prior; SWIP = similarity-weighted informative prior distribution.

increases, the parameter estimates of the SWIP shrink toward the estimates of the pooled Bayesian analysis, and more information is borrowed from the body of available background knowledge. Thus, the standard errors of the parameter estimates decrease, and the estimates are more precise. In this context, the number of previously conducted studies plays a vital role. When the number is small, i.e., when there is less information to borrow, both shrinkage and precision are less distinct.

DISCUSSION

The purpose of this study was to illustrate a novel method to assess the similarity of studies in the context of specifying informative prior distributions for Bayesian multiple regression models. We illustrated the quantification, based on a propensity-score approach and random- and mixed-effects meta-analytic models (e.g., Tipton, 2014; Cheung, 2015), and combination of heterogeneity in samples and predictors, outcomes, and study characteristics in the novel similarity measure ω. We showed how to use the similarity measure ω as a weight for informative prior distributions for the regression coefficients, and investigated the behavior of the similarity measure ω and the similarity–weighted informative prior distribution, comparing its performance to the normalized power prior and meta-analytic predictive prior.

The Performance of the Similarity-Weighted Informative Prior Distribution

The results of our simulation show that the parameter estimates of the similarity-weighted informative prior distribution behave similar to those of hierarchical Bayesian models: as the similarity of the focal and previously conducted studies increases, they shrink toward their pooled, meta-analytic means. Simultaneously, the precision of the parameter estimates increases because more information is borrowed from the previously conducted studies. From the perspective of cumulative knowledge creation, this behavior is desired. As evidence from comparable studies accumulates, our knowledge of the size of an effect becomes incrementally more certain until, over time, it represents the best knowledge we have (unless the evidence contradicts; Kruschke et al., 2012; König and van de Schoot, 2018). The meta-analytic predictive prior, on the one hand, does not provide this increasing certainty in the size of an effect. Compared to the similarity–weighted informative prior distribution, the similarity-dependent shrinkage is much less distinctive. Since the meta-analytic predictive prior only considers the heterogeneity in outcomes, it may be an indication that, echoing Lin et al. (2017), this is not sufficient for an adequate assessment of similarity of the focal and previously conducted studies. Parameter estimates of the normalized power prior, on the other hand, exhibit a stronger, but inconsistent shrinkage toward the pooled, meta-analytic means. From the perspective of cumulative knowledge creation, this is problematic, because the normalized power prior provides parameter estimates that are biased, and the precision of the estimates does not increase consistently as evidence accumulates.

Since the performance of the similarity-weighted informative prior distribution stands or falls with the accuracy of the components of the similarity measure ω, it is essential to estimate the random and mixed-effects meta-analytic models as unbiased as possible. This is usually based on either maximum likelihood (ML) or restricted maximum likelihood (REML) estimation (e.g., Cheung, 2015). These likelihood-based methods, however, exhibit poor performance especially when the number of previously conducted studies is small (Bender et al., 2018), additionally to the general underestimation of the between-study heterogeneity of ML-based random-effects meta-analytic models (Cheung, 2015). Several studies show a superior performance of Bayesian approaches, especially hierarchically specified random and mixed-effects meta-analytic models, in terms of the accuracy of the (residual) variance components (Williams et al., 2018; Seide et al., 2019). Thus, when using the similarity measure ω to specify the similarity-weighted informative prior distributions, we recommend using these Bayesian approaches to estimate both the mean effect size and its variance components, as illustrated in this study.

On the one hand, the similarity-weighted informative prior distribution simplifies the concept of the normalized power prior. The similarity measure is used to weight the informative prior distribution directly, which is more intuitive and less challenging than weighting the likelihood of the data from the previously conducted studies (Ibrahim et al., 2015). The complex calculation of multiple marginal likelihoods by means of bridge sampling approaches (see Carvalho and Ibrahim, 2020) is not necessary. Calculating marginal likelihoods can be complicated and time-consuming especially when the underlying models are complex (for instance, structural equation models), and their likelihood is analytically intractable (Ibrahim et al., 2015). On the other hand, the similarity-weighted informative prior distribution extends both the normalized power prior and meta-analytic predictive prior by taking into account multiple sources of heterogeneity in previously conducted studies, and quantifying these sources in the similarity measure ω. The benefits of this holistic approach are illustrated by the performance of the similarity-weighted informative prior distribution.

