Meta-analyses

Is chocolate consumption associated with health outcomes? An umbrella review of systematic reviews and meta-analyses

Nicola Veronese [a, b, *, 1], Jacopo Demurtas [c, 1], Stefano Celotto [d, 1], Maria Gabriella Caruso [b], Stefania Maggi [a], Francesco Bolzetta [e], Joseph Firth [f, g], Lee Smith [h], Patricia Schofield [i], Ai Koyanagi [j], Lin Yang [k], Marco Solmi [l], Brendon Stubbs [i, m, n]

[a] National Research Council, Neuroscience Institute, Aging Branch, Padova, Italy
[b] Ambulatory of Nutrition, IRCCS "S. de Bellis" National Institute of Gastroenterology-Research Hospital, Castellana Grotte, Bari, Italy
[c] Primary Care Department, Azienda USL Toscana Sud Est, Grosseto, Italy
[d] Primary Care Department, Aziendale AAS3 Alto Friuli — Collina — Medio Friuli, Udine, Italy
[e] Medical Department, Geriatric Unit, Azienda ULSS (Unità Locale Socio Sanitaria) 3 "Serenissima", Dolo-Mirano District, Italy
[f] NICM, School of Science and Health, University of Western Sydney, Australia
[g] Division of Psychology and Mental Health, University of Manchester, UK
[h] The Cambridge Centre for Sport & Exercise Sciences, Anglia Ruskin University, Cambridge, UK
[i] Faculty of Health, Social Care and Education, Anglia Ruskin University, Bishop Hall Lane, Chelmsford CM1 1SQ, UK
[j] Parc Sanitari Sant Joan de Déu/CIBERSAM, Universitat de Barcelona, Fundació Sant Joan de Déu, Sant Boi de Llobregat, Barcelona, Spain
[k] Department of Epidemiology, Center for Public Health, Medical University of Vienna, Austria
[l] University of Padova, Department of Neuroscience, Padova, Italy
[m] Physiotherapy Department, South London and Maudsley NHS Foundation Trust, Denmark Hill, London SE5 8AZ, UK
[n] Health Service and Population Research Department, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK

ARTICLE INFO

SUMMARY

Background & aims: The literature regarding the potential health benefits of chocolate consumption are unclear and the epidemiological credibility has not been systematically scrutinized, while the strength of the evidence is undetermined. We therefore aimed to map and grade the diverse health outcomes associated with chocolate consumption using an umbrella review of systematic reviews.
Methods: Umbrella review of systematic reviews of observational and intervention studies (randomized placebo-controlled trials, RCTs). For each association, random-effects summary effect size, 95% confidence interval, and 95% prediction interval were estimated. We also assessed heterogeneity, evidence for small-study effect and evidence for excess significance bias. For significant outcomes of the RCTs, the GRADE assessment was furtherly used.
Results: From 240 articles returned, 10 systematic reviews were included (8 of which included a meta-analysis), including a total of 84 studies (36 prospective observational studies and 48 interventional). Nineteen different outcomes were included. Among observational studies, including a total of 1,061,637 participants, the best available evidence suggests that chocolate consumption is associated with reduced risk of cardiovascular disease (CVD) death (n = 4 studies), acute myocardial infarction (n = 6), stroke (n = 5) and diabetes (n = 6), although this was based on a weak evidence of credibility. Across meta-analyses of intervention studies, chocolate consumption was positively associated with flow-mediated dilatation at 90—150 min (n = 3) and at 2—18 weeks (n = 3), and insulin resistance markers (n = 2). However, using the GRADE assessment, the evidence for these outcomes was low or very low. Data from two systematic reviews, reported that chocolate consumption was not associated with better depressive mood or cognitive function.
Conclusions: There is weak evidence to suggest that chocolate consumption may be associated with favorable health outcomes.

* Corresponding author. National Research Council, Neuroscience Institute, Aging Branch, Via Giustiniani, 2 35128 Padova, Italy. Fax: +39 0498211218.
  E-mail address: ilmannato@gmail.com (N. Veronese).
[1] These authors equally contributed to this research.

