Check for updates

Risky research? Exploring the potentially detrimental effects of employing stigma scales

Alexander Ort[1]*† and Freya Sukalla[2]†

[1]Faculty of Health Sciences and Medicine, University of Lucerne, Lucerne, Switzerland, [2]Institute for Communication and Media Studies, Leipzig University, Leipzig, Germany

Research on stigma is confronted with major ethical challenges. One potential risk of investigating stigma utilizing self-reports includes the unintentional reinforcement of stigma. Commonly used self-report scales to assess stigma usually include items that directly confront people with the negative stereotypes underlying the stigma. Even though findings in the domain of priming research suggest that such a way of assessing stigma might potentially activate and reinforce existing stigma-relevant beliefs, research to date has neglected the issue of potential detrimental effects. A preregistered online experiment was conducted with a sample of 762 participants (51.1% female, $M_{age} = 49.7$ years, $SD_{age} = 16.4$, 69.9% with some form of tertiary education). The objective of the study was to explore the potential impact of exposure to stigma scales which incorporate negative stereotypes on the development of stigmatizing attitudes toward two specific groups: individuals who use pre-exposure prophylaxis for HIV prevention and patients who undergo weight loss surgery. The findings underline the relevance of the issue by showing that responding to negative stereotypical items on a stigma scale bears the risk of facilitating scale-related stereotype accessibility, negative judgment, and promoting social demarcation from the groups under investigation.

Introduction

Social or public stigma is a prevalent and multifaceted social phenomenon usually described as involving negative evaluations and stereotypes associated with a distinguishing characteristic (Goffman, 1963; Link and Phelan, 2001). Such devalued characteristics range from gender, race, or sexuality (e.g., Frost, 2011) to disability, weight, or mental illness (e.g., Stangl et al., 2019), as well as infectious diseases like COVID-19 (e.g., Saeed et al., 2020). In addition, labeling people as different and linking these differences to negative stereotypes creates a clear separation between the superior "us" and the inferior "them" (Jones et al., 1984; Link and Phelan, 2001; Smith, 2007). Relatedly, stigmatization is a complex process that refers to excluding, discriminating, or marginalizing a person or group based on a particular characteristic stigmatized within a society or social context (Major and O'Brien, 2005). Stigmatization can have far-reaching negative consequences for the affected individuals. Due to its potential to increase social and structural inequalities and lead to severe cognitive and affective consequences for individuals, it constitutes a primary global health concern (Hatzenbuehler et al., 2013; Link et al., 2014; Pescosolido and Martin, 2015).

In light of the potential hazard stigmatization poses for the social life of the individuals being stigmatized as well as for their mental and physical health, research in various

disciplines such as sociology, psychology, or public health has acknowledged the importance of these issues by thoroughly investigating the occurrence of stigma and how it can be overcome (Bos et al., 2013; Pescosolido and Martin, 2015). One crucial and overarching characteristic of quantitative studies examining public stigma in the general population is that they usually include an assessment of the level of the stigma toward a particular group or the tendency to enact stigma, with (validated) self-report scales (Link et al., 2004). Because of their centrality, standardized scales constitute a central instrument and belong to the standard methodological repertoire in quantitative research investigating social or public stigma. To assess the multifaceted nature of stigma as a social phenomenon, most scales consist of multi-item measures that aim to cover the central dimensions involved in the given stigma at hand (Bresnahan and Zhuang, 2011; Pescosolido and Martin, 2015). While scholars regularly note a lack of consistency (partly due to the specificity of different stigmas) and clarity among measures, stereotyping and discrimination belong to the most commonly measured dimensions and can be seen as overarching and unifying elements in stigma research (Link et al., 2004; Pescosolido and Martin, 2015; Fox et al., 2018).

Beyond the challenge of measuring social stigma, research on stigma is confronted with significant ethical challenges. One potential risk of investigating the phenomenon in the general population is the unintentional reinforcement of stigmatization and its adverse outcomes (Millum et al., 2019). This is particularly important, as the central goal of research in this area is to provide evidence on how to fight and overcome stigma. Considering this challenge and risk, it is surprising that studies have not devoted more attention to the issue of unintended effects of stigma research. In contrast to the discussions in the context of, especially qualitative, research with stigmatized groups (Millum et al., 2019; Gabbidon and Chenneville, 2021), quantitative studies investigating public stigma in the general population— if at all—account for the unintentional reinforcement of stigma through research by providing recommendations on avoiding the negative consequences of stigmatizing language. In their article on measuring mental illness stigma, Link et al. (2004), for example, formulate guiding questions to be asked when selecting measures, among them whether "…the words and phrases used to refer to or describe people with mental illnesses [are] appropriately sensitive and respectful" (p. 518).

However, besides respectfully referring to the people affected by stigma, self-report scales usually include items that directly confront people with the negative stereotypes and evaluations underlying the stigma itself. For example, study participants are asked how much they agree that a person with schizophrenia is "dangerous" (e.g., Corrigan et al., 2015), a person with an eating disorder is using it "to gain attention" (e.g., Roehrig and McLean, 2010), or a person experiencing homelessness is "lazy" (e.g., Knecht and Martinez, 2009). By directly confronting participants with stigmatizing statements, the measurement itself might potentially activate and reinforce existing cognitive stigma-relevant patterns or even create new ones. Detrimental effects from these mechanisms likely manifest on different levels, such as increasing implicit and explicit stigma. Considering the lack of evidence about the occurrence and nature of

potential detrimental effects, this study will examine the following research question:

RQ: Do stigma scales assessing negative stereotypes promote stigmatizing tendencies toward the affected group under investigation?

By posing and investigating this research question, we would like to, first and foremost, promote reflection on the possible unintentional negative effects of our research practices in general, but more specifically regarding the use of self-report scales of stigma beliefs. Investigating these risks is not meant to disregard or devalue the invaluable insights previous researchers have generated, nor does it constitute a call for abandoning stigma research altogether. On the contrary, we aim to contribute to responsible research practices in studying stigma. This exploratory study on the possible detrimental effects of stigma scales is one step within the specific context of quantitative research using self-report scales.

