Citation for published version (APA):
Fusar-Poli, P. (Accepted/In press). Can we increase the subjective well-being of the general population? An umbrella review of the evidence. Revista de Psiquiatría y Salud Mental.
Revista de Psiquiatría y Salud Mental
¿Podemos incrementar el bienestar subjetivo de la población en general? Una revisión paraguas de la evidencia
Can we increase the subjective well-being of the general population? An umbrella review of the evidence
--Borrador del manuscrito--

<table>
<thead>
<tr>
<th>Número del manuscrito:</th>
<th>RPM-D-20-00070R2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tipo de artículo:</td>
<td>Review</td>
</tr>
<tr>
<td>Palabras clave:</td>
<td>Actos de bondad, metaanálisis, psicología positiva, ensayo controlado aleatorizado, bienestar subjetivo, revisión paraguas Acts of kindness, meta-analysis, positive psychology, randomized controlled trial, subjective well-being</td>
</tr>
<tr>
<td>Autor correspondiente:</td>
<td>Joaquim Radua, MD PhD Institut d'Investigacions Biomediques August Pi i Sunyer Barcelona, Barcelona SPAIN</td>
</tr>
<tr>
<td>Primer autor:</td>
<td>Aleix Solanes</td>
</tr>
<tr>
<td>Orden de autores:</td>
<td>Aleix Solanes</td>
</tr>
<tr>
<td></td>
<td>Anton Albajes-Eizagirre</td>
</tr>
<tr>
<td></td>
<td>Miquel A Fullana, PhD</td>
</tr>
<tr>
<td></td>
<td>Lydia Fortea</td>
</tr>
<tr>
<td></td>
<td>Paolo Fusar-Poli, MD PhD</td>
</tr>
<tr>
<td></td>
<td>Carla Torrent, PhD</td>
</tr>
<tr>
<td></td>
<td>Brisa Solé, PhD</td>
</tr>
<tr>
<td></td>
<td>Caterina Mar Bonnín, PhD</td>
</tr>
<tr>
<td></td>
<td>Jae Il Shin, MD PhD</td>
</tr>
<tr>
<td></td>
<td>Eduard Vieta, MD PhD</td>
</tr>
<tr>
<td></td>
<td>Joaquim Radua, MD PhD</td>
</tr>
</tbody>
</table>

Resumen:
Introducción: El bienestar subjetivo (SWB) se refiere a estar satisfecho con la vida, tener afecto positivo y tener poco afecto negativo. Podemos entenderlo como una definición subjetiva de la buena vida, o en términos coloquiales, como "felicidad", y se ha asociado con varios beneficios importantes, como una menor mortalidad. En las últimas décadas, varios ensayos controlados aleatorizados (RCT) han investigado la eficacia de varias intervenciones para aumentar el SWB en la población general, pero los resultados de las diferentes disciplinas no se han integrado.
Métodos: Realizamos una revisión paraguas de revisiones sistemáticas y metaanálisis de RCT que evalúan la eficacia de cualquier tipo de intervención para aumentar el SWB en la población general, incluidas tanto las intervenciones de psicología positiva (PPI) como otras intervenciones. (Re)cálculos los estadísticos metaanalíticos necesarios para evaluar objetivamente la calidad de la evidencia de la eficacia cada tipo de intervención para mejorar cada componente del SWB de acuerdo con el Grading of Recommendations Assessment, Development and Evaluation (GRADE).
Resultados: Hubo evidencia de calidad moderada de que las PPI podrían inducir pequeñas disminuciones de afecto negativo y evidencia de baja calidad de que podrían inducir aumentos moderados de afecto positivo. Encontramos resultados similares para aquellas PPI que consistían específicamente en realizar actos de bondad (especialmente gastar dinero en o dar artículos a otros), para las cuales había evidencia de baja calidad de que podrían inducir pequeños aumentos de satisfacción con la vida, pero no para las PPI que consistían específicamente en practicar la gratitud. La calidad de la evidencia de la eficacia para las otras intervenciones incluidas en la revisión paraguas (yoga, entrenamiento de resiliencia, actividad física,
Introduction: Subjective well-being (SWB) refers to being satisfied with one’s life, having positive affect and having little negative affect. We may understand it as a subjective definition of good life, or in colloquial terms “happiness”, and it has been associated with several important benefits such as lower mortality. In the last decades, several randomized controlled trials (RCT) have investigated the efficacy of several interventions in increasing SWB in the general population but results from different disciplines have not been integrated.

Methods: We conducted an umbrella review of systematic reviews and meta-analyses of RCT that assess the efficacy of any kind of interventions in increasing SWB in the general population, including both positive psychology interventions (PPI) and other interventions. We (re)calculated the meta-analytic statistics needed to objectively assess the quality of the evidence of the efficacy of each type of intervention in improving each component of SWB according to the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach.

Results: There was moderate-quality evidence that PPI might induce small decreases of negative affect, and low-quality evidence that they might induce moderate increases of positive affect. We found similar results for those PPI specifically consisting in conducting acts of kindness (especially spending money on or giving items to others), for which there was low-quality evidence that they might induces small increases of life satisfaction, but not for PPI specifically consisting in practicing gratitude. Quality of the evidence of the efficacy for the other interventions included in the umbrella review (yoga, resilience training, physical activity, leisure, control enhancement, psychoeducation, and miscellaneous) was very low.

Conclusion: There is some evidence that PPI, and specially conducting acts of kindness such as spending money on others, may increase the SWB of the general population. The quality of the evidence of the efficacy for other interventions (e.g., yoga, physical activity, or leisure) is still very low.

Revisores sugeridos:

Lara Aknin  
Simon Fraser University, Canada  
laknin@sfu.ca  
Expert in positive psychology interventions

Lalin Anik  
Duke University, USA  
lalin.anik@duke.edu  
Expert in positive psychology interventions

Seth Margolis  
University of California Riverside, USA  
sethmmargolis@gmail.com  
Expert in positive psychology interventions

Michael Norton  
Harvard Business School, USA  
mnorton@hbs.edu  
Expert in positive psychology interventions

Elizabeth Dunn  
University of British Columbia, Canada  
edunn@psych.ubc.ca  
Expert in positive psychology interventions

Eva Walther  
University of Trier, Germany  
walther@uni-trier.de  
Expert in positive psychology interventions

Richard Ryan  
Australian Catholic University, Australia  
richard.ryan@acu.edu.au
Respuesta a los revisores:

REVIEWER #2

Thank you for your thorough review of the paper and for addressing all my concerns. All my comments have been answered, and I think the paper has improved significantly overall. I have some minor comments:

RESPONSE: Thank you very much again for the review, we have addressed all your remaining comments.

1. I understand this might be a reviewer's suggestion, but meta-analysis or reviews do not usually include hypotheses. The objectives are the exploration and summary of the existing data and studies. I therefore encourage authors to delete the sentence: "Our hypothesis was that some of the interventions are efficacious in increasing SWB in the general population, and that these interventions might be both PPI and other interventions."

RESPONSE: We have deleted this sentence.

2. Some of the new sections seem a little bit informal in the writing style. I recommend reviewing sentences and expressions such as: "(...) while at the end of the day an intervention could increase the SWB (...)"; or "(...) different meta-analyses use different methods up to the point that (...)".

RESPONSE: Thank you for the suggestion, we have removed "at the end of the day", and we have replaced "different methods up to the point that two meta-analyses…” by “different methods so that two meta-analyses…”.

3. Some references in the "Search strategy and eligibility criteria" section follow a different format than the rest (e.g. "Lyubomirsky et al (2011)").

RESPONSE: Thank you for noting this typo, we have addressed it.

4. When summarizing the different interventions included in the meta-analysis of Bolier et al., some of them are very similar (they would be considered under the same type of intervention: best possible self): best possible selves exercise, positive future thinking, projecting a positive self in the future, visualizing positive self ("best possible self" studies usually refer to King, 2001 and/or Sheldon & Lyubomirsky, 2006).

RESPONSE: We have grouped these interventions into “best possible self”.

5. Even though it may end up being repetitive, it might help the readers if the authors made references to Table 1 when describing the results of the "other interventions" section.

RESPONSE: We have added the references to Table 1 in the "other interventions" section of results.
Can we increase the subjective well-being of the general population?

An umbrella review of the evidence

Aleix Solanes¹,²,³, Anton Albajes-Eizagirre¹,², Miquel A. Fullana¹,²,⁴, Lydia Fortea¹, Paolo Fusar-Poli⁵,⁶,⁷, Carla Torrent²,⁴,⁸,⁹, Brisa Solé²,⁴,⁸,⁹, Caterina Mar Bonnín²,⁴,⁸,⁹, Jae Il Shin¹⁰, Eduard Vieta²,⁴,⁸,⁹,*, Joaquim Radua¹,²,⁵,¹¹,*,+

¹ Imaging of Mood- and Anxiety-Related Disorders (IMARD) group, Institut d’Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain
² Mental Health Research Networking Center (CIBERSAM), Madrid, Spain
³ Department of Psychiatry and Forensic Medicine, Universitat Autònoma de Barcelona, Barcelona, Spain
⁴ Early Psychosis: Interventions and Clinical-detection (EPIC) lab, Department of Psychosis Studies, Institute of Psychiatry, Psychology & Neuroscience, King’s College London, London, UK
⁵ Department of Brain and Behavioral Sciences, University of Pavia, Pavia, Italy
⁶ OASIS service, South London and Maudsley NHS Foundation Trust, London, UK
⁷ Department of Pediatrics, Yonsei University College of Medicine, Seoul, Republic of Korea
⁸ Bipolar and depressive disorders, Institut d’Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain
⁹ Universitat de Barcelona, Casanova, Barcelona, Spain
¹⁰ Centre for Psychiatric Research and Education, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden
¹¹ Centre for Psychiatric Research and Education, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden

* Shared senior authorship.

Correspondence to:
Joaquim Radua
Institut d’Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS)
Carrer del Rosselló 149
08036 Barcelona, Spain
Email: radua@clinic.cat
ORCID: 0000-0003-1240-5438
Ética de la publicación

1. ¿Su trabajo ha comportado experimentación en animales?:
   No

2. ¿En su trabajo intervienen pacientes o sujetos humanos?:
   No

3. ¿Su trabajo incluye un ensayo clínico?:
   No

4. ¿Todos los datos mostrados en las figuras y tablas incluidas en el manuscrito se recogen en el apartado de resultados y las conclusiones?:
   Sí
Can we increase the subjective well-being of the general population?

An umbrella review of the evidence

Abstract

**Introduction:** Subjective well-being (SWB) refers to being satisfied with one’s life, having positive affect and having little negative affect. We may understand it as a subjective definition of good life, or in colloquial terms “happiness”, and it has been associated with several important benefits such as lower mortality. In the last decades, several randomized controlled trials (RCT) have investigated the efficacy of several interventions in increasing SWB in the general population but results from different disciplines have not been integrated.

**Methods:** We conducted an umbrella review of systematic reviews and meta-analyses of RCT that assess the efficacy of any kind of interventions in increasing SWB in the general population, including both positive psychology interventions (PPI) and other interventions. We (re)calculated the meta-analytic statistics needed to objectively assess the quality of the evidence of the efficacy of each type of intervention in improving each component of SWB according to the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach.