1. Introduction

The cocoa tree provides the basis for one of the world's most popular food products, i.e. chocolate. In 2015/2016, about 7.3 million tons of retail chocolate confectionery were consumed worldwide, with this figure expected to reach approximately 7.7 million tons by 2018/2019 [1]. Whilst there is evidence that excessive chocolate consumption is harmful for health [2], there is some evidence that eaten in small-moderate amounts, chocolate may have a number of health benefits [2]. Indeed, chocolate has been known from the ancient era as "kakawa", meaning "Food of the Gods" possibly for its health benefits [3]. Moreover, an ecological article reported that countries with a higher chocolate intake have a higher percentage of Nobel prize winners [4]. However, the findings of this study should be taken very cautiously due to the inherent biases of this approach.

There is increasing research from laboratory experiments and human studies suggesting that chocolate consumption may be beneficial for several health outcomes, particularly cardiovascular health [5]. Among the components present in chocolate, particular importance is given to flavonoids [6]. Flavonoids might be protective against cardiovascular disease (CVD) through several pathways, including their influence as antioxidant, antiplatelet, and anti-inflammatory agents [6] Similarly, flavonoids in chocolate might be able to improve other potential risk factors for CVD, such as hypercholesterolemia, hypertension as well as improve endothelial function [7].

Furthermore, chocolate might also have beneficial effects on other diseases such as on neurological diseases [2]. Similarly, it is commonly thought that chocolate has important anti-depressant effects, possibly through its anti-inflammatory mechanism and since it promotes the production of some neurotransmitters, such as serotonin [8,9].

A number of previous efforts to systematically appraise the evidence on chocolate have been undertaken, yet have focused on single disease endpoints, with a particular focus on CVD. However, the epidemiological credibility of this evidence for the health benefits of chocolate are unclear across the totality of the evidence. Here we used the umbrella review methodology (i.e. the syntheses and appraisal of existing systematic reviews) [10–12] to capture the breadth of outcomes associated with dietary chocolate intake. To identify health outcomes or medical conditions with the strongest evidence, we systematically assessed the quality and strength of the evidence across all health outcomes or medical conditions.

2. Materials and methods

2.1. Literature search and selection criteria

An umbrella review was carried out following standardized procedures [13,14]. Umbrella reviews provide important information that can be utilized by decision makers in healthcare to understand a broad topic area [14]. We systematically searched the MEDLINE/PubMed, Scopus, Embase, Cochrane Library/DARE databases from inception until 07th January 2018 with the following search strategy: "(chocolate OR cocoa) and (meta-an* or systematic review)". Next, we searched reference lists of eligible articles and so on until no further papers could be identified. We included formal systematic review with or without meta-analyses of observational and intervention studies investigating chocolate and any health outcome. No language restrictions were applied.

The primary screening (i.e., title/abstract screening) was carried out by two authors (NV, BS) and any disagreements were resolved via screening of the disagreed title/abstract by a third author (SM). Full-texts were sourced of all potentially eligible articles, and screened by two investigators (NV, BS) who determined the final references to be included. Conference abstracts were also considered.

We included: 1) peer-reviewed systematic reviews, with or without meta-analysis, that assessed chocolate intake using validated dietary questionnaires (e.g. food-frequency questionnaire [FFQ], 24 h recall, 7 days questionnaire etc.); 2) meta-analyses of observational studies (case–control or prospective cohort studies) that investigated the association of chocolate intake at baseline with any incident health-related outcomes (e.g. CVD, cancer, death, diabetes etc.); 3) meta-analyses of randomized controlled trials (RCTs), including at least one group using placebo. Studies that reported effect sizes - odds ratio (OR), relative risk (RR), or hazard ratio (HR) at follow-up or mean difference (MD) or standardized mean differences (SMD) — for the outcomes of interest were analyzed through a meta-analytic approach. The others were summarized descriptively.