Future Directions

The similarity measure ω and the similarity-weighted informative prior distribution offer various opportunities for further research. First, the inconsistent behavior of the normalized power prior may be due to the limited number of available small-sample studies (Neuenschwander et al., 2009). Thus, a limitation of this study is that we only considered sample sizes of the focal and previously conducted studies that are of a comparable order of magnitude. Investigating the performance of the similarity-weighted informative prior distribution in situations where these sample sizes differ by orders of magnitude, and where the sample sizes of the previously conducted studies vary considerably, is an important topic for further research. If the sample sizes of the focal and previously conducted studies vary considerably in size (especially when $N_P \gg N_F$), it is possible to multiply the scale parameter of the informative prior distribution SE_p^2 by the ratio N_P/N_F. This can be understood as a mechanism to avoid that the prior information overwhelms the likelihood, because it flattens the distribution and makes it less informative. Second, the similarity measure can be used as the a_0-parameter of the normalized power prior. Investigating the behavior of the normalized power prior in the context of a fixed–a_0 approach, where the study-specific a_0-parameters are fixed to the values of the study-specific similarity measures may be an interesting topic for future research. Especially because the fixed–a_0 approach is considered superior to the random–a_0 approach, where the comparability of the focal and previously conducted studies is inferred from the data, and the prior distribution for the a_0-parameter has to be chosen carefully (Neuenschwander et al., 2009; Ibrahim et al., 2015). Third, comparing ML-based and Bayesian meta-analytic or other approaches in the context of assessing the similarity of studies, i.e., regarding their impact on the behavior of the similarity-weighted informative prior distribution, is another important topic for future studies. As mentioned above, the precision of the average effect sizes that are used as the hyperparameters of the informative prior distributions, are pivotal for the accuracy of these distributions. Identifying the correct approach, especially when the number of previously conducted studies is small (Bender et al., 2018), is crucial for the performance of the similarity-weighted informative prior distribution. Fourth, the calculation of the modified generalizability index $\overline{B}$ still requires the availability of the raw data of the previously conducted studies. This remains a limitation for the applicability of the similarity measure. Extending its applicability is a question of being able to calculate the modified generalizability index $\overline{B}$ in situations when only summary data are available. It is possible, however, to simulate a number of datasets based on correlation matrices, or means and standard deviations, and to calculate $\overline{B}$ for each of the simulated datasets. The pooled $\overline{B}$ can then be used to calculate the similarity measure. Such an approach, similar to multiple imputation or the estimation of plausible values, will be addressed and investigated in a future study. Fifth, both the similarity measure and the similarity-weighted informative prior distribution are currently only available for multiple regression models, i.e., univariate methods. It may be fruitful to extend and adapt both to multivariate methods, for example structural equation models.

Concluding Remarks

As mentioned in the introduction to this study, specifying accurate informative prior distributions is a question of carefully selecting studies that comprise the body of comparable background knowledge. Given the considerable heterogeneity of studies that are being conducted in Psychological research (different circumstances, with different samples and instruments), the results of these studies are heterogeneous, and not all available results can and should contribute equally to an informative prior distribution. The similarity measure ω and the similarity-weighted informative prior distribution developed in this study provide researchers with tools to (a) justify the selection of studies that contribute to the informative prior distribution, and (b) to accomplish the necessary similarity-based weighting of the available background knowledge. On the one hand, the quantification of the similarity of studies, and the similarity-based weighting of prior information, are important elements of a systematization of the specification and use of informative prior distribution. Being able to justify empirically the use of previously conducted studies for the specification of informative prior distributions, on the other hand, helps building confidence in the use of informative prior distributions. The theoretical rationale of the similarity measure ω and the evidence-based nature of the similarity-weighted informative prior distribution may help to supersede the subjective notion of informative prior distributions. We hope that the similarity measure ω and the similarity-weighted informative prior distribution stimulates further research, eventually helping researchers in Psychology to move beyond non-informative prior distributions, and to finally exploit the full potential of Bayesian statistics for cumulative knowledge creation.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: Open Science Framework—http://doi.org/10.17605/OSF.IO/8AEF4.