**Results:** There was moderate-quality evidence that PPI might induce small decreases of negative affect, and low-quality evidence that they might induce moderate increases of positive affect. We found similar results for those PPI specifically consisting in conducting acts of kindness (especially spending money on or giving items to others), for which there was low-quality evidence that they might induces small increases of life satisfaction, but not for PPI specifically consisting in practicing gratitude. Quality of the evidence of the efficacy for the other interventions included in the umbrella review (yoga, resilience training, physical activity, leisure, control enhancement, psychoeducation, and miscellaneous) was very low.

**Conclusion:** There is some evidence that PPI, and specially conducting acts of kindness such as spending money on others, may increase the SWB of the general population. The quality of the evidence of the efficacy for other interventions (e.g., yoga, physical activity, or leisure) is still very low.

**Registration number:** PROSPERO CRD42020111681.

**Keywords:** acts of kindness, meta-analysis, positive psychology, randomized controlled trial, subjective well-being, umbrella review
¿Podemos incrementar el bienestar subjetivo de la población en general?

Una revisión paraguas de la evidencia

Resumen

Introducción: El bienestar subjetivo (SWB) se refiere a estar satisfecho con la vida, tener afecto positivo y tener poco afecto negativo. Podemos entenderlo como una definición subjetiva de la buena vida, o en términos coloquiales, como "felicidad", y se ha asociado con varios beneficios importantes, como una menor mortalidad. En las últimas décadas, varios ensayos controlados aleatorizados (RCT) han investigado la eficacia de varias intervenciones para aumentar el SWB en la población general, pero los resultados de las diferentes disciplinas no se han integrado.

Métodos: Realizamos una revisión paraguas de revisiones sistemáticas y metaanálisis de RCT que evalusen la eficacia de cualquier tipo de intervención para aumentar el SWB en la población general, incluidas tanto las intervenciones de psicología positiva (PPI) como otras intervenciones. (Re)calculamos los estadísticos metaanalíticos necesarios para evaluar objetivamente la calidad de la evidencia de la eficacia cada tipo de intervención para mejorar cada componente del SWB de acuerdo con el Grading of Recommendations Assessment, Development and Evaluation (GRADE).

Resultados: Hubo evidencia de calidad moderada de que las PPI podrían inducir pequeñas disminuciones de afecto negativo y evidencia de baja calidad de que podrían inducir aumentos moderados de afecto positivo. Encontramos resultados similares para aquellas PPI que consistían específicamente en realizar actos de bondad (especialmente gastar dinero en o dar artículos a otros), para las cuales había evidencia de baja calidad de que podrían inducir pequeños aumentos de satisfacción con la vida, pero no para las PPI que consistían específicamente en practicar la gratitud. La calidad de la evidencia de la eficacia para las otras intervenciones incluidas en la revisión paraguas (yoga, entrenamiento de resiliencia, actividad física, ocio, mejora del control, psicoeducación y miscelánea) fue muy baja.

Conclusión: Existe alguna evidencia de que las PPI, y especialmente la realización de actos de bondad como gastar dinero en otros, pueden aumentar el SWB de la población general. La calidad de la evidencia de la eficacia para otras intervenciones (p. ej., yoga, actividad física u ocio) sigue siendo muy baja.

Número de registro: PROSPERO CRD42020111681.

Palabras clave: actos de bondad, metaanálisis, psicología positiva, ensayo controlado aleatorizado, bienestar subjetivo, revisión paraguas
Introduction

Ed Diener defined subjective well-being (SWB) as an overall satisfaction with one’s life, having many pleasant emotions and moods (positive affect), and having few unpleasant ones (negative affect). It may be understood as a subjective definition of good life, or colloquially as “happiness”, but we will avoid these terms because they may have other meanings that are out of the scope of the present work. In the following we will refer to the definition of SWB presented above, known as the Diener tripartite model.

A high SWB has been associated with several important positive outcomes, such as substantially lower risk of exercise stress-induced myocardial ischemia (logistic regression-derived odds ratio = 0.55) or lower mortality (hazard ratio = 0.92 when comparing the group with highest SWB with the group with lowest SWB). SWB has also been linked to higher self-esteem and self-efficacy, and it is a relevant component of mental health. Some abnormalities of SWB (e.g., depressed mood and anxiety as extreme forms of few pleasant emotions and many unpleasant emotions) are core symptoms of mood/anxiety disorders (the most common mental/psychiatric disorders) and a component of many other mental/psychiatric disorders (e.g., schizophrenia is characterised by decreased ability to experience pleasure). Lower SWB has indeed been associated to higher suicidal ideation and behavior.

SWB consists of three main components: positive affect, negative affect and life satisfaction. Positive affect refers to pleasant emotions and moods, such as interest, enthusiasm or pride, and negative affect to unpleasant ones, such as guilt, irritability or fear. Life satisfaction refers to thinking that one’s life is close to one’s ideal, that the important things one wants in life have been already achieved. Positive and negative affect are sometimes combined in an “affective” component to distinguish them from the “cognitive” component (life satisfaction). All components show a weak to moderate correlation, for what the improvement of one component (e.g., having more positive emotions) may be accompanied by improvements of other components (e.g., feeling more satisfied). However, this should not be taken for granted. There are common exceptions where individuals with “many” pleasant emotions also have “many” unpleasant emotions or feel unsatisfied. Or similarly, individuals with a depressive episode (associated with high negative affect and low positive affect) may also feel “high” (which should be considered as positive affect), known as mixed features. It is thus more desirable to study each component separately.

Individuals with mental disorders could receive interventions that increase SWB but the promotion of mental health should also include interventions that increase SWB in healthy individuals. Note that the aims of mental
health prevention interventions include targeting risk factors and strengthen abilities to prevent the development of one or more conditions, while mental health promotion interventions aim to promote psychological wellbeing, increase the ability to achieve developmental milestones, strengthen abilities to adapt to adversity and build resilience and competence.19

The major predictor of SWB is by far personality, especially neuroticism (strongly correlated with negative affect and moderately inversely correlated with life satisfaction), and extraversion (moderately correlated with positive affect).20 Thus, interventions that were able to accurately “modulate” our personality would potentially increase our SWB quite permanently. However, while personality does change with time,21 interventions that modulate the personality effectively are still scarce.22

Another kind of interventions that could increase our SWB would be the improvement of our circumstances, e.g., a salary increase.23 However, circumstances cannot be always improved, and when they do, their effect on SWB fades with time. Few months after being fired or promoted, our SWB returns mostly to normal.24 We use to think that an achievement will make us very happy, or a misfortune very unhappy, yet when it happens we realize that our prediction was wrong, that our happiness has changed less than expected.25 To integrate this evidence, Headey and Wearing26 proposed, in their dynamic equilibrium model, that people have levels of SWB determined by their personality, and that changes in circumstances produce increases or decreases in SWB, but over time individuals tend to return to their baseline SWB. That said, this adaptation to circumstances is not total, e.g., richer people are still slightly happier than poorer people.27

Beyond changes in personality and circumstances, various intervention frameworks have investigated whether there are other ways to increase individual SWB. Unfortunately, these interventions have been traditionally disintegrated in different disciplines, which we may be globally divide in those within the umbrella of positive psychology,28 and those with other global aims but with interventions that still may increase SWB, such as mindfulness,29 physical activity30 or even diet31. Positive psychology is a branch of psychology that instead of being focused on treating mental disorders, it is focused on the improvement of SWB and other valued subjective experiences such as optimism, as well as on positive traits (e.g., the capacity for love) and virtues (e.g., altruism).32 For improving SWB, positive psychologists have created a series of varied interventions like conducting acts of
kindness (e.g., prosocial purchases), thinking about positive experiences, practicing gratitude, cultivating sacred moments or savoring the moment, to cite some.  

Unfortunately, while great synthesis work exists within the field of PPI, to our knowledge there are no wider, umbrella syntheses that combine both PPI and other interventions, while an intervention could increase the SWB independently of whether it belongs to the positive psychology or not. In addition, many meta-analyses did not assess the efficacy of these interventions separately for positive affect, negative affect, and life satisfaction.  

We present here an umbrella review of the systematic reviews and meta-analyses of randomized controlled trials (RCT) that have assessed the efficacy of any kind of interventions in increasing SWB, as compared to control conditions, in the general population. The greatest advantage of umbrella reviews is that they summarize and systematically assess and grade the existing evidence on a specific topic only including the highest level of evidence, namely other systematic reviews and meta-analyses. This systematic integration of evidence from multiple meta-analyses is necessary because when there are many types of intervention to choose from, a meta-analysis typically assesses only one type of intervention (e.g., positive psychology), and because different meta-analyses use different methods so that two meta-analyses on the same intervention may reach different conclusions even when published within the same year. An integrated, systematic, umbrella review is thus necessary to provide an objective picture of the wide range of interventions from different disciplines that might potentially increase SWB.  

The aim of this review was thus to synthetize the evidence of the different interventions that might improve the SWB.

**Materials and methods**

We pre-registered the study protocol with the International Prospective Register of Systematic Reviews (PROSPERO; CRD42020111681) and we completed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist (tables available upon request).

**Search strategy and eligibility criteria**

We searched *PubMed*, *Web of Science* and *Scopus* from inception to August 20, 2018 for systematic reviews and meta-analyses of RCT of any kind of interventions assessing the increase of any of the following measures of SWB:
positive affect, negative affect, and life satisfaction. The search terms were ["systematic review" or "meta-analysis"] and “subjective well-being”), and we did not restrict them to appear in a specific section (e.g., in the title). With the aim of conducting a balanced umbrella review of both PPI and other interventions of any nature, we did not use search terms that would detect only interventions of a specific discipline (e.g., “optimism” for PPI).

The inclusion criteria for the individual RCT were: a) they were published in peer-reviewed journals; b) they were conducted in non-clinical populations; c) they assessed the efficacy of interventions in increasing positive affect, negative affect or life satisfaction; and d) the effects were compared to control groups (e.g., waiting list or psychological placebo interventions). Conversely, we excluded non-systematic reviews, studies not using a control group, studies not randomizing individuals to the control and intervention arms, and studies on patients or caregivers. We did not impose language restrictions. Two investigators performed the search independently (either AAE, AS and/or JR; we distributed the work in different combinations of peers), and we resolved disagreements by consensus.

We initially intended to include pooled measures of SWB, but we later decided to exclude them because they were too heterogeneous. For example, Lyubomirsky et al\(^{38}\) created a composite of life satisfaction, happiness, pleasant affect and unpleasant affect by averaging their z-scores (after reverse-coding unpleasant affect), Page et al\(^{39}\) summed life satisfaction and positive affect and subtracted negative affect, or King et al\(^{40}\), Sheldon et al\(^{41}\), Aknin et al\(^{42}\) and Donnelly et al\(^{43}\) subtracted negative affect from positive affect (i.e., the affect balance). Similarly, we initially also intended to include measures of subjective “happiness”, but this outcome was discarded during the review process because the multiple meanings of this word might lead to confusion. In any case, we still include these studies if they reported separate scores for positive affect, negative affect, and/or life satisfaction.

**Data extraction and selection**

We used a systematic approach to extract and select the data. First, we checked the inclusion criteria for each systematic review or meta-analysis. Second, we checked the inclusion criteria for each individual study in the included systematic reviews and meta-analyses. Third, we extracted the following data (from the systematic review or meta-analysis or, when not reported, from the individual RCT): reference, type of intervention and control group, time of assessment (e.g., 1-month follow-up), specific population under investigation, number of participants in each group, age, measure of SWB (positive affect, negative affect, or life satisfaction), assessment instrument (e.g.,
Positive and Negative Affect Schedule\textsuperscript{14}, effect size (Hedges’ g) of the comparison between intervention and control groups and the corresponding 95% confidence interval or data to estimate them, and any potential study limitation (e.g., unclear blinding or substantial loss to follow-up). Fourth, we rated the quality of the systematic review or meta-analysis using the Assessment of Multiple Systematic Reviews (AMSTAR 2) tool,\textsuperscript{44} and assessed the risk of bias of the studies included using the Grading of Recommendations Assessment, Development, and Evaluation (GRADE).\textsuperscript{45}

Two investigators (either AAE, AS and/or JR; we distributed the work in different combinations of peers) conducted these steps independently, and we resolved disagreements by consensus.