Conversely, cross-sectional and RCTs studies using only active groups (e.g. people taking different forms of chocolate or other dietary products) were excluded. We finally excluded meta-analyses based on individual data without a systematic review of the literature.

2.2. Data extraction

The following information were extracted from each article by two independent investigators (SC, JD): (1) first author name; (2) year of publication; (3) journal; (4) type of chocolate (e.g. dark, milk, all types) and the categorization of chocolate intake (e.g. quintiles, quartiles, continuous); (5) number of included studies and participants in each meta-analysis; (6) the inclusion criteria for studied population; (7) effect size used in each meta-analysis; (8) study design (case–control, prospective, RCTs); (9) number of cases and controls in each study.

We then extracted the study-specific estimated relative risk, adjusted for the maximum number of covariates available for health outcome (RR, OR, HR), along with the 95% confidence intervals (CI), and the number of cases for each study by chocolate intake. For meta-analyses including RCTs, we extracted MDs and the number of people randomized to chocolate or placebo groups. For observational studies, assuming a linearity for the association between chocolate intake and health outcomes, lowest and highest quantiles were used in the final analyses and thus this data was extracted. Where two meta-analyses were found for the same association the analyses with the largest number of studies was included.

2.3. Data analysis

For each meta-analysis, summary effect size and 95% CI were estimated using fixed and random-effects models [15] To further consider between-study effects prediction intervals and corresponding 95% CI were calculated, this also estimates the certainty of the association if another study addresses that same association [16]. The standard error (SE) of the effect size was calculated for the largest dataset of each meta-analysis. If the SE was less than 0.10 then the 95% CI would be lower than 0.20 (which is less than the magnitude of a small effect size). To estimate between study associations the I^2 metric was used; high heterogeneity was indicated by values $\geq$50% and very high heterogeneity $\geq$75% [17,18].

We used the regression asymmetry test suggested by Egger and coworkers [19] to calculate evidence of small-study effects. A P value < 0.10 with more conservative effects in larger studies than in random-effects meta-analysis was considered as indicative of small-study effects [20].

The excess significance test [21] was used to evaluate whether the number of studies with nominally significant results (P < 0.05) included in a meta-analysis is too great based on the power of these data sets to detect effects at $\alpha = 0.05$. We calculated the power estimate for each data set. The sum of the power estimates of each study provides the expected (E) number of data sets with nominal statistical significance. The number of expected 'positive' (significant data sets) studies can be compared with the observed (O) number of significant studies through a χ^2-based test [21]. Greater the difference between O and E, the greater the degree of excess of significance bias. The true effect size of a meta-analysis is unknown. We considered the effect size of the largest dataset (i.e., with the lower SE), from this we calculated effect sizes of each constituent study utilizing an algorithm that employs a non-central t distribution. Where P < 0.10 excess significance for single meta-analysis was considered. For each meta-analysis O versus E comparison were carried out, separately. However, comparisons were extended to groups including a large number of meta-analyses after summing the O and E values of each individual meta-analysis.

2.4. Quality of the meta-analyses

We assessed the methodological quality of the included meta-analyses using AMSTAR [22]. We categorized the overall AMSTAR score as high (8–11 items achieved), moderate (4–7 items) and low (0–3 items) [22].

2.5. Credibility assessment

The credibility assessment criteria employed in the present review are based on established tools applied to observations evidence [10,13,23]. Evidence from meta-analyses of observational studies with nominally significant summary results (P < 0.05) were classified into four categories (class I, II, III, and IV). We considered as: 1) convincing (class I) associations with a statistical significance of $P < 10^{-6}$, included >1000 cases (or for continuous outcomes >20,000 participants), had the largest component study reporting a significant result (P < 0.05), had a 95% prediction interval excluding the null, large heterogeneity ($I^2 < 50\%$) was not present, and no evidence of small study effects (P > 0.10) was observed and of excess significance bias (P > 0.10); 2) highly suggestive (class II) evidence when associations reported a significance of P < 0.001, included >1000 cases (or for continuous outcomes >20,000 participants), and had the largest component study reporting a statistically significant result (P < 0.05); 3) suggestive (class III) evidence when associations that reported a significance of P < 0.01 with >1000 cases (or for continuous outcomes >20,000); 4) weak (class IV) evidence for significant associations with P < 0.05.