AUTHOR CONTRIBUTIONS

CK developed the conceptual background, designed, programmed, and ran the simulation, analyzed the data, and wrote the manuscript.

REFERENCES

Aloe, A., and Thompson, C. (2013). The synthesis of partial effect sizes. *J. Soc. Soc. Work Res.* 4, 390–405. doi: 10.5243/jsswr.2013.24

Bender, R., Friede, T., Koch, A., Kuss, O., Schlattmann, P., Schwarzer, G., et al. (2018). Methods for evidence synthesis in the case of very few studies. *Res. Synthesis Methods* 9, 382–392. doi: 10.1002/jrsm.1297

Carvalho, L. M., and Ibrahim, J. (2020). On the normalized power prior. *arxiv [Preprint]*

Cheung, M. W.-L. (2015). metaSEM: an R package for meta-analysis using structural equation modeling. *Front. Psychol.* 5:1521. doi: 10.3389/fpsyg.2014.01521

Cronbach, L. J. (1982). *Designing Evaluations of Educational and Social Programs*. San Francisco, CA: Jossey Bass.

Cumming, G. (2014). The new statistics: why and how. *Psychol. Sci.* 25, 7–29.

Finch, W. H., and Miller, J. E. (2019). The use of incorrect informative priors in the estimation of MIMIC Model parameters with small sample sizes. *Struct. Equation Modeling Multidisciplinary J.* 26, 497–508. doi: 10.1080/10705511.2018.1553111

Gelman, A., and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Stat. Sci.* 7, 457–511. doi: 10.1214/ss/1177011136

Goldstein, M. (2006). Subjective bayesian analysis: principles and practice. *Bayesian Anal.* 1, 403–420. doi: 10.1214/06-BA116

Ibrahim, J., Chen, M.-H., Gwon, Y., and Chen, F. (2015). The power prior: theory and applications. *Stat. Med.* 34, 3724–3749. doi: 10.1002/sim.6728

Jak, S. (2015). *Meta-Analytic Structural Equation Modeling.* Berlin: Springer.

Kaplan, D. (2014). *Bayesian Statistics for the Social Sciences.* New York, NY: Guilford.

Kenny, D. A., and Judd, C. M. (2019). The unappreciated heterogeneity of effect sizes: implications for power, precision, planning of research, and replication. *Psychol. Methods* 24, 578–589. doi: 10.1037/met0000209

König, C., and van de Schoot, R. (2018). Bayesian statistics in educational research: a look at the current state of affairs. *Educ. Rev.* 70, 486–509. doi: 10.1080/00131911.2017.1350636

Kruschke, J., Aguinis, H., and Joo, H. (2012). The time has come: bayesian methods for data analysis in the organizational sciences. *Organ. Res. Methods* 15, 722–752. doi: 10.1177/0956797613504966

Lin, L., Chu, H., and Hodges, J. (2017). Alternative measures of between-study heterogeneity in meta-analysis: reducing the impact of outlying studies. *Biometrics* 73, 156–166. doi: 10.1111/biom.12543

Makel, M. C., Plucker, J. A., and Hegarty, B. (2012). Replications in psychology research: how often do they really occur? *Perspect. Psychol. Sci.* 7, 537–542. doi: 10.1177/1745691612460688

McNeish, D. (2016). On using bayesian methods to address small sample problems. *Struct. Equation Modeling Multidisciplinary J.* 23, 750–773. doi: 10.1080/10705511.2016.1186549