If the RCT had used more than one control condition, we preferred the more emotionally neutral one and we discarded the others. For example, Froh et al\textsuperscript{46} compared a group conducting a gratitude intervention with two groups: one group instructed to recall hassles and burdens, and one group instructed to only complete the measurements; we preferred the latter and discarded the hassles condition group. If the RCT reported more than one effect size (e.g., at different follow-up times), we averaged them. For example, Froh et al\textsuperscript{46} reported two measurements of negative affect after their gratitude vs. control intervention: one at post-test and one at 3-week follow-up; we calculated the standardized mean difference between groups at each time point (post-test: 0.09, follow-up: 0.14), and then we averaged them (0.11). This averaged effect size simply represents the mean effect size, it is not the effect size of a combination of the different measurements. To understand the meaning of this averaged effect size, it is important to observe that differences in the reported effect sizes are not only due to potential differences between post-test and 3-week follow-up but also due to measurement error. In other words, even if there were no true differences between post-test and 3-week follow-up, the effect sizes would be possibly different just by chance. The average of the two effect sizes, thus, may represent the effect size at a time-point between post-test and 3-week follow-up but, also, it represents an effect size with lower measurement error. If the sample sizes used to estimate the two reported effect sizes differed slightly (e.g., due to few loss-to-follow-up before the second measurement), we selected the largest. We preferred the comparison of the changes in measures of SWB from pre-to post-intervention (or the slope in linear models), but we extracted the comparison of post-intervention measures of SWB in those studies that did not report other data. For the latter studies, we used the effect size of the comparison between post-intervention scores. If a manuscript reported more than one SWB measure (e.g., life satisfaction and positive affect), we analyzed them separately.
**Statistical analyses and grading of the evidence**

For each included systematic review (or meta-analysis) and measure, we conducted a random-effects meta-analysis to estimate the summary effect size of the comparison between groups, its 95% confidence interval, and the between-study heterogeneity $I^2$ statistic (values > 50% might indicate high heterogeneity, and values > 75% very high). The use of a random-effects model with its estimation of heterogeneity accommodates potential differences between studies, e.g., related to the use of different scales. We also conducted a random-effects meta-regression by the standard error to detect potential publication bias in small studies. We conducted the calculations with in-house umbrella-review scripts for R and the packages “meta” and “metansue”. The latter allows to include unbiasedly studies reporting “non-statistically significant results” but not reporting any statistic; note that excluding these studies would inflate the effect size, whereas including them assuming a null effect size would cause a bias towards zero. When reporting the results, we refer to Hedges’ $g<0.3$ as small, $g=0.3-0.4$ as small-to-medium, $g=0.4-0.6$ as medium, $g=0.6-0.7$ as medium-to-large, and $g>0.7$ as large. This classification is based on Cohen’s recommendations (0.2: small; 0.5: medium; 0.8: large).

We graded the evidence of the efficacy of each intervention in improving each measure according to the GRADE guidelines. Specifically, we assessed the risk of bias, inconsistency, indirectness, imprecision, and publication bias, and derived the quality of the evidence from these assessments. To assess the risk of bias we looked for limitations of the included studies such as lack of blinding, significant loss of follow-up or lack of intention-to-treat analysis. Note that we only included trials that had randomized the participants (or clusters) to different arms, thus minimizing other sources of bias such as lack of allocation concealment, for what we did not consider any study to have very serious limitations. We assessed inconsistency with the $I^2$ heterogeneity statistic complemented with a description of the proportion of studies reporting contradictory findings. For example, we only rated very serious inconsistency when heterogeneity was very high plus the findings were contradictory (i.e., at least >10% studies showing results in the opposite direction from that of most studies). When there was only one study, we could note rate inconsistency, but we conservatively scored the quality of the evidence as if there was serious inconsistency. Indirect evidence would refer to that from trials conducted in special population groups or from trials that had not measured SWB but surrogates thereof. However, we had decided not to include any special population group or any surrogate of SWB. We had only included direct measures of SWB (positive affect, negative affect, and life satisfaction) in the general population. Meta-analyses were rated as imprecise if the 95% confidence interval included
both null and large effects (or could include if unknown), or if they did not meet the optimal information size (i.e., the sample size required to detect small or medium effects with 80% statistical power). Specifically, we rated serious imprecision if the confidence interval included both null and large effects (or could include if unknown), or if the overall number of individuals in each arm was inferior to 394. This number corresponds to the sample size required to detect small effects ($g = 0.2$) with 80% statistical power according to standard formula (R function “power.t.test”). We rated very serious imprecision if the overall number of individuals in each arm was inferior to 394 and the confidence interval included both null and large effects, or if the overall number of individuals in each arm was inferior to 64. This number corresponds to the sample size required to detect medium effects ($g = 0.5$) with 80% statistical power. Finally, we rated likely potential publication bias when the corresponding test when the Egger test p-value <0.1. Again, when there were too few studies to conduct the test, we conservatively scored the quality of the evidence as if there was likely publication bias.

To summarize and grade the quality of the evidence, we first gave each intervention four pluses, and subsequently we subtracted one plus if there were serious limitations, one plus if there was serious inconsistency or this could not be assessed, two pluses if there was very serious inconsistency, one plus if there was serious imprecision, two pluses if there was very serious imprecision, and one plus if there was likely publication bias or this could not be assessed. If the final score was lower than one plus, we gave the intervention a final score of one plus.

We would have considered interventions with four pluses to have high-quality evidence (i.e., we would be very confident that the true effect is approximately the effect reported here). We considered interventions with three pluses to have moderate-quality evidence (the true effect is likely to be like the effect reported here, but there is a possibility that it is substantially different). We considered interventions with two pluses to have low-quality evidence (the effect may be substantially different from the effect reported here). Finally, we considered interventions with one plus to have very low-quality evidence (the true effect is likely to be substantially different from the effect reported here).

For the sake of completeness, we also report an analysis of all PPI combined, an analysis of all other interventions combined, and an analysis of all interventions (PPI and other).

**Results**
The initial search yielded 132 manuscripts. Of these, we discarded 51 because they were in clinical populations or their caregivers, 34 because they did not include RCT, 21 because they were not systematic reviews or meta-analyses, 15 because they did not report separate measures of SWB, and three because they used duplicated datasets (Figure 1). Some of these excluded manuscripts met more than one exclusion criterion. We finally included eight systematic reviews or meta-analyses. Their AMSTAR 2 score ranged from 6.5 to 14.5. They contained 136 RCT but we excluded 17 because they were in clinical populations or their caregivers, six because the randomization was unclear, 44 because they did not report separate measures of SWB, 12 because they had not been published in a scientific journal, and 12 because they were old studies without the required data to derive effect sizes (Figure 1 and Supplement for details). We therefore included 45 RCT, of which one had been included in two reviews.

With few exceptions, all the RCT that had assessed differences in positive or negative affect used the Positive and Negative Affect Schedule (PANAS) to ask for these affects currently, in the past day, week, 2 weeks or few weeks. Burton et al, Spence et al, Layous et al, Aknin et al, and Martela et al asked the participants the degree to which they felt each of a number of adjectives (e.g., joyful, pleased or upset), analogue to the PANAS. Finally, the assessment instruments used for measuring life satisfaction included the Satisfaction with Life Scale (SWLS), the Life Satisfaction Index A (LSIA) and the Brief Multidimensional Students' Life Satisfaction Scale (BMSLSS).

We grouped the interventions as they were presented in the included systematic reviews of meta-analyses. For this reason, we first present a meta-analysis of the overall efficacy of PPI, and subsequently we present separate meta-analyses of the efficacy of PPI consisting in conducting acts of kindness and the efficacy of PPI consisting in practicing gratitude. Similarly, we divide the interventions in the meta-analysis by Okun et al according to the groups they did: control enhancement, psychoeducational, social activity and miscellaneous.

Positive psychology
The meta-analysis by Bolier et al.\textsuperscript{33} of the overall efficacy of PPI included 39 RCT, of which 17 met our inclusion criteria.\textsuperscript{54,56,57,62,68,77-83} The interventions included were rather diverse: best possible self, positive writing, solution-focused coaching, life coaching and attainment of goals, writing about positive experiences, thinking about positive life experiences, doing acts of kindness, an optimism and gratitude exercise, practicing gratitude by counting one’s blessings, using own strengths in a new way, savoring the moment, and cultivating sacred moments. Participants in the control groups maintained their lifestyle (e.g., waiting list) or conducted psychological placebo interventions. We found moderate-quality evidence that PPI might induce small decreases of negative affect, and low-quality evidence that they might induce medium increases of positive affect (Table 1). The quality of the evidence for potential increases in life satisfaction was very low.

- Table 1 about here -

The meta-analysis by Curry et al.\textsuperscript{55} of the efficacy of PPI consisting in conducting acts of kindness included 26 RCT, of which 12 met our inclusion criteria.\textsuperscript{43,54,58-60,84-90} The interventions included prosocial purchases, social recycling, benevolence, and other acts of kindness. Participants in the control groups maintained their lifestyle or conducted psychological placebo interventions. We found low-quality evidence that acts of kindness might induce small to medium increases in life satisfaction (Table 1). The quality of the evidence for potential changes of positive or negative affect was very low.

The meta-analysis by Renshaw and Steeves\textsuperscript{75} of the efficacy of PPI consisting in practicing gratitude in youth included six RCT, of which three met our inclusion criteria.\textsuperscript{46,69,72} Participants in the control groups maintained their lifestyle or conducted psychological placebo interventions. We found low-quality evidence for a potential non-statistically significant change of life satisfaction with gratitude interventions. The quality of the evidence for potential changes of positive or negative affect was very low.

Finally, when we combined all PPI from these meta-analyses, the results were similar to those of the meta-analysis of the overall efficacy of PPI.\textsuperscript{33} We found moderate-quality evidence that PPI might induce small decreases of negative affect, and low-quality evidence that they might induce small increases of life satisfaction (Table 2). The quality of the evidence for potential increases in positive affect was very low.
Other interventions

Yoga: The systematic review by Mansfield et al\textsuperscript{91} of the efficacy of sport and dance participation among healthy young people included 11 RCT, of which one met our inclusion criteria.\textsuperscript{70} It was about the efficacy of yoga, and participants in the control group conducted usual physical education. The yoga program consisted of physical postures, breathing exercises, relaxation, and meditation for 30 minutes, two or three times a week for ten weeks.\textsuperscript{70} The quality of the evidence for potential changes of positive or negative affect with yoga was very low (Table 1).

Resilience training: The systematic review by Robertson et al\textsuperscript{92} of the efficacy of work-based resilience training included eight RCT, of which one met our inclusion criteria.\textsuperscript{71} Participants in the control group maintained their lifestyle (waiting list). The quality of the evidence for potential changes of positive or negative affect with resilience training was very low (Table 1).