Stigma-relevant priming and stereotype activation through self-report scales

Applying priming theory and related empirical findings provides a reasonable basis for investigating this phenomenon in more detail (for an extensive overview, see Carpentier, 2020). Priming theory has its roots in free associations techniques and thought associations and focuses on cognitive structures and how concepts in memory are stored, linked with one another, and accessed (e.g., Tulving, 1972). Further developments in this domain resulted in specific work on political priming (Iyengar et al., 1982), media priming (Berkowitz, 1984), but also—and in the light of this research most relevant—stereotype priming (Devine, 1989). Building on these findings and more recent knowledge about stereotype activation and application (Hilton and von Hippel, 1996; Bargh, 1999; Macrae and Bodenhausen, 2000; Greenwald et al., 2002), it can be assumed that exposure to stigma scales assessing negative stereotypes (i.e., words, statements, or questions) to measure the phenomenon creates or (re-)activates (existing) associations between those characteristics and the respective stigmatized group. This, in turn, might increase implicit and explicit stigma. To assess effects on explicit stigma without directly asking about stigma beliefs, we focus on downward social comparison as an indicator of outgroup devaluation and desired social distance as an indicator of discrimination tendencies. We therefore hypothesize:

H1: Exposure to a stigma scale assessing negative stereotypes will increase (a) implicit stigma, (b) downward social comparisons, and (c) desired social distance.

The role of knowledge and topic involvement

In addition, it is vital to consider an individual's knowledge about a topic and its personal relevance to them. More

knowledgeable individuals are likely—by way of spreading activation (e.g., Collins and Loftus, 1975)—to have a wider variety of information for both judgment and control of stereotype application compared to less knowledgeable individuals who should be more likely to apply the negative stereotypes primed by scale exposure (Devine, 1989). Research has shown that education interventions can reduce stigma (e.g., Corrigan et al., 2012). On the other hand, individuals who perceive a topic as less relevant personally are less likely to engage in motivated inhibition of stereotype application (Kunda and Sinclair, 1999; Devine and Sharp, 2009; Rees et al., 2019). Consequently, prior knowledge and topic involvement can moderate the effect of stigma scale exposure. Therefore, we assume that:

H2: *The effect of stigma scale exposure will be stronger for participants who (a) know less or (b) care less about the concerned topic.*

Methods

Research design and procedure

To test the proposed hypotheses and answer our research question, we conducted a preregistered[1] online experiment with a 2 (*stigma scale exposure: no/yes*) × 2 (*topic: PrEP[2] user/WLS[3] patients*) between-subject design. Data were collected using the *SoSci Panel*, a German non-commercial online convenience respondent panel with participants from the general population (Leiner, 2016). The panel requires researchers to apply for access which involves a review of the questionnaire and procedures by at least two experts. After approval, respondents were invited via email to participate in a study on perceptions of health issues. After informed consent and warm-up questions about demographics, social comparison tendencies, and eHealth literacy, participants were randomly assigned to one of the four experimental groups and asked to respond to items referring to either PrEP users or weight loss surgery patients (*topic*). After knowledge and perceived relevance questions, we provided background information on the respective health topics before participants continued to the relevant section of the stigma-related dependent variables. At the beginning of this part of the questionnaire, participants were asked to respond to a stigma scale or not (*stigma scale exposure*). In the end, participants were thoroughly debriefed regarding the study's purpose, during which we also explicitly explained and dismantled the stigma and stereotypes surrounding WLS and PrEP. On average, participants took 10 min and 23 s to complete the study (*SD* = 3 m 22 s). The ethical review board of the University of Leipzig has approved the study protocol (reference #: 2021.04.01_eb_82).

---

1   https://osf.io/adegr/?view_only=a5bac8b1eb0049a79eeb25ccb7fa85f6

2   Pre-exposure prophylaxis (HIV prevention).

3   Weight-loss surgery.

Sample

Concerning the planned analyses, i.e., analyses of covariance and multiple linear regressions including main effects and interactions, a statistical power analysis was performed for sample size estimation with G*Power (Version 3.1.9.4; Faul et al., 2009). Based on conservative criteria, the necessary sample size was computed for small effect sizes, i.e., $f = 0.10$. With an alpha = 0.05 and power = 0.80, the estimated required sample size is $N = 787$.

A total of 805 individuals from the general population participated in the study. Of those, 12 were excluded for finishing in less than five minutes, and 31 were excluded for missing values on the main dependent variables (social comparison, social distance). The final sample comprises 762 participants, of which 51.1% are women. Age ranges from 17 to 87 years, with a mean of 49.7 years ($SD = 16.4$). The sample is skewed toward higher education, with 69.9% reporting some form of tertiary education.

Experimental manipulation

Topic selection: weight loss surgery and pre-exposure prophylaxis

To examine the possible detrimental effects of stigma scale exposure, we chose to do so for two different topics, allowing us to examine whether these effects can be found across topics. Ultimately, we chose weight loss surgery (WLS), referring to surgical procedures to tackle overweight and obesity, and pre-exposure prophylaxis (PrEP), a drug to prevent HIV infection without using condoms because even though they are both health-related, they vary in several relevant ways. On the one hand, WLS might trigger prejudices related to people exceeding a certain weight or not fulfilling a specific body ideal, e.g., being lazy or lacking control concerning food intake (Zhu et al., 2022). On the other hand, PrEP is often associated with sexuality-related ideas, e.g., irresponsibility and promiscuity (Dubov et al., 2018). Even though participants can be expected to have existing knowledge, beliefs, and attitudes about both topics, they were not expected to be too familiar with or have excessive knowledge about WLS and PrEP.

The results from our sample indicate that a larger proportion of participants were familiar with WLS compared to PrEP. Specifically, only 20.6% of the participants reported having no knowledge of WLS, whereas 61.3% indicated that they were not familiar with PrEP. In addition, participants asked about PrEP users agreed more with stigma items than those asked about WLS patients, though the agreement is generally low. WLS patients are seen as warmer but less competent than PrEP users. Table 1 provides an overview of the topics' attributes.

Scale exposure

For the experimental manipulation of scale exposure, comparable 8-item scales (7-pt. Likert scale) were created for PrEP users and WLS patients. We adapted items from existing scales (WLS: Hansen and Dye, 2018; PrEP: Klein and Washington, 2019) and streamlined them so that the two versions are comparable in structure (see Table 2). All items include negative stereotypes,

TABLE 1 Topic characteristics.