Moss, J., and Tveten, M. (2019). kdensity: an R package for kernel density estimation with parametric starts and asymmetric kernels. *J. Open Sour. Softw.* 4:1566. doi: 10.21105/joss.01566

Neuenschwander, B., Branson, M., and Spiegelhalter, D. (2009). A note on the power prior. *Stat. Med.* 28, 3562–3566. doi: 10.1002/sim.3722

Neuenschwander, B., Capkun-Niggli, G., Branson, M., and Spiegelhalter, D. (2010). Summarizing historical information on controls in clinical trials. *Clin. Trials* 7, 5–18. doi: 10.1177/1740774509356002

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Rev. General Psychol.* 13, 90–100. doi: 10.1037/a0015108

Seide, S., Röver, C., and Friede, T. (2019). Likelihood-based random-effects meta-analysis with few studies: empirical and simulation studies. *BMC Med. Res. Methodol.* 19:16. doi: 10.1186/s12874-018-0618-3

Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton-Mifflin.

Smid, S. C., McNeish, D., Miocevic, M., and van de Schoot, R. (2020). Bayesian versus frequentist estimation for structural equation models in small sample contexts: a systematic review. *Struct. Equation Modeling* 27, 131–161. doi: 10.1080/10705511.2019.1577140

Stan Development Team, (2020). *Rstan: The R interface to Stan, Version 2.19.3*. Available online at: http://mc-stan.org/users/interfaces/rstan.html (accessed September 1, 2020).

Stanley, T. D., Carter, E. C., and Doucouliagos, H. (2018). What meta-analyses reveal about the replicability of psychological research. *Psychol. Bull.* 144, 1325–1346. doi: 10.1037/bul0000169

Tipton, E. (2014). How generalizable is your experiment? An index for comparing experimental samples and populations. *J. Educ. Behav. Stat.* 39, 478–501. doi: 10.3102/1076998614558486

Tipton, E., and Olsen, R. B. (2018). A review of statistical methods for generalizing from evaluations of educational interventions. *Educ. Res.* 47, 516–524. doi: 10.3102/0013189X1878152

van de Schoot, R., Winter, S. D., Ryan, O., Zondervan-Zwijnenburg, M., and Depaoli, S. (2017). A systematic review of Bayesian articles in psychology: the last 25 years. *Psychol. Methods* 22, 217–239. doi: 10.1037/met0000100

van Erp, S., Verhagen, A. J., Grasman, R. P. P. P., and Wagenmakers, E.-J. (2017). Estimates of be-tween-study heterogeneity for 705 meta-analyses reported in Psychological Bulletin from 1990–2013. *J. Open Psychol. Data* 5:4. doi: 10.5334/jopd.33

Vanpaemel, W. (2011). Constructing informative model priors using hierarchical methods. *J. Math. Psychol.* 55, 106–117. doi: 10.1016/j.jmp.2010.08.005

Veroniki, A. A., Jackson, D., Viechtbauer, W., Bender, R., Bowden, J., Knapp, G., et al. (2016). Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Res. Synthesis Methods* 7, 55–79. doi: 10.1002/jrsm.1164

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.* 36, 1–48.

Weber, S., Li, Y., Seaman Iii, J. W., Kakizume, T., and Schmidli, H. (2019). Applying meta-analytic predictive priors with the R Bayesian evidence synthesis tools. *arxiv [Preprint]*

Williams, D. R., Rast, P., and Bürkner, P.-C. (2018). Bayesian meta-analysis with weakly informative prior distributions. *PsyArXiv [Preprint]* doi: 10.31234/osf.io/7tbrm

Zhang, Z., Jiang, K., Liu, H., and Oh, I.-S. (2017). Bayesian meta-analysis of correlation coefficients through power prior. *Commun. Stat. Theory Methods* 46, 11988–12007. doi: 10.1080/03610926.2017.1288251

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.