Physical activity: The systematic review by Zhang and Chen\textsuperscript{93} of the efficacy of physical activity interventions included six RCT, of which two met our inclusion criteria\textsuperscript{94,95}. They were in elder individuals, and participants in the control groups did moderate or stretching and toning exercise. The quality of the evidence for potential changes in life satisfaction with physical activity was very low (Table 1).

Leisure: The meta-analysis by Kuykendall et al\textsuperscript{96} of the efficacy of leisure interventions included six RCT, of which one met our inclusion criteria.\textsuperscript{97} It was in elder individuals, and participants in the control group maintained their lifestyle. Leisure interventions consisted of a variety of activities such as discussion exercises, paper and pencil exercises, role playing, and recreation activity participation.\textsuperscript{97} The quality of the evidence for a potential increase in life satisfaction with leisure was very low (Table 1).

Control enhancement, psychoeducation, and part-time work: The meta-analysis by Okun et al\textsuperscript{76} of the efficacy of several interventions for elders included 34 RCT, grouped in control-enhancement, psychoeducational, social activity, and miscellaneous interventions. We could include three RCT about the efficacy of control enhancement interventions (e.g., offering the responsibility for taking care of bird feeders, education about stress management, nutritional...
awareness, immediate environment, self-responsibility, physical fitness and spirituality),\textsuperscript{98-100} four about the efficacy of psychoeducational interventions (e.g., increasing the knowledge and skills),\textsuperscript{101-104} and one in the miscellaneous category (part-time work).	extsuperscript{105} Participants in the control groups maintained their lifestyle or conducted psychological placebo interventions. The quality of the evidence for a potential increase in life satisfaction with control enhancement interventions, or potential changes of life satisfaction with psychoeducational or part-time work interventions, was very low (Table 1).

When we combined all interventions from these meta-analyses, the quality of the evidence for potential changes in positive affect, negative affect or life satisfaction was very low (Table 2).

\textit{All interventions combined}

When we combined all PPI and other interventions, we found moderate-quality evidence that they might induce small decreases of negative affect, and low-quality evidence that they might induce small-to-medium increases of positive affect (Table 2). The quality of the evidence for potential increases in life satisfaction was very low.

\textbf{Discussion}

This umbrella review systematically assessed the evidence of the efficacy of any kind of interventions in increasing individual SWB, including both PPI and other interventions. We aimed to provide the picture of the wide range of interventions from different disciplines that might potentially increase SWB, and thus we included any intervention as far as the inclusion criteria were met.

The main finding was that there is low- to moderate-quality evidence that PPI might increase positive affect and decrease negative affect. The larger effect size was the increase in positive affect (Hedges’ $g \sim 0.4$), but the quality of its evidence was low due to serious inconsistency (i.e., heterogeneity between studies). Indeed, effect size ranged from very large increases to even a (small) decrease depending on the RCT. The decrease in negative affect associated to PPI was more consistent but also smaller (Hedges’ $g \sim -0.2$).
A potential source of the heterogeneity in the increase in positive affect may be related to the different types of PPI, although this was not apparent when looking at the specific PPI associated to different effect sizes. Specifically, PPI associated to very large increases of positive affect (Hedges’ g 0.9 to 1.3) consisted of conducting acts of kindness, of writing about an intensely positive experience, of solution-focused life coaching group program, and of writing about oneself in the future imagining that everything has gone as well as it possibly could.\textsuperscript{56,59,63,68} PPI associated to moderate increases of positive affect (Hedges’ g 0.3 to 0.7) consisted again of conducting acts of kindness, of other solution-focused life coaching programs, of similar writing about oneself in the future imaging that everything has gone as well as it possibly could, of education about the appraisal of benefit exchanges, and of writing things for which one could feel grateful.\textsuperscript{57,60,64,66,72,77,83,84} Finally, PPI associated to week or null increases of positive affect (Hedges’ g 0.2 or lower) consisted once again of conducting acts of kindness, once again of writing about oneself in the future imaging that everything has gone as well as it possibly could, again of writing things for which one could feel grateful, of savoring the moment (e.g., smiling while enjoying time well spent with friends), of writing about happy experiences, and of practicing one’s character strengths (e.g., creativity, social intelligence or humor) \textsuperscript{46,58,62,65,67,69,80}. That said, two additional meta-analyses showed different effect sizes on subjective wellbeing for two of these types of PPI. We found low-quality evidence that PPI consisting in conducting acts of kindness might increase life satisfaction, whereas we did not find statistically significant effects for PPI consisting in practicing gratitude. A possible beneficial effect of conducting acts kindness on SWB might be associated with the increased social support as opposed to loneliness.\textsuperscript{106} Other factors such as the duration of the PPI might play a role in the heterogeneity. In their meta-analysis of PPI, Bolier et al\textsuperscript{33} did not find statistically significant moderator effects for this, but those analyses may have been under-powered. Future evidence syntheses may provide larger evidence databases that would allow a better-powered exploration of the sources of heterogeneity.

Besides the inconsistency, there are some other reasons why the quality of the evidence of the PPI included in this umbrella review is relatively low. First, most studies had unclear blinding, which may be difficult to implement in these settings. Second, most studies used a per-protocol analysis instead of conducting intention-to-treat analyses, a study limitation that may exaggerate the effects, may introduce bias, and it may even violate the principle of randomization.\textsuperscript{107}

The quality of the evidence of the studies assessing other interventions (e.g., leisure, physical activity, or yoga) was in general very low, especially due to a lack of large overall sample sizes. However, great caution is
warranted until the overall sample sizes are larger and the quality of the evidence higher. Indeed, some interventions only included one study. Future studies might provide evidence that these interventions do increase the SWB. At this regard, we want to highlight the important differences between quality of the evidence and effect size. The quality of the evidence helps us know how much we can trust in the overall results, independently of whether the results are that the interventions are efficacious or not. Therefore, we may have high-quality evidence that an intervention has no effect, or we may have low-quality evidence that an intervention has huge beneficial effects. In addition, here we only studied the effects on SWB. It is well-known, for example, that physical activity has other important positive outcomes.108

Most individuals from highly developed countries report that their SWB is in the upper range, e.g., that they already feel very satisfied with life.109 This may result in a ceiling effect because an intervention will hardly increase the life satisfaction of an individual who already feels very satisfied with life. Therefore, it is possible that the effect size of the interventions would be larger in individuals who reported that their SWB is in the middle or lower ranges. One observation supporting this possibility is the finding that the efficacy of antidepressants is higher in patients with more severe depression,110 although this observation has been questioned.111 A related issue is that one could even question whether it is sensible to try to increase SWB in individuals with already high SWB. When comparing individuals with high and very high SWB, the latter are more successful in close relationships and volunteer work, but less successful in income, education, and political participation.112

We would like to highlight some limitations of our review. First, umbrella reviews do not include those studies that have not been included in a systematic review or meta-analysis. The possibility that there are no systematic reviews or meta-analyses on the effects of a given intervention seems unlikely, because meta-analyses are nowadays performed massively.113 However, published systematic reviews and meta-analyses might not include recent trials. Second, our search strategy was designed to find interventions from any discipline, but it was probably sub-optimal for finding interventions of a specific discipline. For example, we probably failed to include some meta-analyses of PPI that we could possibly have identified using positive psychology terms such as “optimism” or “gratitude”. However, we considered that including additional positive psychology terms would bias the inclusion of studies towards PPI to the detriment of other interventions. We refer the reader to a recently published synthesis in the field of positive psychology28 that may have included some works that we failed to include. We would also want to note that for any kind of intervention for which we were able to find a meta-analysis, we should have theoretically been
able to include any RCT published until the date of the systematic search of the meta-analysis. Third, for studies that did not report pre-intervention scores or pre-post statistics, we used the effect size of the comparison between post-intervention scores. While suboptimal, this is equivalent to the effect size of the comparison between pre-post score differences under the general assumptions that pre-intervention mean scores are similar between the two groups, variances are similar, and pre-post correlation is about 0.5. Fourth, when the study reported several post-intervention and follow-up effect sizes, we averaged them. We considered that follow-up information is very relevant, given that we understand that the interest of these interventions is that the increase in SWB lasts a time, and thus we even considered using only the last effect size. However, effect size usually decreases with follow-up, and it would have been unfair to include the last, usually smaller effect size from studies conducting a long follow-up, while the initial, usually larger effect size from studies not conducting any follow-up after the intervention. To balance the situation, we preferred a medium consisting on averaging the effect sizes of the different follow-up points. In addition, as we noted earlier, such average has lower measurement error. Fifth, we focused on the common assessments of SWB, but other assessments are possible and may have indeed some advantages. For example, experience-sampling methods or ecological momentary assessments ask the participants their SWB at random moments of their everyday lives, potentially circumventing memory biases. Sixth, even when measuring the same components of SWB (positive affect, negative affect, and life satisfaction), the studies used different scales, which may have introduced heterogeneity in the analysis. Seventh, PPI were rather varied, a factor that may have increased heterogeneity. Finally, as we noted earlier, some other interventions included only one RCT and should be thus taken with more caution.

Future umbrella syntheses with more studies may be able to better model the heterogeneity between interventions. This may be achieved by stratifying the interventions, or more sophisticatedly, by conducting meta-regressions that model the relevant characteristics of each intervention. Additionally, there have been few studies in population subgroups scoring low in specific SWB components but not in others (which should benefit more from these interventions)\textsuperscript{114}, and little research on potential moderators.\textsuperscript{115,116} The study of subgroups and moderators when there are enough data may be of great interest, as it is entirely plausible that an intervention works for an individual but not for another. Finally, given the importance of personality for SWB, the creation of interventions that modulate the personality (e.g., decreasing neuroticism and increasing extraversion) could be also promising.
In conclusion, despite its limitations, this umbrella review shows that there is moderate-quality evidence that PPI increase SWB. Conversely, the evidence for other interventions (e.g., yoga, physical activity, or leisure) is still very low.
Funding details

This work was supported by Miguel Servet Research Contract (CP14/00041 and CPII19/00009) to J.R., PFIS Predoctoral Contract FI16/00311 to A.A.E. and Research Projects PI14/00292 and PI19/00394 from the Plan Nacional de I+D+i 2013–2016, the Instituto de Salud Carlos III-Subdirección General de Evaluación y Fomento de la Investigación and the European Regional Development Fund (FEDER, ‘Investing in your future’). The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Disclosure statement

The authors have no interests to declare in relation to this manuscript.
References


Figure captions

Figure 1. Flow diagram of the inclusion of systematic reviews and meta-analyses and the inclusion of randomized controlled trials (RCT).
Can we increase the subjective well-being of the general population?

An umbrella review of the evidence

Abstract

Introduction: Subjective well-being (SWB) refers to being satisfied with one’s life, having positive affect and having little negative affect. We may understand it as a subjective definition of good life, or in colloquial terms “happiness”, and it has been associated with several important benefits such as lower mortality. In the last decades, several randomized controlled trials (RCT) have investigated the efficacy of several interventions in increasing SWB in the general population but results from different disciplines have not been integrated.

Methods: We conducted an umbrella review of systematic reviews and meta-analyses of RCT that assess the efficacy of any kind of interventions in increasing SWB in the general population, including both positive psychology interventions (PPI) and other interventions. We (re)calculated the meta-analytic statistics needed to objectively assess the quality of the evidence of the efficacy of each type of intervention in improving each component of SWB according to the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach.