	M (SD)		
Variables	WLS stigma	PrEP stigma	Test statistics
Knowledge ($n = 761$, 7-pt. scale)	3.04 (1.67)	2.04 (1.62)	$t_{(759)} = -8.34$, $p < 0.001$, $d = 0.60$
Social comparison ($n = 762$, 10-pt. scale)	6.39 (1.05)	6.06 (1.16)	$t_{(760)} = -4.12$, $p < 0.001$, $d = 0.30$
Social distance ($n = 762$, 7-pt. scale)	4.36 (1.36)	4.32 (1.38)	$t_{(760)} = -0.36$, $p = 0.719$, $d = 0.03$
Stigma agreement ($n = 656$, 0/1)	0.20 (0.28)	0.31 (0.24)	Welch's $t_{(653)} = 5.62$, $p < 0.001$, $d = 0.44$
Warmth ($n = 656$, 0/1)	0.83 (0.30)	0.74 (0.32)	Welch's $t_{(634)} = -3.71$, $p < 0.001$, $d = 0.29$
Competence ($n = 656$, 0/1)	0.69 (0.32)	0.75 (0.30)	Welch's $t_{(653)} = 2.51$, $p < 0.001$, $d = 0.20$

TABLE 2 Scale items of stigma scales.

WLS stigma	PrEP stigma
People who have WLS lack willpower.[a]	People using PrEP lack willpower.
Having WLS is a sign of personal failure.	Using PrEP is a sign of personal failure.[b]
People who have WLS are looking for a magic bullet.[a]	People who take PrEP are looking for a magic bullet.
People who have WLS are lazy.[a]	People using PrEP cannot take personal responsibility.
People who have WLS want to eat excessively.	People using PrEP like to have sex with a lot of different people.[b]
People who have WLS have strange eating habits.	People using PrEP like to have strange types of sex.[b]
People who have WLS can't watch their weight.	People using PrEP can't use condoms.[b]
People that have WLS are cowards.[a]	People using PrEP are cowards.

[a] Origin WLS stigma scale (Hansen and Dye, 2018).
[b] Origin PrEP stigma scale (Klein and Washington, 2019).

including the lack of willpower, cowardice, and either laziness (WLS) or promiscuity (PrEP).

Measures

Dependent variables
Implicit stigma along the warmth and competence dimensions

Implicit stigma was measured using a reaction time (RT) task. Participants were asked to rate as quickly as possible whether they would attribute a specific adjective to a person who uses PrEP or has had weight loss surgery. Each adjective appeared in the middle of the screen, and participants were instructed to press the L key for "more likely applies" and D for "more likely does not apply". All adjectives were presented in random order.

Before the trial, participants were given a practice round of rating 10 adjectives regarding their applicability to a modern skyscraper building shown at the top of the screen throughout the trial. Five of those adjectives applied to the building, e.g., modern

and tall, and five did not, e.g., historical and windowless. In the main trial, participants were presented with 18 adjectives from three categories (see Table 3). The first category consisted of three adjectives that were part of the stigma scale items. The second category measured more generalized stigma using five adjectives for the warmth and competence dimensions of the stereotype content model (Fiske et al., 2002), respectively. In addition, a third category included stigma-unrelated adjectives of positive and negative valence as filler items.

Before analyses, reaction times were preprocessed according to established conventions (e.g., Ratcliff, 1993; Wittenbrink et al., 2001): In the first step, RTs faster than 150 ms and slower than 3,000 ms were excluded. They were then submitted to an inverse transformation and multiplied by 1,000. Consequently, higher values represent faster reaction times, thus, higher accessibility.

Participants had to be excluded for having no RT data ($n = 2$), no response variation in practice or main trials ($n = 32$), more than six incorrect responses in the practice round ($n = 32$), or not enough valid RT data to calculate indices ($n = 40$). The resulting sample size for analyses with implicit measures is $n = 656$ (53.1% female; $M_{age} = 49.4$, $SD_{age} = 16.2$; 70.6% with tertiary education). The excluded participants do not significantly differ from the final implicit sample on age, education, or any of the relevant study variables, i.e., social comparison, desired social distance, and prior knowledge (all $p > 0.19$). However, the percentage of men is significantly higher for the excluded participants, 61.3 vs. 46.9%., $\chi^2 = 7.62$, $df = 1$, $p = 0.006$. Therefore, we included gender as a covariate in all our analyses.

After data preparation, mean indices were calculated for each category, i.e., RT warmth ($M = 1.02$, $SD = 0.36$; Cronbach's $\alpha = 0.81$), and RT competence ($M = 0.96$, $SD = 0.35$; Cronbach's $\alpha = 0.82$). In addition, we calculated means for the agreement with these categories, to be able to determine whether it is stigma agreement (1) that is facilitated by stigma scale exposure or disagreement (0): warmth ($M = 0.79$, $SD = 0.31$; Cronbach's $\alpha = 0.83$), and competence ($M = 0.72$, $SD = 0.31$; Cronbach's $\alpha = 0.73$).

Agreement to the three items we selected from the stigma scales did not form a reliable index (Cronbach's $\alpha < 0.49$). We, therefore, analyze reaction times to and agreement with these items separately (weak-minded: mean RT = 0.84, $SD = 0.42$, agreement = 22.3%; cowardly: mean RT = 1.00, $SD = 0.48$, agreement = 14.2%; PrEP

TABLE 3 Adjectives for the reaction time task.

Specific stigma	Stereotype content model		Unrelated	
	Warmth	Competence	Positive	Negative
Weak-minded	Likable	Competent	Creative	Forgetful
Cowardly	Sincere	Confident	Musical	Mean
Promiscuous[a]/lazy[b]	Warm	Competitive	Polite	Promiscuous[b]/lazy[a]
	Good-natured	Independent		
	Tolerant	Intelligent		

[a] PrEP, Pre-exposure prophylaxis.
[b] WLS, Weight-loss surgery.

group: promiscuous: mean RT = 0.94, SD = 0.52, agreement = 64.1%; WLS group: lazy: mean RT = 1.02, SD = 0.53, agreement = 16.9%).

Downward social comparison

The Social Comparison Scale (Allan and Gilbert, 1995) was used to measure downward social comparison on a 10-pt bipolar scale from 1 indicating upward comparison to 10 indicating downward comparison. The 11 items, such as whether participants feel more attractive, superior, or confident compared to a person using PrEP/having WLS, were combined into a mean index (M = 6.20, SD = 1.09, Cronbach's α = 0.89).

Desired social distance

Four items based on Hoffner and Cohen (2015) and Bartsch et al. (2018) asked participants how desirable it is to spend an evening, be friends with, work closely with or be in a romantic relationship with a person using PrEP/having WLS. The items were measured on a 7-point scale and combined into a mean index (M = 4.34, SD = 1.37, Cronbach's α = 0.90).