Evidence from meta-analyses of randomized controlled trials was assessed in terms of the significance of the summary effect (P < 0.01, $0.01 \leq P < 0.05$, $P \geq 0.05$), 95% prediction interval (excluding the null or not), and presence of large heterogeneity ($I^2 \geq 50\%$), small study effects (P > 0.10), and excess significance (P > 0.10) [23]. When the P-value for the random effect was <0.05, we evaluated the evidence using the GRADE (Grading of Recommendations, Assessment, Development and Evaluation) assessment [24].

3. Results

3.1. Literature review

As shown in Fig. 1, the literature review identified 240 unduplicated papers. After applying the inclusion or exclusion criteria, 22 full-texts were identified and of them 10 were eligible, including 8 meta-analyses and two systematic reviews without meta-analysis. Across the included studies was a total of 84 eligible original studies [25–34]. These systematic reviews and meta-analyses included 19 health outcomes (Table 1). Ten of these outcomes were derived from meta-analyses including RCTs, and the other nine from observational studies, all from longitudinal cohort studies.

3.2. Meta-analyses of observational studies

As reported in Table 1, seven outcomes were included, the median number of studies for each outcome was 5 (range 4–6), the median number of participants was 144,823 (57,709 to 322,732) for a total of 1,061,637 people, and the median number of cases was 8749 (4553 to 16,626), although this last information was available only for three meta-analyses. All the outcomes included cardiovascular diseases or cardiovascular mortality, except one which considered incident diabetes as the outcome (Table 1).

Overall, 4 out of the 7 outcomes reported nominally significant summary results (P < 0.05), but none of them survived the application of a more stringent P value ($P < 1 \times 10^{-6}$), the lowest P-value being 0.001.

The study with the largest sample size for each database had a SE of less than 0.10 for all the outcomes (except for stroke) and a more conservative effect compared to the random-effects model in six outcomes. Heterogeneity among studies was present for two outcomes, diabetes and CVD death. Only one association (stroke) presented a 95% prediction interval excluding the null value. The evidence for excess statistical significance (i.e. whether smaller studies tend to give substantially larger estimates of effect size compared with larger studies) was present for two outcomes, i.e. atrial fibrillation and heart failure.

Based on the above criteria, all the significant outcomes (CVD death, acute myocardial infarction, stroke and diabetes) have a weak level of evidence (Table 1) and for the other three outcomes (coronary artery disease, atrial fibrillation and heart failure), no significant association was found.

The median AMSTAR score was 6 (range: 2–7), indicating a moderate quality of these meta-analyses (Supplementary Table 1).

3.3. Findings from systematic review without meta-analysis

Two systematic reviews (without formal meta-analysis) investigating depression/depressive mood and cognitive disorders as outcomes were included. The results of the first systematic review [26] indicate that consumption of chocolate had no protective effects against depression, even if these findings were limited only to one longitudinal study, whilst the RCTs included in this review did not use placebo and consequently were not eligible for our work [26]. The other systematic review reported no significant effect on cognitive function in one small placebo-RCT [25].

3.4. Meta-analyses of randomized placebo-controlled trials

As reported in Table 2, ten outcomes were included, the median number of studies included for each outcome was 4 (range 2–11), the median number of participants was 177 (median 89 randomized to chocolate and 88 to placebo) for a total of 2601 people. All the outcomes included cardiovascular or metabolic outcomes (Table 2).