Results: There was moderate-quality evidence that PPI might induce small decreases of negative affect, and low-quality evidence that they might induce moderate increases of positive affect. We found similar results for those PPI specifically consisting in conducting acts of kindness (especially spending money on or giving items to others), for which there was low-quality evidence that they might induces small increases of life satisfaction, but not for PPI specifically consisting in practicing gratitude. Quality of the evidence of the efficacy for the other interventions included in the umbrella review (yoga, resilience training, physical activity, leisure, control enhancement, psychoeducation, and miscellaneous) was very low.

Conclusion: There is some evidence that PPI, and specially conducting acts of kindness such as spending money on others, may increase the SWB of the general population. The quality of the evidence of the efficacy for other interventions (e.g., yoga, physical activity, or leisure) is still very low.

Registration number: PROSPERO CRD42020111681.

Keywords: acts of kindness, meta-analysis, positive psychology, randomized controlled trial, subjective well-being, umbrella review
¿Podemos incrementar el bienestar subjetivo de la población en general?

Una revisión paraguas de la evidencia

Resumen

Introducción: El bienestar subjetivo (SWB) se refiere a estar satisfecho con la vida, tener afecto positivo y tener poco afecto negativo. Podemos entenderlo como una definición subjetiva de la buena vida, o en términos coloquiales, como “felicidad”, y se ha asociado con varios beneficios importantes, como una menor mortalidad. En las ultimas décadas, varios ensayos controlados aleatorizados (RCT) han investigado la eficacia de varias intervenciones para aumentar el SWB en la población general, pero los resultados de las diferentes disciplinas no se han integrado.

Métodos: Realizamos una revisión paraguas de revisiones sistemáticas y metaanálisis de RCT que evaluasen la eficacia de cualquier tipo de intervención para aumentar el SWB en la población general, incluidas tanto las intervenciones de psicología positiva (PPI) como otras intervenciones. (Re)calculamos los estadísticos metaanalíticos necesarios para evaluar objetivamente la calidad de la evidencia de la eficacia cada tipo de intervención para mejorar cada componente del SWB de acuerdo con el Grading of Recommendations Assessment, Development and Evaluation (GRADE).

Resultados: Hubo evidencia de calidad moderada de que las PPI podrían inducir pequeñas disminuciones de afecto negativo y evidencia de baja calidad de que podrían inducir aumentos moderados de afecto positivo. Encontramos resultados similares para aquellas PPI que consistían específicamente en realizar actos de bondad (especialmente gastar dinero en o dar artículos a otros), para las cuales había evidencia de baja calidad de que podrían inducir pequeños aumentos de satisfacción con la vida, pero no para las PPI que consistían específicamente en practicar la gratitud. La calidad de la evidencia de la eficacia para las otras intervenciones incluidas en la revisión paraguas (yoga, entrenamiento de resiliencia, actividad física, ocio, mejora del control, psicooeducación y miscelánea) fue muy baja.

Conclusión: Existe alguna evidencia de que las PPI, y especialmente la realización de actos de bondad como gastar dinero en otros, pueden aumentar el SWB de la población general. La calidad de la evidencia de la eficacia para otras intervenciones (p. ej., yoga, actividad física u ocio) sigue siendo muy baja.

Número de registro: PROSPERO CRD42020111681.

Palabras clave: actos de bondad, metaanálisis, psicología positiva, ensayo controlado aleatorizado, bienestar subjetivo, revisión paraguas
Introduction

Ed Diener defined subjective well-being (SWB) as an overall satisfaction with one’s life, having many pleasant emotions and moods (positive affect), and having few unpleasant ones (negative affect).\(^1\) It may be understood as a subjective definition of good life, or colloquially as “happiness”, but we will avoid these terms because they may have other meanings that are out of the scope of the present work. In the following we will refer to the definition of SWB presented above, known as the Diener tripartite model.

A high SWB has been associated with several important positive outcomes, such as substantially lower risk of exercise stress-induced myocardial ischemia (logistic regression-derived odds ratio = 0.55)\(^2\) or lower mortality (hazard ratio = 0.92 when comparing the group with highest SWB with the group with lowest SWB).\(^3\) SWB has also been linked to higher self-esteem and self-efficacy, and it is a relevant component of mental health.\(^4\)–\(^7\) Some abnormalities of SWB (e.g., depressed mood and anxiety as extreme forms of few pleasant emotions and many unpleasant emotions) are core symptoms of mood/anxiety disorders (the most common mental/psychiatric disorders)\(^8\)–\(^10\), and a component of many other mental/psychiatric disorders (e.g., schizophrenia is characterised by decreased ability to experience pleasure\(^11\)). Lower SWB has indeed been associated to higher suicidal ideation and behavior.\(^12\),\(^13\)

SWB consists of three main components: positive affect, negative affect and life satisfaction\(^1\) Positive affect refers to pleasant emotions and moods, such as interest, enthusiasm or pride, and negative affect to unpleasant ones, such as guilt, irritability or fear.\(^14\) Life satisfaction refers to thinking that one’s life is close to one’s ideal, that the important things one wants in life have been already achieved.\(^15\) Positive and negative affect are sometimes combined in an “affective” component to distinguish them from the “cognitive” component (life satisfaction). All components show a weak to moderate correlation,\(^16\),\(^17\) for what the improvement of one component (e.g., having more positive emotions) may be accompanied by improvements of other components (e.g., feeling more satisfied). However, this should not be taken for granted. There are common exceptions where individuals with “many” pleasant emotions also have "many" unpleasant emotions or feel unsatisfied. Or similarly, individuals with a depressive episode (associated with high negative affect and low positive affect) may also feel “high” (which should be considered as positive affect), known as mixed features.\(^11\) It is thus more desirable to study each component separately.\(^1\)

Individuals with mental disorders could receive interventions that increase SWB\(^18\) but the promotion of mental health should also include interventions that increase SWB in healthy individuals. Note that the aims of mental
health prevention interventions include targeting risk factors and strengthen abilities to prevent the development of one or more conditions, while mental health promotion interventions aim to promote psychological wellbeing, increase the ability to achieve developmental milestones, strengthen abilities to adapt to adversity and build resilience and competence.  

The major predictor of SWB is by far personality, specially neuroticism (strongly correlated with negative affect and moderately inversely correlated with life satisfaction), and extraversion (moderately correlated with positive affect). Thus, interventions that were able to accurately “modulate” our personality would potentially increase our SWB quite permanently. However, while personality does change with time, interventions that modulate the personality effectively are still scarce.  

Another kind of interventions that could increase our SWB would be the improvement of our circumstances, e.g., a salary increase. However, circumstances cannot be always improved, and when they do, their effect on SWB fades with time. Few months after being fired or promoted, our SWB returns mostly to normal. We use to think that an achievement will make us very happy, or a misfortune very unhappy, yet when it happens we realize that our prediction was wrong, that our happiness has changed less than expected. To integrate this evidence, Headey and Wearing proposed, in their dynamic equilibrium model, that people have levels of SWB determined by their personality, and that changes in circumstances produce increases or decreases in SWB, but over time individuals tend to return to their baseline SWB. That said, this adaptation to circumstances is not total, e.g., richer people are still slightly happier than poorer people.  

Beyond changes in personality and circumstances, various intervention frameworks have investigated whether there are other ways to increase individual SWB. Unfortunately, these interventions have been traditionally disintegrated in different disciplines, which we may be globally divide in those within the umbrella of positive psychology, and those with other global aims but with interventions that still may increase SWB, such as mindfulness, physical activity or even diet. Positive psychology is a branch of psychology that instead of being focused on treating mental disorders, it is focused on the improvement of SWB and other valued subjective experiences such as optimism, as well as on positive traits (e.g., the capacity for love) and virtues (e.g., altruism). For improving SWB, positive psychologists have created a series of varied interventions like conducting acts of
kindness (e.g., prosocial purchases), thinking about positive experiences, practicing gratitude, cultivating sacred moments or savoring the moment, to cite some.33

Unfortunately, while great synthesis work exists within the field of PPI, to our knowledge there are no wider, umbrella syntheses that combine both PPI and other interventions, while an intervention could increase the SWB independently of whether it belongs to the positive psychology or not. In addition, many meta-analyses did not assess the efficacy of these interventions separately for positive affect, negative affect, and life satisfaction.33

We present here an umbrella review of the systematic reviews and meta-analyses of randomized controlled trials (RCT) that have assessed the efficacy of any kind of interventions in increasing SWB, as compared to control conditions, in the general population. The greatest advantage of umbrella reviews is that they summarize and systematically assess and grade the existing evidence on a specific topic only including the highest level of evidence, namely other systematic reviews and meta-analyses.34 This systematic integration of evidence from multiple meta-analyses is necessary35,36 because when there are many types of intervention to choose from, a meta-analysis typically assesses only one type of intervention (e.g., positive psychology), and because different meta-analyses use different methods so that two meta-analyses on the same intervention may reach different conclusions even when published within the same year. An integrated, systematic, umbrella review is thus necessary to provide an objective picture of the wide range of interventions from different disciplines that might potentially increase SWB.

The aim of this review was thus to synthetize the evidence of the different interventions that might improve the SWB.

Materials and methods

We pre-registered the study protocol with the International Prospective Register of Systematic Reviews (PROSPERO; CRD42020111681) and we completed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist17 (tables available upon request).

Search strategy and eligibility criteria

We searched PubMed, Web of Science and Scopus from inception to August 20, 2018 for systematic reviews and meta-analyses of RCT of any kind of interventions assessing the increase of any of the following measures of SWB:
positive affect, negative affect, and life satisfaction. The search terms were ["systematic review" or "meta-analysis"] and "subjective well-being"), and we did not restrict them to appear in a specific section (e.g., in the title). With the aim of conducting a balanced umbrella review of both PPI and other interventions of any nature, we did not use search terms that would detect only interventions of a specific discipline (e.g., “optimism” for PPI).

The inclusion criteria for the individual RCT were: a) they were published in peer-reviewed journals; b) they were conducted in non-clinical populations; c) they assessed the efficacy of interventions in increasing positive affect, negative affect or life satisfaction; and d) the effects were compared to control groups (e.g., waiting list or psychological placebo interventions). Conversely, we excluded non-systematic reviews, studies not using a control group, studies not randomizing individuals to the control and intervention arms, and studies on patients or caregivers. We did not impose language restrictions. Two investigators performed the search independently (either AAE, AS and/or JR; we distributed the work in different combinations of peers), and we resolved disagreements by consensus.

We initially intended to include pooled measures of SWB, but we later decided to exclude them because they were too heterogeneous. For example, Lyubomirsky et al\textsuperscript{38} created a composite of life satisfaction, happiness, pleasant affect and unpleasant affect by averaging their z-scores (after reverse-coding unpleasant affect), Page et al\textsuperscript{39} summed life satisfaction and positive affect and subtracted negative affect, or King et al\textsuperscript{40}, Sheldon et al\textsuperscript{41}, Aknin et al\textsuperscript{42} and Donnelly et al\textsuperscript{43} subtracted negative affect from positive affect (i.e., the affect balance). Similarly, we initially also intended to include measures of subjective “happiness”, but this outcome was discarded during the review process because the multiple meanings of this word might lead to confusion. In any case, we still include these studies if they reported separate scores for positive affect, negative affect, and/or life satisfaction.