Moderator variables
Prior knowledge

One item assessed participants' general knowledge of weight loss surgery or PrEP on a 7-point scale (M = 2.56, SD = 1.72).

Personal relevance

Participants reporting at least a little bit of knowledge were asked one item about the personal relevance of weight loss surgery (n = 314, M = 1.86, SD = 1.43) or PrEP (n = 139, M = 2.97, SD = 1.85).

Results

Scale exposure effects on implicit and explicit stigma

To test the effect of stigma scale exposure on implicit stigma (H1a), we conducted analyses of covariance (ANCOVA[4]) with the reaction times (RT) for the single stigma items as well as

---

4 Tukey's HSD was used to adjust for multiple comparisons.

the warmth and competence dimensions as dependent variables. Age, gender, and, to account for individual differences in reaction times, mean RTs during the practice round served as covariates. Results indicate significant main effects of exposure to a scale containing stigmatizing statements on reaction times for weak-minded [$F_{(1,616)}$ = 6.56, p = 0.011, part. η^2 = 0.01], and cowardly [$F_{(1,639)}$ = 10.29, p = 0.001, part. η^2 = 0.02]. Participants responded faster to these items when they were exposed to the scale (weak-minded: EM = 0.88, SE = 0.02; cowardly: EM = 1.05, SE = 0.02) compared to when not (weak-minded: EM = 0.79, SE = 0.02; cowardly: EM = 0.93, SE = 0.03). Scale exposure did not facilitate other reaction times, regarding neither the topic-specific items of promiscuous and lazy, nor the warmth and competence dimensions, all p > 0.105.

We used binary logistic regressions to assess the effect of stigma scale exposure on agreement with the single stigma items, and ANCOVAs for the warmth and competence indices. There were no effects of scale exposure, all p > 0.118, except for the topic-specific items, promiscuous and lazy. Interestingly, the effects occurred in the opposite direction: While participants were more likely to agree that PrEP users are promiscuous after they were exposed to the scale, b = 0.74, SE = 0.25, Z = 3.00, p = 0.003, Odds-Ratio = 2.10, 95% CI (1.29, 3.40), participants in the other group were less likely to agree that WLS patients are lazy b = −0.74, SE = 0.30, Z = −2.45, p = 0.014, Odds-Ratio = 0.48, 95% CI (0.27, 0.86). Effects on explicit stigma were analyzed using ANCOVAs with either downward social comparison or social distance as dependent variables (H1b-c). Besides the experimental factors as independent factors, we included age and gender as covariates. Results indicate no significant main effects of exposure to a scale containing stigmatizing statements, neither for social comparison [$F_{(1,754)}$ = 0.02, p = 0.896, part. η^2 = 0.000], nor for social distance [$F_{(1,754)}$ = 0.77, p = 0.379, part. η^2 = 0.001]. We did not find any interaction effects with the topic; for social comparison [$F_{(1,754)}$ = 0.79, p = 0.375, part. η^2 = 0.001], for social distance [$F_{(1,754)}$ = 2.35, p = 0.125, part. η^2 = 0.003].

Therefore, H1 was rejected. While scale exposure activated some concepts contained in the stigma scales, it had divergent effects on agreement with topic-specific stigma items. Scale exposure did not affect other components of stigma, such as the warmth and competence dimensions of the stereotype content model, social comparison, or desired social distance.

Moderating effects of prior knowledge and personal relevance

H2 postulated that an individual's knowledge *(H2a)* and personal relevance *(H2b)* of the topic moderate the effects of stigma scale exposure. Regarding the role of knowledge, no moderating effect of scale exposure's relationships with reaction times emerged; both overall and including interactions with topic, all $p > 0.201$. Similarly, there were no effects on agreement, all $p > 0.174$.

Concerning downward social comparison, we find a three-way interaction between scale exposure, topic, and knowledge [$F_{(1,749)}$ = 7.55, $p = 0.006$, part. $\eta^2 = 0.010$]. While exposure to a stigma scale about PrEP users increased downward social comparison for increasingly knowledgeable individuals, the opposite is observed for WLS stigma scale exposure. Specifically, the effects of scale exposure are only significant for highly knowledgeable individuals (see Figure 1). There was no moderation effect of knowledge for social distance; overall and including interactions with the topic, all $p > 0.680$.

Examining the moderating role of personal relevance, we only found significant interaction effects between importance and scale exposure for competence RTs [$F_{(1,380)} = 3.93$, $p = 0.048$, part. $\eta^2 = 0.010$], and agreement [$F_{(1,381)} = 5.64$, $p = 0.018$, part. $\eta^2 = 0.015$]; all other $p > 0.062$. Specifically, scale exposure positively affected reaction times and agreement with competence items for people to whom the topic was highly relevant and had a negative effect on agreement for those low in relevance (see Figure 2). In other words, participants to whom the topic was not personally relevant judged PrEP users/WLS patients as less competent when exposed to a stigma scale. If the topic was highly relevant to the participants, they responded faster to the competence items and judged PrEP users/WLS patients as more competent after scale exposure.