Overall, only three (flow-mediated dilatation 90–150 min, flow-mediated dilatation 2–18 weeks and insulin resistance markers) out of the 10 outcomes reported nominally significant summary results (P < 0.05) and, of them, no one survived the application of a more stringent P value ($P < 1 \times 10^{-6}$), being the lowest P-
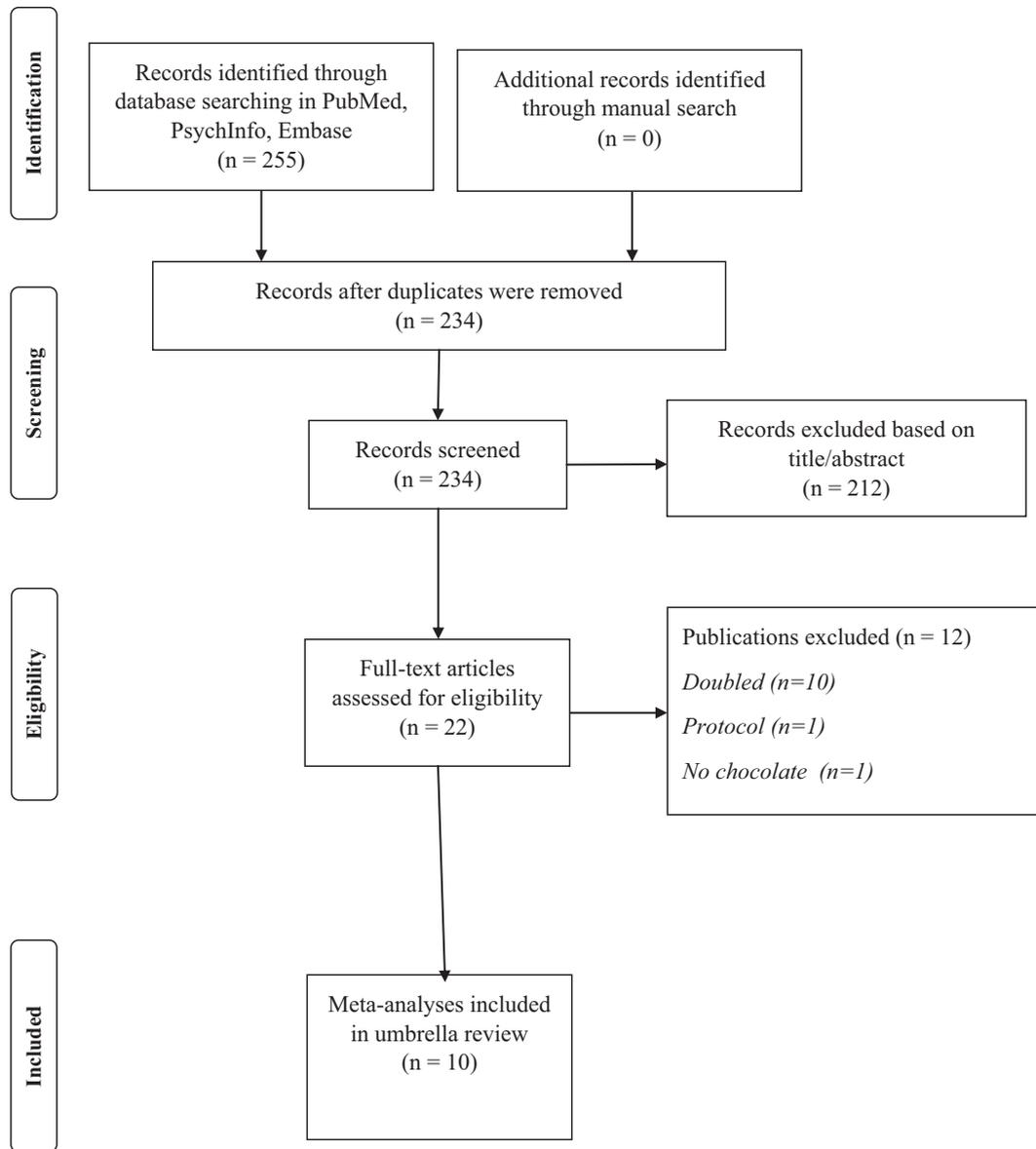
Fig. 1. PRISMA flow-chart.

value = 0.002. However, using the GRADE assessment the certainty of evidence for these outcomes was very low or low, mainly due to limited sample sizes included (Table 3).