**Data extraction and selection**

We used a systematic approach to extract and select the data. First, we checked the inclusion criteria for each systematic review or meta-analysis. Second, we checked the inclusion criteria for each individual study in the included systematic reviews and meta-analyses. Third, we extracted the following data (from the systematic review or meta-analysis or, when not reported, from the individual RCT): reference, type of intervention and control group, time of assessment (e.g., 1-month follow-up), specific population under investigation, number of participants in each group, age, measure of SWB (positive affect, negative affect, or life satisfaction), assessment instrument (e.g.,
Positive and Negative Affect Schedule\textsuperscript{14}, effect size (Hedges’ \( g \)) of the comparison between intervention and control groups and the corresponding 95% confidence interval or data to estimate them, and any potential study limitation (e.g., unclear blinding or substantial loss to follow-up). Fourth, we rated the quality of the systematic review or meta-analysis using the Assessment of Multiple Systematic Reviews (AMSTAR 2) tool,\textsuperscript{44} and assessed the risk of bias of the studies included using the Grading of Recommendations Assessment, Development, and Evaluation (GRADE).\textsuperscript{45}

Two investigators (either AAE, AS and/or JR; we distributed the work in different combinations of peers) conducted these steps independently, and we resolved disagreements by consensus.

If the RCT had used more than one control condition, we preferred the more emotionally neutral one and we discarded the others. For example, Froh et al\textsuperscript{46} compared a group conducting a gratitude intervention with two groups: one group instructed to recall hassles and burdens, and one group instructed to only complete the measurements; we preferred the latter and discarded the hassles condition group. If the RCT reported more than one effect size (e.g., at different follow-up times), we averaged them. For example, Froh et al\textsuperscript{46} reported two measurements of negative affect after their gratitude vs. control intervention: one at post-test and one at 3-week follow-up; we calculated the standardized mean difference between groups at each time point (post-test: 0.09, follow-up: 0.14), and then we averaged them (0.11). This averaged effect size simply represents the mean effect size, it is not the effect size of a combination of the different measurements. To understand the meaning of this averaged effect size, it is important to observe that differences in the reported effect sizes are not only due to potential differences between post-test and 3-week follow-up but also due to measurement error. In other words, even if there were no true differences between post-test and 3-week follow-up, the effect sizes would be possibly different just by chance. The average of the two effect sizes, thus, may represent the effect size at a time-point between post-test and 3-week follow-up but, also, it represents an effect size with lower measurement error. If the sample sizes used to estimate the two reported effect sizes differed slightly (e.g., due to few loss-to-follow-up before the second measurement), we selected the largest. We preferred the comparison of the changes in measures of SWB from pre-to post-intervention (or the slope in linear models), but we extracted the comparison of post-intervention measures of SWB in those studies that did not report other data. For the latter studies, we used the effect size of the comparison between post-intervention scores. If a manuscript reported more than one SWB measure (e.g., life satisfaction and positive affect), we analyzed them separately.
Statistical analyses and grading of the evidence

For each included systematic review (or meta-analysis) and measure, we conducted a random-effects meta-analysis to estimate the summary effect size of the comparison between groups, its 95% confidence interval, and the between-study heterogeneity $I^2$ statistic (values > 50% might indicate high heterogeneity, and values > 75% very high). The use of a random-effects model with its estimation of heterogeneity accommodates potential differences between studies, e.g., related to the use of different scales. We also conducted a random-effects meta-regression by the standard error to detect potential publication bias in small studies. We conducted the calculations with in-house umbrella-review scripts for R and the packages “meta” and “metansue”. The latter allows to include unbiasedly studies reporting “non-statistically significant results” but not reporting any statistic; note that excluding these studies would inflate the effect size, whereas including them assuming a null effect size would cause a bias towards zero. When reporting the results, we refer to Hedges’ $g$ as small, $g=0.3$–0.4 as small–medium, $g=0.4$–0.6 as medium, $g=0.6$–0.7 as medium-to-large, and $g>0.7$ as large. This classification is based on Cohen’s recommendations (0.2: small; 0.5: medium; 0.8: large).

We graded the evidence of the efficacy of each intervention in improving each measure according to the GRADE guidelines. Specifically, we assessed the risk of bias, inconsistency, indirectness, imprecision, and publication bias, and derived the quality of the evidence from these assessments. To assess the risk of bias we looked for limitations of the included studies such as lack of blinding, significant loss of follow-up or lack of intention-to-treat analysis. Note that we only included trials that had randomized the participants (or clusters) to different arms, thus minimizing other sources of bias such as lack of allocation concealment, for what we did not consider any study to have very serious limitations. We assessed inconsistency with the $I^2$ heterogeneity statistic complemented with a description of the proportion of studies reporting contradictory findings. For example, we only rated very serious inconsistency when heterogeneity was very high plus the findings were contradictory (i.e., at least >10% studies showing results in the opposite direction from that of most studies). When there was only one study, we could note rate inconsistency, but we conservatively scored the quality of the evidence as if there was serious inconsistency. Indirect evidence would refer to that from trials conducted in special population groups or from trials that had not measured SWB but surrogates thereof. However, we had decided not to include any special population group or any surrogate of SWB. We had only included direct measures of SWB (positive affect, negative affect, and life satisfaction) in the general population. Meta-analyses were rated as imprecise if the 95% confidence interval included
both null and large effects (or could include if unknown), or if they did not meet the optimal information size (i.e., the sample size required to detect small or medium effects with 80% statistical power). Specifically, we rated serious imprecision if the confidence interval included both null and large effects (or could include if unknown), or if the overall number of individuals in each arm was inferior to 394. This number corresponds to the sample size required to detect small effects ($g = 0.2$) with 80% statistical power according to standard formula (R function “power.t.test”). We rated very serious imprecision if the overall number of individuals in each arm was inferior to 394 and the confidence interval included both null and large effects, or if the overall number of individuals in each arm was inferior to 64. This number corresponds to the sample size required to detect medium effects ($g = 0.5$) with 80% statistical power. Finally, we rated likely potential publication bias when the corresponding test when the Egger test p-value <0.1. Again, when there were too few studies to conduct the test, we conservatively scored the quality of the evidence as if there was likely publication bias.

To summarize and grade the quality of the evidence, we first gave each intervention four pluses, and subsequently we subtracted one plus if there were serious limitations, one plus if there was serious inconsistency or this could not be assessed, two pluses if there was very serious inconsistency, one plus if there was serious imprecision, two pluses if there was very serious imprecision, and one plus if there was likely publication bias or this could not be assessed. If the final score was lower than one plus, we gave the intervention a final score of one plus.

We would have considered interventions with four pluses to have high-quality evidence (i.e., we would be very confident that the true effect is approximately the effect reported here). We considered interventions with three pluses to have moderate-quality evidence (the true effect is likely to be like the effect reported here, but there is a possibility that it is substantially different). We considered interventions with two pluses to have low-quality evidence (the effect may be substantially different from the effect reported here). Finally, we considered interventions with one plus to have very low-quality evidence (the true effect is likely to be substantially different from the effect reported here).

For the sake of completeness, we also report an analysis of all PPI combined, an analysis of all other interventions combined, and an analysis of all interventions (PPI and other).

**Results**
The initial search yielded 132 manuscripts. Of these, we discarded 51 because they were in clinical populations or their caregivers, 34 because they did not include RCT, 21 because they were not systematic reviews or meta-analyses, 15 because they did not report separate measures of SWB, and three because they used duplicated datasets (Figure 1). Some of these excluded manuscripts met more than one exclusion criterion. We finally included eight systematic reviews or meta-analyses. Their AMSTAR 2 score ranged from 6.5 to 14.5. They contained 136 RCT but we excluded 17 because they were in clinical populations or their caregivers, six because the randomization was unclear, 44 because they did not report separate measures of SWB, 12 because they had not been published in a scientific journal, and 12 because they were old studies without the required data to derive effect sizes (Figure 1 and Supplement for details). We therefore included 45 RCT, of which one had been included in two reviews.33,55

With few exceptions,66-60 all the RCT that had assessed differences in positive or negative affect used the Positive and Negative Affect Schedule (PANAS)14,61 to ask for these affects currently,62-64 in the past day,46 week,65,66 2 weeks67 or few weeks,68,72 Burton et al56, Spence et al57, Layous et al58, Aknin et al59, and Martela et al60 asked the participants the degree to which they felt each of a number of adjectives (e.g., joyful, pleased or upset), analogue to the PANAS. Finally, the assessment instruments used for measuring life satisfaction included the Satisfaction with Life Scale (SWLS)15, the Life Satisfaction Index A (LSIA)73 and the Brief Multidimensional Students' Life Satisfaction Scale (BMSLSS).74

We grouped the interventions as they were presented in the included systematic reviews of meta-analyses. For this reason, we first present a meta-analysis of the overall efficacy of PPI 33, and subsequently we present separate meta-analyses of the efficacy of PPI consisting in conducting acts of kindness and the efficacy of PPI consisting in practicing gratitude.75 Similarly, we divide the interventions in the meta-analysis by Okun et al76 according to the groups they did: control enhancement, psychoeducational, social activity and miscellaneous.

Positive psychology
The meta-analysis by Bolier et al\textsuperscript{33} of the overall efficacy of PPI included 39 RCT, of which 17 met our inclusion criteria\textsuperscript{54,56,57,62,68,77-83}. The interventions included were rather diverse: \textbf{best possible self}, positive writing, solution-focused coaching, life coaching and attainment of goals, writing about positive experiences, thinking about positive life experiences, doing acts of kindness, an optimism and gratitude exercise, practicing gratitude by counting one’s blessings, using own strengths in a new way, savoring the moment, and cultivating sacred moments. Participants in the control groups maintained their lifestyle (e.g., waiting list) or conducted psychological placebo interventions. We found moderate-quality evidence that PPI might induce small decreases of negative affect, and low-quality evidence that they might induce medium increases of positive affect (Table 1). The quality of the evidence for potential increases in life satisfaction was very low.

The meta-analysis by Curry et al\textsuperscript{55} of the efficacy of PPI consisting in conducting acts of kindness included 26 RCT, of which 12 met our inclusion criteria\textsuperscript{43,54,58-60,84-90}. The interventions included prosocial purchases, social recycling, benevolence, and other acts of kindness. Participants in the control groups maintained their lifestyle or conducted psychological placebo interventions. We found low-quality evidence that acts of kindness might induce small to medium increases in life satisfaction (Table 1). The quality of the evidence for potential changes of positive or negative affect was very low.

The meta-analysis by Renshaw and Steeves\textsuperscript{75} of the efficacy of PPI consisting in practicing gratitude in youth included six RCT, of which three met our inclusion criteria\textsuperscript{46,69,72}. Participants in the control groups maintained their lifestyle or conducted psychological placebo interventions. We found low-quality evidence for a potential non-statistically significant change of life satisfaction with gratitude interventions. The quality of the evidence for potential changes of positive or negative affect was very low.

Finally, when we combined all PPI from these meta-analyses, the results were similar to those of the meta-analysis of the overall efficacy of PPI\textsuperscript{33}. We found moderate-quality evidence that PPI might induce small decreases of negative affect, and low-quality evidence that they might induce small increases of life satisfaction (Table 2). The quality of the evidence for potential increases in positive affect was very low.
Other interventions

Yoga: The systematic review by Mansfield et al91 of the efficacy of sport and dance participation among healthy young people included 11 RCT, of which one met our inclusion criteria.70 It was about the efficacy of yoga, and participants in the control group conducted usual physical education. The yoga program consisted of physical postures, breathing exercises, relaxation, and meditation for 30 minutes, two or three times a week for ten weeks.70 The quality of the evidence for potential changes of positive or negative affect with yoga was very low (Table 1).

Resilience training: The systematic review by Robertson et al92 of the efficacy of work-based resilience training included eight RCT, of which one met our inclusion criteria.71 Participants in the control group maintained their lifestyle (waiting list). The quality of the evidence for potential changes of positive or negative affect with resilience training was very low (Table 1).