Personal relevance of the topic did not moderate the effects of scale exposure on social comparison, all $p > 0.092$, nor on desired social distance, all $p > 0.808$. *H2* was not supported. However, topic knowledge was found to moderate the effect of stigma scale exposure on social comparison, though in a direction opposing our assumptions and with effects only occurring for more knowledgeable participants. Similarly, personal relevance emerged as a moderator of scale exposure effects for competence reaction times and agreement, showing opposite effects for participants with low and high topic relevance.
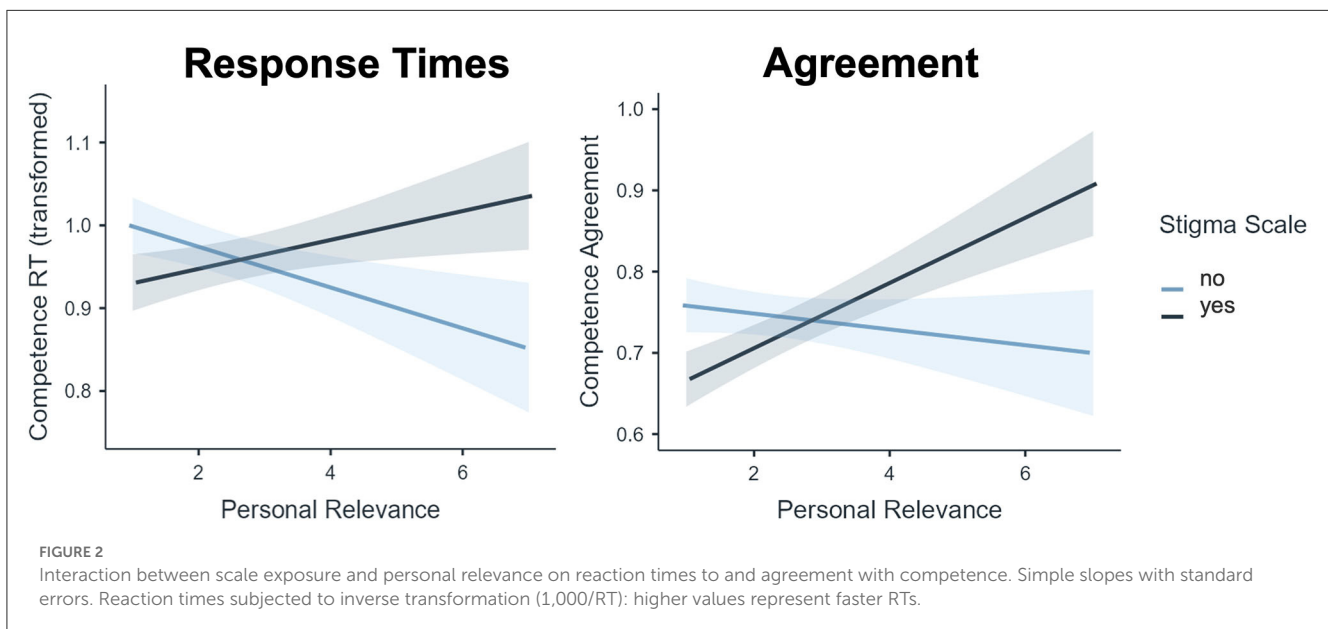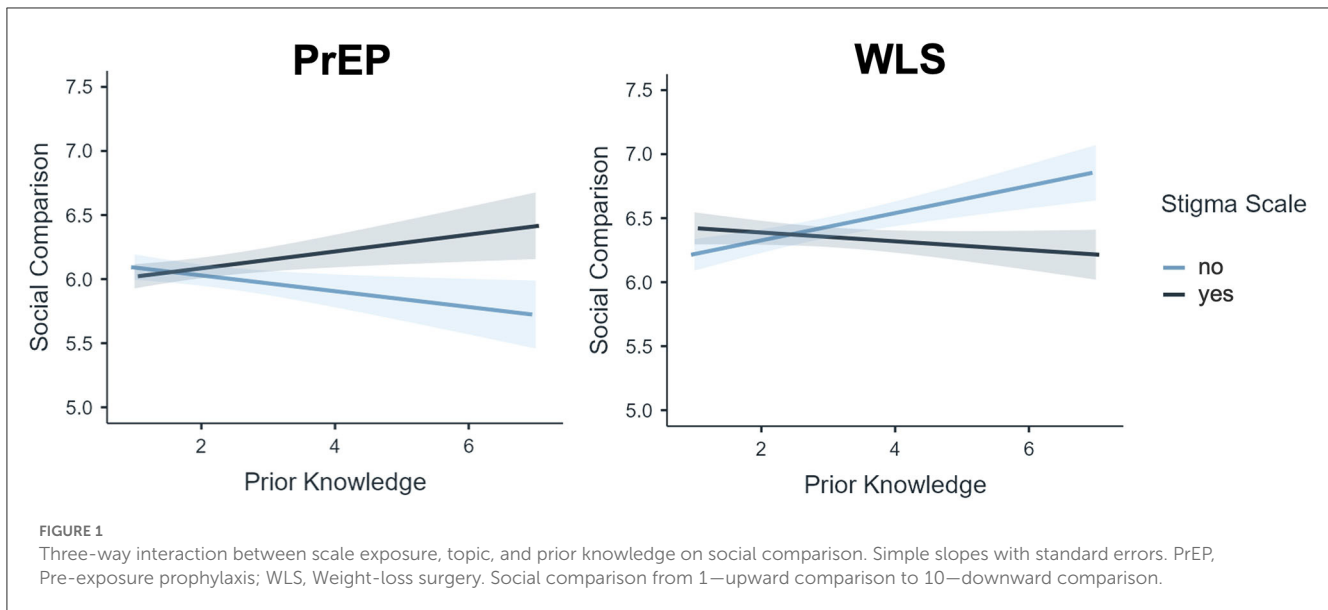
Discussion

This study investigated the potentially detrimental effects of employing self-report scales to measure public stigma in research projects using a general population sample. Within the contexts of WLS and PrEP we tested whether the use of scales that explicitly include negative stereotypes can promote stigmatizing tendencies in participants who are asked to respond to such statements. Based on theory and research in priming (Carpentier, 2020) and stereotype activation/application (Macrae and Bodenhausen, 2000), we assumed that stigma scale exposure will have a positive effect on indicators of stigmatization (i.e., implicit stigma, downward social comparison, and desired social distance). Moreover, we expected this effect to be moderated by individuals' knowledge about the

topic as well as their issue involvement. The results demonstrate that the underlying mechanisms of potential detrimental effects of stigma scale exposure are more complex and as multifaceted as the concept of stigma itself.

First, regarding general scale exposure effects on stigma *(H1)*, we found increased accessibility of stigma scale items (weak-minded and cowardly) after exposure. Even though there was no influence on participants' tendency to agree with those statements, the increased accessibility could strengthen the connection between these negative stereotypes and the target group if activated repeatedly over time. Of course, the accessibility effect on these two items should not be overrated: This type of priming effect was expected for those stigma adjectives contained in the scale whose exposure was manipulated. At the same time, there was no heightened accessibility effect for negative stereotypes most closely associated with the respective stigma (PrEP: promiscuous, WLS: lazy), but scale exposure did influence agreement with these stereotypes. Curiously, while it increased endorsement of the stigma for PrEP, scale exposure reduced endorsement for WLS patients. In terms of the lack of similar accessibility effects for weak-minded and cowardly, this might result from stronger existing associations between lazy/promiscuous and WLS patients/PrEP users that are less affected by one exposure. An explanation for the opposite effect of scale exposure on agreement with these negative stereotypes is more complex. The stigmatizing effect for PrEP users, which aligns with our hypothesis, might be rooted in the nature of the stigma, i.e., differences between the topics, with PrEP being less familiar and more stigmatized. The de-stigmatizing effect for WLS patients might result from more explicit processing, e.g., a form of impression management or even a protest against perceived stigma. These explanations are only speculative, and future research needs to disentangle such short-term accessibility effects of exposure to such statements, including longer-term accessibility and conditions for translation into an agreement. Besides effects on the directly related stigma items, scale exposure effects did not extend to more general dimensions of the stereotype content model, downward social comparison, or desired social distance.