Physical activity: The systematic review by Zhang and Chen93 of the efficacy of physical activity interventions included six RCT, of which two met our inclusion criteria94,95. They were in elder individuals, and participants in the control groups did moderate or stretching and toning exercise. The quality of the evidence for potential changes in life satisfaction with physical activity was very low (Table 1).

Leisure: The meta-analysis by Kuykendall et al96 of the efficacy of leisure interventions included six RCT, of which one met our inclusion criteria.97 It was in elder individuals, and participants in the control group maintained their lifestyle. Leisure interventions consisted of a variety of activities such as discussion exercises, paper and pencil exercises, role playing, and recreation activity participation.97 The quality of the evidence for a potential increase in life satisfaction with leisure was very low (Table 1).

Control enhancement, psychoeducation, and part-time work: The meta-analysis by Okun et al76 of the efficacy of several interventions for elders included 34 RCT, grouped in control-enhancement, psychoeducational, social activity, and miscellaneous interventions. We could include three RCT about the efficacy of control enhancement interventions (e.g., offering the responsibility for taking care of bird feeders, education about stress management, nutritional
awareness, immediate environment, self-responsibility, physical fitness and spirituality), four about the efficacy of psychoeducational interventions (e.g., increasing the knowledge and skills) and one in the miscellaneous category (part-time work). Participants in the control groups maintained their lifestyle or conducted psychological placebo interventions. The quality of the evidence for a potential increase in life satisfaction with control enhancement interventions, or potential changes of life satisfaction with psychoeducational or part-time work interventions, was very low (Table 1).

When we combined all interventions from these meta-analyses, the quality of the evidence for potential changes in positive affect, negative affect or life satisfaction was very low (Table 2).

**All interventions combined**

When we combined all PPI and other interventions, we found moderate-quality evidence that they might induce small decreases of negative affect, and low-quality evidence that they might induce small-to-medium increases of positive affect (Table 2). The quality of the evidence for potential increases in life satisfaction was very low.

**Discussion**

This umbrella review systematically assessed the evidence of the efficacy of any kind of interventions in increasing individual SWB, including both PPI and other interventions. We aimed to provide the picture of the wide range of interventions from different disciplines that might potentially increase SWB, and thus we included any intervention as far as the inclusion criteria were met.

The main finding was that there is low- to moderate-quality evidence that PPI might increase positive affect and decrease negative affect. The larger effect size was the increase in positive affect (Hedges’ $g \sim 0.4$), but the quality of its evidence was low due to serious inconsistency (i.e., heterogeneity between studies). Indeed, effect size ranged from very large increases to even a (small) decrease depending on the RCT. The decrease in negative affect associated to PPI was more consistent but also smaller (Hedges’ $g \sim -0.2$).
A potential source of the heterogeneity in the increase in positive affect may be related to the different types of PPI, although this was not apparent when looking at the specific PPI associated to different effect sizes. Specifically, PPI associated to very large increases of positive affect (Hedges’ $g$ 0.9 to 1.3) consisted of conducting acts of kindness, of writing about an intensely positive experience, of solution-focused life coaching group program, and of writing about oneself in the future imagining that everything has gone as well as it possibly could.\textsuperscript{56,59,63,68} PPI associated to moderate increases of positive affect (Hedges’ $g$ 0.3 to 0.7) consisted again of conducting acts of kindness, of other solution-focused life coaching programs, of similar writing about oneself in the future imaging that everything has gone as well as it possibly could, of education about the appraisal of benefit exchanges, and of writing things for which one could feel grateful.\textsuperscript{57,60,64,66,72,77,83,84} Finally, PPI associated to week or null increases of positive affect (Hedges’ $g$ 0.2 or lower) consisted once again of conducting acts of kindness, once again of writing about oneself in the future imaging that everything has gone as well as it possibly could, again of writing things for which one could feel grateful, of savoring the moment (e.g., smiling while enjoying time well spent with friends), of writing about happy experiences, and of practicing one’s character strengths (e.g., creativity, social intelligence or humor)\textsuperscript{46,58,62,65,67,69,80}. That said, two additional meta-analyses showed different effect sizes on subjective wellbeing for two of these types of PPI. We found low-quality evidence that PPI consisting in conducting acts of kindness might increase life satisfaction, whereas we did not find statistically significant effects for PPI consisting in practicing gratitude. A possible beneficial effect of conducting acts kindness on SWB might be associated with the increased social support as opposed to loneliness.\textsuperscript{106} Other factors such as the duration of the PPI might play a role in the heterogeneity. In their meta-analysis of PPI, Bolier et al\textsuperscript{33} did not find statistically significant moderator effects for this, but those analyses may have been under-powered. Future evidence syntheses may provide larger evidence databases that would allow a better-powered exploration of the sources of heterogeneity.

Besides the inconsistency, there are some other reasons why the quality of the evidence of the PPI included in this umbrella review is relatively low. First, most studies had unclear blinding, which may be difficult to implement in these settings. Second, most studies used a per-protocol analysis instead of conducting intention-to-treat analyses, a study limitation that may exaggerate the effects, may introduce bias, and it may even violate the principle of randomization.\textsuperscript{107}

The quality of the evidence of the studies assessing other interventions (e.g., leisure, physical activity, or yoga) was in general very low, especially due to a lack of large overall sample sizes. However, great caution is
warranted until the overall sample sizes are larger and the quality of the evidence higher. Indeed, some interventions only included one study. Future studies might provide evidence that these interventions do increase the SWB. At this regard, we want to highlight the important differences between quality of the evidence and effect size. The quality of the evidence helps us know how much we can trust in the overall results, independently of whether the results are that the interventions are efficacious or not. Therefore, we may have high-quality evidence that an intervention has no effect, or we may have low-quality evidence that an intervention has huge beneficial effects. In addition, here we only studied the effects on SWB. It is well-known, for example, that physical activity has other important positive outcomes.\(^\text{108}\)

Most individuals from highly developed countries report that their SWB is in the upper range, e.g., that they already feel very satisfied with life.\(^\text{109}\) This may result in a ceiling effect because an intervention will hardly increase the life satisfaction of an individual who already feels very satisfied with life. Therefore, it is possible that the effect size of the interventions would be larger in individuals who reported that their SWB is in the middle or lower ranges. One observation supporting this possibility is the finding that the efficacy of antidepressants is higher in patients with more severe depression,\(^\text{110}\) although this observation has been questioned.\(^\text{111}\) A related issue is that one could even question whether it is sensible to try to increase SWB in individuals with already high SWB. When comparing individuals with high and very high SWB, the latter are more successful in close relationships and volunteer work, but less successful in income, education, and political participation.\(^\text{112}\)

We would like to highlight some limitations of our review. First, umbrella reviews do not include those studies that have not been included in a systematic review or meta-analysis. The possibility that there are no systematic reviews or meta-analyses on the effects of a given intervention seems unlikely, because meta-analyses are nowadays performed massively.\(^\text{113}\) However, published systematic reviews and meta-analyses might not include recent trials. Second, our search strategy was designed to find interventions from any discipline, but it was probably sub-optimal for finding interventions of a specific discipline. For example, we probably failed to include some meta-analyses of PPI that we could possibly have identified using positive psychology terms such as “optimism” or “gratitude”. However, we considered that including additional positive psychology terms would bias the inclusion of studies towards PPI to the detriment of other interventions. We refer the reader to a recently published synthesis in the field of positive psychology \(^\text{28}\) that may have included some works that we failed to include. We would also want to note that for any kind of intervention for which we were able to find a meta-analysis, we should have theoretically been
able to include any RCT published until the date of the systematic search of the meta-analysis. Third, for studies that did not report pre-intervention scores or pre-post statistics, we used the effect size of the comparison between post-intervention scores. While suboptimal, this is equivalent to the effect size of the comparison between pre-post score differences under the general assumptions that pre-intervention mean scores are similar between the two groups, variances are similar, and pre-post correlation is about 0.5. Fourth, when the study reported several post-intervention and follow-up effect sizes, we averaged them. We considered that follow-up information is very relevant, given that we understand that the interest of these interventions is that the increase in SWB lasts a time, and thus we even considered using only the last effect size. However, effect size usually decreases with follow-up, and it would have been unfair to include the last, usually smaller effect size from studies conducting a long follow-up, while the initial, usually larger effect size from studies not conducting any follow-up after the intervention. To balance the situation, we preferred a medium consisting on averaging the effect sizes of the different follow-up points. In addition, as we noted earlier, such average has lower measurement error. Fifth, we focused on the common assessments of SWB, but other assessments are possible and may have indeed some advantages. For example, experience-sampling methods or ecological momentary assessments ask the participants their SWB at random moments of their everyday lives, potentially circumventing memory biases. Sixth, even when measuring the same components of SWB (positive affect, negative affect, and life satisfaction), the studies used different scales, which may have introduced heterogeneity in the analysis. Seventh, PPI were rather varied, a factor that may have increased heterogeneity. Finally, as we noted earlier, some other interventions included only one RCT and should be thus taken with more caution.

Future umbrella syntheses with more studies may be able to better model the heterogeneity between interventions. This may be achieved by stratifying the interventions, or more sophisticatedly, by conducting meta-regressions that model the relevant characteristics of each intervention. Additionally, there have been few studies in population subgroups scoring low in specific SWB components but not in others (which should benefit more from these interventions)\textsuperscript{114}, and little research on potential moderators.\textsuperscript{115,116} The study of subgroups and moderators when there are enough data may be of great interest, as it is entirely plausible that an intervention works for an individual but not for another. Finally, given the importance of personality for SWB, the creation of interventions that modulate the personality (e.g., decreasing neuroticism and increasing extraversion) could be also promising.
In conclusion, despite its limitations, this umbrella review shows that there is moderate-quality evidence that PPI increase SWB. Conversely, the evidence for other interventions (e.g., yoga, physical activity, or leisure) is still very low.
**Funding details**

This work was supported by Miguel Servet Research Contract (CP14/00041 and CPII19/00009) to J.R., PFIS Predoctoral Contract FI16/00311 to A.A.E. and Research Projects PI14/00292 and PI19/00394 from the Plan Nacional de I+D+i 2013–2016, the Instituto de Salud Carlos III-Subdirección General de Evaluación y Fomento de la Investigación and the European Regional Development Fund (FEDER, ‘Investing in your future’). The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

**Disclosure statement**

The authors have no interests to declare in relation to this manuscript.


Figure captions

Figure 1. Flow diagram of the inclusion of systematic reviews and meta-analyses and the inclusion of randomized controlled trials (RCT).
Otros archivos (Vídeo, etc.)

Click here to access/download

**Otros archivos (Vídeo, etc.)**
supplement.docx
Table 1. GRADE evidence profile for interventions aimed to increase individual subjective wellbeing (SWB), separately for each systematic review or meta-analysis

<table>
<thead>
<tr>
<th>Measure (N of RCT)</th>
<th>Limitations</th>
<th>Inconsistency</th>
<th>Indirectness</th>
<th>Imprecision</th>
<th>Publication bias</th>
<th>Number of participants</th>
<th>Hedges’ g (95% CI)</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Overall</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PA (12)</td>
<td>Serious limitations (some unclear blinding and losses to follow-up analyzed per protocol)</td>
<td>Serious inconsistency ($I^2=76%$, but only ~8% Hedges’ $g&lt;0$)</td>
<td>No serious indirectness</td>
<td>No serious imprecision</td>
<td>Undetected</td>
<td>652</td>
<td>581</td>
<td>0.44 (0.19, 0.68)</td>
</tr>
<tr>
<td>NA (11)</td>
<td>Serious limitations (some unclear blinding and losses to follow-up analyzed per protocol)</td>
<td>No serious inconsistency ($I^2=0%$)</td>
<td>No serious indirectness</td>
<td>No serious imprecision</td>
<td>Undetected</td>
<td>572</td>
<td>544</td>
<td>-0.19 (-0.31, 0.06)</td>
</tr>
<tr>
<td>LS (8)</td>
<td>Serious limitations (some unclear blinding and losses to follow-up analyzed per protocol)</td>
<td>No serious inconsistency ($I^2=20%$)</td>
<td>No serious indirectness</td>
<td>Serious imprecision (small overall sample size)</td>
<td>Suspected (funnel plot asymmetry)</td>
<td>385</td>
<td>327</td>
<td>0.22 (0.05, 0.39)</td>
</tr>
<tr>
<td><strong>Positive psychology interventions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conducting acts of kindness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PA (4)</td>
<td>Serious limitations (some unclear blinding and losses to follow-up analyzed per protocol)</td>
<td>Very serious inconsistency ($I^2=81%$, with 25% Hedges’ $g&lt;0$)</td>
<td>No serious indirectness</td>
<td>Very serious imprecision (small overall sample size and effect not excluded)</td>
<td>Suspected (funnel plot asymmetry)</td>
<td>355</td>
<td>362</td>
<td>n.s. (-0.05, 0.77)</td>
</tr>
<tr>
<td>NA (1)</td>
<td>Serious limitations (unclear blinding and losses to follow-up analyzed per protocol)</td>
<td>-</td>
<td>No serious indirectness</td>
<td>Very serious imprecision (small overall sample size and effect not excluded)</td>
<td>-</td>
<td>34</td>
<td>42</td>
<td>n.s. (-0.87, 0.04)</td>
</tr>
<tr>
<td>LS (5)</td>
<td>Serious limitations (some unclear blinding and losses to follow-up analyzed per protocol)</td>
<td>No serious inconsistency ($I^2=36%$)</td>
<td>No serious indirectness</td>
<td>No serious imprecision</td>
<td>Suspected (funnel plot asymmetry)</td>
<td>394</td>
<td>402</td>
<td>0.23 (0.02, 0.43)</td>
</tr>
<tr>
<td><strong>Practicing gratitude</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PA (3)</td>
<td>Serious limitations (unclear blinding)</td>
<td>Serious inconsistency ($I^2=54%$)</td>
<td>No serious indirectness</td>
<td>Serious imprecision (small overall sample size)</td>
<td>Undetected</td>
<td>164</td>
<td>148</td>
<td>n.s. (-0.14, 0.57)</td>
</tr>
<tr>
<td>NA (3)</td>
<td>Serious limitations (unclear blinding)</td>
<td>No serious inconsistency ($I^2=0%$)</td>
<td>No serious indirectness</td>
<td>Serious imprecision (small overall sample size)</td>
<td>Undetected</td>
<td>164</td>
<td>148</td>
<td>n.s. (-0.09, 0.36)</td>
</tr>
<tr>
<td>LS (2)</td>
<td>Serious limitations (unclear blinding)</td>
<td>No serious inconsistency ($I^2=7%$)</td>
<td>No serious indirectness</td>
<td>Serious imprecision (small overall sample size)</td>
<td>-</td>
<td>120</td>
<td>103</td>
<td>n.s. (-0.46, 0.26)</td>
</tr>
<tr>
<td><strong>Yoga</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PA (1)</td>
<td>Serious limitations (unclear blinding)</td>
<td>-</td>
<td>No serious indirectness</td>
<td>Very serious imprecision (very small overall sample size)</td>
<td>-</td>
<td>36</td>
<td>15</td>
<td>n.s. (-0.54, 0.66)</td>
</tr>
<tr>
<td>NA (1)</td>
<td>Serious limitations (unclear blinding)</td>
<td>-</td>
<td>No serious indirectness</td>
<td>Very serious imprecision (very small overall sample size and large effect not excluded)</td>
<td>-</td>
<td>36</td>
<td>15</td>
<td>n.s. (-0.87, 0.34)</td>
</tr>
<tr>
<td>Resilience training</td>
<td>PA (1)</td>
<td>Serious limitations (unclear blinding)</td>
<td>-</td>
<td>No serious indirectness</td>
<td>Very serious imprecision (very small overall sample size and large effect not excluded)</td>
<td>-</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>Treatment</td>
<td>NA (1)</td>
<td>Serious limitations (unclear blinding)</td>
<td>-</td>
<td>No serious indirectness</td>
<td>Very serious imprecision (very small overall sample size and large effect not excluded)</td>
<td>-</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>------------------------------------------------</td>
<td>--------</td>
<td>----------------------------------------</td>
<td>---</td>
<td>--------------------------</td>
<td>------------------------------------------------------------------------------------------</td>
<td>---</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>Physical activity</td>
<td>LS (1)</td>
<td>Serious limitations (unclear blinding)</td>
<td>-</td>
<td>No serious indirectness</td>
<td>Very serious imprecision (small overall sample size and large effect not excluded)</td>
<td>-</td>
<td>85</td>
<td>89</td>
</tr>
<tr>
<td>Leisure</td>
<td>LS (1)</td>
<td>Serious limitations (unclear blinding and losses to follow-up analyzed per protocol)</td>
<td>-</td>
<td>No serious indirectness</td>
<td>Very serious imprecision (very small overall sample size)</td>
<td>-</td>
<td>13</td>
<td>15</td>
</tr>
<tr>
<td>Control enhancement</td>
<td>LS (2)</td>
<td>Serious limitations (unclear blinding and losses to follow-up analyzed per protocol)</td>
<td>No serious inconsistency ($I^2=36%$)</td>
<td>No serious indirectness</td>
<td>Very serious imprecision (very small overall sample size)</td>
<td>-</td>
<td>34</td>
<td>34</td>
</tr>
<tr>
<td>Psychoeducational</td>
<td>LS (4)</td>
<td>Serious limitations (some unclear blinding and losses to follow-up analyzed per protocol)</td>
<td>Very serious inconsistency ($I^2=88%$, with ~25% Hedges' $g&lt;0$)</td>
<td>No serious indirectness</td>
<td>Very serious imprecision (small overall sample size and large effect not excluded)</td>
<td>Undetected</td>
<td>81</td>
<td>81</td>
</tr>
<tr>
<td>Miscellaneous (part-time work)</td>
<td>LS (1)</td>
<td>Serious limitations (unclear blinding and losses to follow-up analyzed per protocol)</td>
<td>-</td>
<td>No serious indirectness</td>
<td>Very serious imprecision (very small overall sample size and large effect not excluded)</td>
<td>-</td>
<td>23</td>
<td>23</td>
</tr>
</tbody>
</table>

LS: life satisfaction; NA: negative affect; n.s.: non-statistically significant; PA: positive affect; RCT: randomized controlled trials
# Table 2. GRADE evidence profile for interventions aimed to increase individual subjective wellbeing (SWB), combining all systematic reviews and meta-analyses

<table>
<thead>
<tr>
<th>Measure (N of RCT)</th>
<th>Limitations</th>
<th>Inconsistency</th>
<th>Indirectness</th>
<th>Imprecision</th>
<th>Publication bias</th>
<th>Number of participants</th>
<th>Hedges’ g (95% CI)</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Positive psychology interventions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PA (19)</td>
<td>Serious limitations (some unclear blinding and losses to follow-up analyzed per protocol)</td>
<td>Very serious inconsistency ($I^2=75%$, with $\sim 11%$ Hedges’ g&lt;0)</td>
<td>No serious indirectness</td>
<td>No serious imprecision</td>
<td>Suspected (funnel plot asymmetry)</td>
<td>1170</td>
<td>1092</td>
<td>-0.38 (0.20, 0.56)</td>
</tr>
<tr>
<td>NA (15)</td>
<td>Serious limitations (some unclear blinding and losses to follow-up analyzed per protocol)</td>
<td>No serious inconsistency ($I^2=17%$)</td>
<td>No serious indirectness</td>
<td>No serious imprecision</td>
<td>Undetected</td>
<td>770</td>
<td>734</td>
<td>-0.12 (-0.24, -0.002)</td>
</tr>
<tr>
<td>LS (14)</td>
<td>Serious limitations (some unclear blinding and losses to follow-up analyzed per protocol)</td>
<td>No serious inconsistency ($I^2=0%$)</td>
<td>No serious indirectness</td>
<td>No serious imprecision</td>
<td>Suspected (funnel plot asymmetry)</td>
<td>871</td>
<td>804</td>
<td>0.14 (0.04, 0.23)</td>
</tr>
<tr>
<td><strong>Other interventions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PA (2)</td>
<td>Serious limitations (unclear blinding)</td>
<td>No serious inconsistency ($I^2=0%$)</td>
<td>No serious indirectness</td>
<td>Very serious imprecision (very small overall sample size)</td>
<td>-</td>
<td>61</td>
<td>40</td>
<td>n.s. (-0.26, 0.56)</td>
</tr>
<tr>
<td>NA (2)</td>
<td>Serious limitations (unclear blinding)</td>
<td>No serious inconsistency ($I^2=0%$)</td>
<td>No serious indirectness</td>
<td>Very serious imprecision (very small overall sample size)</td>
<td>-</td>
<td>61</td>
<td>40</td>
<td>n.s. (-0.62, 0.20)</td>
</tr>
<tr>
<td>LS (9)</td>
<td>Serious limitations (some unclear blinding and losses to follow-up analyzed per protocol)</td>
<td>Very serious inconsistency ($I^2=79%$, but only 25% Hedges’ g&lt;0)</td>
<td>No serious indirectness</td>
<td>Very serious imprecision (small overall sample size and large effect not excluded)</td>
<td>Undetected</td>
<td>236</td>
<td>242</td>
<td>n.s. (-0.02, 0.85)</td>
</tr>
<tr>
<td><strong>All interventions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PA (21)</td>
<td>Serious limitations (some unclear blinding and losses to follow-up analyzed per protocol)</td>
<td>Serious inconsistency ($I^2=72%$)</td>
<td>No serious indirectness</td>
<td>No serious imprecision</td>
<td>Undetected</td>
<td>1231</td>
<td>1132</td>
<td>0.36 (0.20, 0.53)</td>
</tr>
<tr>
<td>NA (17)</td>
<td>Serious limitations (some unclear blinding and losses to follow-up analyzed per protocol)</td>
<td>No serious inconsistency ($I^2=11%$)</td>
<td>No serious indirectness</td>
<td>No serious imprecision</td>
<td>Undetected</td>
<td>831</td>
<td>774</td>
<td>-0.13 (-0.24, -0.02)</td>
</tr>
<tr>
<td>LS (23)</td>
<td>Serious limitations (some unclear blinding and losses to follow-up analyzed per protocol)</td>
<td>Serious inconsistency ($I^2=70%$)</td>
<td>No serious indirectness</td>
<td>No serious imprecision</td>
<td>Suspected (funnel plot asymmetry)</td>
<td>1107</td>
<td>1046</td>
<td>0.27 (0.09, 0.44)</td>
</tr>
</tbody>
</table>

LS: life satisfaction; NA: negative affect; n.s.: non-statistically significant; PA: positive affect; RCT: randomized controlled trials