However, the findings concerning *H2*, the moderating effects of prior knowledge and personal relevance, show that scale exposure may affect these variables depending on prior knowledge and personal relevance. Independent of the topic, participants to whom PrEP or WLS was less personally relevant judged PrEP users/WLS patients as less competent when previously being confronted with a stigma scale compared to when not. In contrast, individuals who reported greater personal relevance responded faster and judged PrEP users/WLS patients as more competent when exposed to the scale. Thus, the more relevant the topic was to the participants, the more positively they responded to a more generalized dimension of stigma under time pressure after scale exposure. Exposure to the stigma scale resulted in more stigmatized responses by participants with low relevance. While in line with hypothesis two, this negative effect for people who regard the topics as less relevant is quite concerning. Further research is needed to examine the underlying mechanisms.

Like the topic-specific effects for the stigma-specific items, lazy and promiscuous, we found different effects of scale exposure for different levels of prior knowledge about the topic on social comparison as a dependent variable. In contrast to the

FIGURE 1
Three-way interaction between scale exposure, topic, and prior knowledge on social comparison. Simple slopes with standard errors. PrEP, Pre-exposure prophylaxis; WLS, Weight-loss surgery. Social comparison from 1—upward comparison to 10—downward comparison.

FIGURE 2
Interaction between scale exposure and personal relevance on reaction times to and agreement with competence. Simple slopes with standard errors. Reaction times subjected to inverse transformation (1,000/RT): higher values represent faster RTs.

expected stronger effects for individuals with less knowledge, exposure to a stigma scale about PrEP users increased downward social comparison for increasingly knowledgeable individuals, while the opposite is observed for WLS stigma scale exposure. The findings for WLS correspond to the de-stigmatizing effects of scale exposure on the topic-specific stereotype and point to the potential that a more consolidated attitude makes the negative influence of exposure to stereotyping content less likely but rather leads to the opposite effect. Whether that is a genuinely positive effect or the result of short-term impression management processes or protesting is a question for future research. An even more puzzling question to be addressed by future studies is the detrimental effect of scale exposure on participants with high knowledge of PrEP. A very preliminary, speculative explanation could be that there are self-stigmatizing

processes activated by scale exposure, as participants with very high knowledge of PrEP are more likely to be part of the targeted stigma group. Again, more research is needed to disentangle these relationships.

In light of our findings, it is crucial to consider the wider practical implications for anti-stigma communication, particularly regarding the use of negative stereotypes in interventions such as public health campaigns. Some campaigns employ stereotypes as a strategy to convey the message that labeling individuals in certain ways, such as dangerous (in the case of mental disorders) or lazy (pertaining to overweight and obesity), is both morally wrong and socially unacceptable. However, our study reveals that this approach might lead to unintended effects and emphasizes the need for careful consideration and sensitivity when employing stereotypes in anti-stigma interventions.

Limitations

There are limitations of this study that need to be addressed. This includes the single-exposure nature of the study design. As such, we cannot be certain whether the emerging short-term effects also translate into long-term effects or even accumulate, especially when participants are repeatedly confronted with such statements, i.e., in longitudinal investigations or other settings, such as on social media. Therefore, especially longitudinal designs with measurements weeks or months after the main study are needed to examine whether scale exposure effects are persistent.

Moreover, even though using the panel service to distribute the survey provides more heterogeneity than regular student samples, it must be noted that the sample is still a convenience sample skewed to higher education. Future studies should recruit samples with more varied levels of education.

Our study was conducted within the context of two topics with their respective stigma contexts and the question remains whether the findings can be replicated for other topics. Of note, by choosing two topics, we aimed to increase the generalizability of the underlying effects. However, the topic-specific findings show that more factors need to be considered when investigating the potential detrimental effects of self-report scales measuring stigma beliefs. Thus, it is relevant for future research to systematically consider topic-specific or stigma context-specific characteristics.

Relatedly, our attempt to streamline the specific stigma items to make the self-report scales of stigma beliefs comparable across the two topics and avoid confounding due to differences in scale composition might have resulted in a less specific fit of the applied stigma scales and the low reliability for the stigma items in the implicit measure task. The decision to use single-item dependent variables for the stigma adjectives is not ideal. On a more general note about our implicit measure, two aspects need to be discussed, its validity as a measure of implicit stigma generally and the exploratory nature of our measure. In the methodological literature there is a continued debate about the validity of implicit measures, especially in terms of what they measure (De Houwer et al., 2009; Gawronski and Hahn, 2019). For example, in our case, reaction times for a decision under time pressure of whether or not an adjective applies or not are first and foremost an assessment of how long it takes participants to make that decision. However, by comparing two groups of participants randomly assigned to the experimental conditions, differences between participants exposed to the stigma scale and those not can be interpreted regarding the accessibility of the association between the adjective and the target group. Regarding the agreement decision, this is likely not an implicit measure in the pure sense of automaticity (De Houwer et al., 2009), because participants are aware of the nature of the judgment. However, due to the time pressure under which they are asked to make their decision, their responses are less controlled and more likely to represent their "gut reaction" to the task. At the same time, we need to emphasize that we have created this measure for this study, and it needs to be validated and refined, especially with respect to the reliability issues mentioned above. This underlines the need for future research to develop different sensitive measures to tap into implicit stigma.

Our research aimed to investigate the potential negative effects of standardized stigma scales, including items associated with negative stereotypes. In this context, it is important to consider the potentially harmful impacts of other measures used in this study, such as social comparison, social distance, and the reaction time task. We argue that these three measures do not meet the criteria of stigma messages as defined by Smith (2007) because they lack explicit statements connecting the stigmatized group to negative stereotypes. Moreover, these measures include either exclusively positive items or a combination of positive and negative items. Finally, all three measures are not exclusive to stigma research but can be used to assess perceptions of any group or individual.

However, it is still possible that, within the context of stigma research, a question regarding whether one can imagine being friends with or having a romantic relationship with someone from a stigmatized group may trigger a thought process leading to a desire for greater distance from that person. Nevertheless, it is practically impossible to eliminate all stigmatizing effects of every measure for all participants. The importance of comprehensive debriefing procedures in stigma research cannot be overstated. Nonetheless, it is essential to identify and address unintended effects that are more substantial and consistent if we discover them.

Conclusion

In conclusion, this study is the first exploratory attempt at identifying potential adverse effects of utilizing stigma scales assessing negative stereotypes in quantitative studies of public stigma in the general population. Overall, our results underline the relevance of the issue by showing that being confronted with judgmental or stereotypical items on a stigma scale—even only for a short time—can facilitate scale-related stereotype accessibility, both increasing and decreasing stigmatizing tendencies. Specifically, scale exposure increased endorsement of a negative stereotype for PrEP, increased downward social comparison for individuals highly knowledgeable about PrEP, and reduced perceptions of competence for individuals with low personal relevance for both topics. At the same time, scale exposure reduced endorsement of a negative stereotype of WLS, increased upward social comparison for individuals highly knowledgeable about WLS, and increased perceptions of competence for high personal relevance. In short, stigma scale exposure can have both adverse and positive effects on stigmatizing tendencies. Future research must replicate and extend these results to determine the range and strength of significant adverse effects. A systematic investigation would allow the examination of the mechanisms and conditions under which such detrimental effects occur and how they can be avoided.

At the same time, researchers need to discuss and develop alternative ways of measurement or strategies to avoid or counter any stigmatizing effects of self-report scales assessing stigma beliefs. For example, scales could be constructed with only positive wordings. Alternatively, specific debriefing methods might be developed and tested to counter any negative effects of scale exposure. Our exploratory study is a first step to raise awareness and contribute to ethical research practices that avoid

unintended negative consequences of exacerbating the stigma under examination.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Ethikbeirat Universität Leipzig, 2021.04.01_eb_82. The patients/participants provided their written informed consent to participate in this study.

Author contributions

AO and FS contributed to the conception, design of the study, secured data access, performed the statistical analysis, wrote the first draft of the manuscript, and contributed to the manuscript revision. Both authors read and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Allan, S., and Gilbert, P. (1995). A social comparison scale: psychometric properties and relationship to psychopathology. *Personal. Indiv. Diff.* 19, 293–299. doi: 10.1016/0191-8869(95)00086-L

Bargh, J. A. (1999). "The cognitive monster: the case against the controllability of automatic stereotype effects," in *Dual-Process Theories in Social Psychology*, eds. S. Chaiken and Y. Trope (New York, NY: The Guilford Press), 361–382.

Bartsch, A., Oliver, M. B., Nitsch, C., and Scherr, S. (2018). Inspired by the paralympics: effects of empathy on audience interest in para-sports and on the destigmatization of persons with disabilities. *Commun. Res.* 45, 525–553. doi: 10.1177/0093650215626984

Berkowitz, L. (1984). Some effects of thoughts on anti- and prosocial influences of media events: a cognitive-neoassociation analysis. *Psychol. Bull.* 95, 410–427. doi: 10.1037/0033-2909.95.3.410

Bos, A. E. R., Pryor, J. B., Reeder, G. D., and Stutterheim, S. E. (2013). Stigma: advances in theory and research. *Basic Appl. Soc. Psychol.* 35, 1–9. doi: 10.1080/01973533.2012.746147

Bresnahan, M., and Zhuang, J. (2011). Exploration and validation of the dimensions of stigma. *J. Health Psychol.* 16, 421–429. doi: 10.1177/1359105310382583

Carpentier, F. R. D. (2020). Priming. *Int. Encyclop. Media Psychol.* 1–14. doi: 10.1002/9781119011071.iemp0069

Collins, A. M., and Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychol. Rev.* 82, 407–428. doi: 10.1037/0033-295X.82.6.407

Corrigan, P. W., Gause, M., Michaels, P. J., Buchholz, B. A., and Larson, J. E. (2015). The California assessment of stigma change: a short battery to measure improvements in the public stigma of mental illness. *Commun. Ment. Health J.* 51, 635–640. doi: 10.1007/s10597-014-9797-5

Corrigan, P. W., Morris, S. B., Michaels, P. J., Rafacz, J. D., and Rüsch, N. (2012). Challenging the public stigma of mental illness: a meta-analysis of outcome studies. *Psychiatr. Serv.* 63, 963–973. doi: 10.1176/appi.ps.201100529

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., and Moors, A. (2009). Implicit measures: a normative analysis and review. *Psychological. Bull.* 135, 347–368. doi: 10.1037/a0014211

Devine, P. G. (1989). Stereotypes and prejudice: their automatic and controlled components. *J. Pers. Soc. Psychol.* 56, 5–18. doi: 10.1037/0022-3514.56.1.5

Devine, P. G., and Sharp, L. B. (2009). "Automaticity and control in stereotyping and prejudice," in *Handbook of Prejudice, Stereotyping, and Discrimination*, ed. T. D. Nelson (New York, NY: Psychology Press), 61–87.

Dubov, A., Galbo, P. Jr, Altice, F. L., and Fraenkel, L. (2018). Stigma and shame experiences by MSM who take PrEP for HIV prevention: a qualitative study. *Am. J. Mens. Health* 12, 1843–1854. doi: 10.1177/1557988318797437

Faul, F., Erdfelder, E., Buchner, A., and Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behav. Res. Methods* 41, 1149–1160. doi: 10.3758/BRM.41.4.1149

Fiske, S. T., Cuddy, A. J. C., Glick, P., and Xu, J. (2002). A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. *J. Person. Soc. Psychol.* 82, 878–902. doi: 10.1037/0022-3514.82.6.878

Fox, A. B., Earnshaw, V. A., Taverna, E. C., and Vogt, D. (2018). Conceptualizing and measuring mental illness stigma: the mental illness stigma framework and critical review of measures. *Stigma Health* 3, 348–376. doi: 10.1037/sah0000104

Frost, D. (2011). Social stigma and its consequences for the socially stigmatized. *Soc. Person. Psychol. Comp.* 5, 824–839. doi: 10.1111/j.1751-9004.2011.00394.x

Gabbidon, K., and Chenneville, T. (2021). Strategies to minimize further stigmatization of communities experiencing stigma: a guide for qualitative researchers. *Stigma Health* 6, 32–42. doi: 10.1037/sah0000280

Gawronski, B., and Hahn, A. (2019). "Implicit measures: procedures, use, and interpretation," in *Measurement in Social Psychology*, eds H. Blanton, J. M. LaCroix, and G. D. Webster (New York, NY: Routledge; Taylor and Francis Group), 29–55.

Goffman, E. (1963). *Stigma: Notes on the Management of Spoiled Identity*. Englewood Cliffs, NJ: Prentice Hall.

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., and Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychol. Rev.* 109, 3–25. doi: 10.1037/0033-295X.109.1.3

Hansen, B., and Dye, M. H. (2018). Damned if you do, damned if you don't: the stigma of weight loss surgery. *Deviant Behav.* 39, 137–147. doi: 10.1080/01639625.2016.1263081

Hatzenbuehler, M. L., Phelan, J. C., and Link, B. G. (2013). Stigma as a fundamental cause of population health inequalities. *Am. J. Public Health* 103, 813–821. doi: 10.2105/AJPH.2012.301069

Hilton, J. L., and von Hippel, W. (1996). Stereotypes. *Ann. Rev. Psychol.* 47, 237–271. doi: 10.1146/annurev.psych.47.1.237

Hoffner, C. A., and Cohen, E. L. (2015). Portrayal of mental illness on the tv series monk: presumed influence and consequences of exposure. *Health Commun.* 30, 1046–1054. doi: 10.1080/10410236.2014.917840

Iyengar, S., Peters, M. D., and Kinder, D. R. (1982). Experimental demonstrations of the "not-so-minimal" consequences of television news programs. *Am. Polit. Sci. Rev.* 76, 848–858. doi: 10.2307/1962976

Jones, E. E., Farina, A., Hastorf, A., Markus, H., Miller, D. T., and Scott, R. (1984). *Social Stigma: The Psychology of Marked Relationships*. New York, NY: W. H. Freeman.

Klein, H., and Washington, T. A. (2019). The pre-exposure prophylaxis (PrEP) stigma scale: preliminary findings from a pilot study. *Int. Public Health J.* 11, 185–195.

Knecht, T., and Martinez, L. M. (2009). Humanizing the homeless: does contact erode stereotypes? *Soc. Sci. Res.* 38, 521–534. doi: 10.1016/j.ssresearch.2009.01.009

Kunda, Z., and Sinclair, L. (1999). Motivated reasoning with stereotypes: activation, application, and inhibition. *Psychol. Inq.* 10, 12–22. doi: 10.1207/s15327965pli1001_2

Leiner, D. (2016). Convenience samples from online respondent pools: A case study of the SoSci Panel. *Stud. Commun. Media* 5, 367–396. doi: 10.5771/2192-4007-2016-4

Link, B. G., and Phelan, J. C. (2001). Conceptualizing stigma. *Ann. Rev. Sociol.* 27, 363–385. doi: 10.1146/annurev.soc.27.1.363

Link, B. G., Phelan, J. C., and Hatzenbuehler, M. L. (2014). "Stigma and social inequality," in *Handbook of the Social Psychology of Inequality Handbooks of Sociology and Social Research*, eds. J. D. McLeod, E. J. Lawler, and M. Schwalbe (Dordrecht: Springer), 49–64.

Link, B. G., Yang, L. H., Phelan, J. C., and Collins, P. Y. (2004). Measuring mental illness stigma. *Schizophr. Bull.* 30, 511–541. doi: 10.1093/oxfordjournals.schbul.a007098

Macrae, C. N., and Bodenhausen, G. V. (2000). Social cognition: thinking categorically about others. *Annu. Rev. Psychol.* 51, 93–120. doi: 10.1146/annurev.psych.51.1.93

Major, B., and O'Brien, L. T. (2005). The social psychology of stigma. *Ann. Rev. Psychol.* 56, 393–421. doi: 10.1146/annurev.psych.56.091103.070137

Millum, J., Campbell, M., Luna, F., Malekzadeh, A., and Karim, Q. A. (2019). Ethical challenges in global health-related stigma research. *BMC Med.* 17, 84. doi: 10.1186/s12916-019-1317-6

Pescosolido, B. A., and Martin, J. K. (2015). The stigma complex. *Annu. Rev. Sociol.* 41, 87–116. doi: 10.1146/annurev-soc-071312-145702

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychol. Bull.* 114, 510–532. doi: 10.1037/0033-2909.114.3.510

Rees, H. R., Rivers, A. M., and Sherman, J. W. (2019). Implementation intentions reduce implicit stereotype activation and application. *Pers. Soc. Psychol. Bull.* 45, 37–53. doi: 10.1177/0146167218775695

Roehrig, J. P., and McLean, C. P. (2010). A comparison of stigma toward eating disorders versus depression. *Int. J. Eat. Disord.* 43, 671–674. doi: 10.1002/eat.20760

Saeed, F., Mihan, R., Mousavi, S. Z., Reniers, R. L., Bateni, F. S., Alikhani, R., et al. (2020). A narrative review of stigma related to infectious disease outbreaks: what can be learned in the face of the Covid-19 pandemic? *Front. Psychiatry* 11, e565919. doi: 10.3389/fpsyt.2020.565919

Smith, R. A. (2007). Language of the lost: an explication of stigma communication. *Commun. Theory* 17, 462–485. doi: 10.1111/j.1468-2885.2007.00307.x

Stangl, A. L., Earnshaw, V. A., Logie, C. H., van Brakel, W., C., Simbayi, L., et al. (2019). The health stigma and discrimination framework: a global, crosscutting framework to inform research, intervention development, and policy on health-related stigmas. *BMC Med.* 17, 31. doi: 10.1186/s12916-019-1271-3

Tulving, E. (1972). "Episodic and semantic memory," in *Organization of Memory*, eds. E. Tulving and W. Donaldson (New York, NY: Academic Press), 381–402.

Wittenbrink, B., Judd, C. M., and Park, B. (2001). Evaluative versus conceptual judgments in automatic stereotyping and prejudice. *J. Exp. Soc. Psychol.* 37, 244–252. doi: 10.1006/jesp.2000.1456

Zhu, X., Smith, R. A., and Buteau, E. (2022). A meta-analysis of weight stigma and health behaviors. *Stigma Health* 7, 1–13. doi: 10.1037/sah0000352