The study with the largest sample size of each database reported significant results only for one outcome (i.e. flow mediated dilatation at 90−150 min). Heterogeneity among studies was absent ($I^2 < 50\%$) for all the outcomes, except two (HDL cholesterol and flow-mediated dilatation at 90−150 min). No outcome presented 95% prediction interval excluding the null value. The evidence for excess statistical significance (i.e. whether smaller studies tend to give substantially larger estimates of effect size compared with larger studies) was absent in all the outcomes included.

The median AMSTAR score was 5 (range: 3−10), indicating a moderate quality of these meta-analyses (Supplementary Table 2).

4. Discussion

In this study, including 84 original studies and over 1 million people that consumed chocolate, we found weak evidence that

chocolate consumption is associated with lower CVD death, acute myocardial infarction, stroke and diabetes. We arrived at this position by evaluating the epidemiological credibility of the evidence, an approach which has been used in other research specialties [12,35], including nutritional epidemiology [36]. This critical appraisal of the literature is necessary, since often researchers widely use a nominal significance level at P < 0.05 to claim novel associations. However, increasing research is showing that findings based on this threshold can only constitute a weak evidence and there are ongoing debates on redefining the level of statistical significance using more stringent criteria [37]. For example, in our umbrella review, seven (four observational and three interventional) outcomes were statistically significant (as P < 0.05), but no convincing or highly suggestive evidence was evident for any of the outcomes included.

In observational studies, we found a weak evidence that chocolate could be beneficial for the prevention of CVD and for diabetes. Altogether, these findings suggest that, even if these outcomes are significant, they are affected by some limitations and or biases that

Table 1
Health outcomes and evidence class reported in included meta-analyses of observational studies.

Outcome (reference)	Population	Level of comparison	N of studies	N of participants	N of cases	Risk ratio (95% CI)	P-value	95% CI prediction intervals	I^2	Publication bias	P-value for excess significance test	Small study effect	Largest study significant	Evidence class[a]
CVD death [28]	Healthy	Highest vs. lowest category and continuous	4	57,709	NA	0.650 (0.465–0.909)	0.01	0.15–2.74	82.6	Yes	NA	Yes	Yes	IV
AMI [27]	Healthy and general population	Highest vs. lowest category	6	144,823	8749	0.895 (0.822–0.975)	0.01	0.74–1.08	24.3	Yes	No	No	Yes	IV
Stroke [32]	General population	Highest vs. lowest category	5	322,732	NA	0.858 (0.786–0.937)	0.001	0.74–0.99	0	No	NA	No	No	IV
Diabetes [32]	General population	Highest vs. lowest category	6	146,385	NA	0.842 (0.725–0.978)	0.02	0.56–1.26	54.2	Yes	NA	Yes	No	IV
Coronary artery disease [32]	General population	Highest vs. lowest category	4	104,514	NA	0.885 (0.775–1.010)	0.07	0.55–1.42	47.3	No	NA	No	Yes	NS
Atrial fibrillation [31]	General population	Highest vs. lowest category	6	180,534	16,626	0.963 (0.898–1.032)	0.284	0.87–1.06	0	No	Yes	No	No	NS
Heart failure [30]	General population	Highest vs. lowest category	4	104,940	4553	0.908 (0.815–1.013)	0.08	0.72–1.15	0	No	Yes	No	No	NS
Summary statistics			**Median = 5 (range: 4–6)**	**Median = 144,823 (range: 57,709 to 322,732)**	**Median = 8749 (range: 4553 to 16,626)**									

Abbreviations: AMI: acute myocardial infarction; CVD: cardiovascular disease; CI: confidence intervals; NA: not available.

[a] Evidence class criteria: class I (convincing): statistical significance with $P < 10^{-6}$, more than 1000 cases (or >20,000 participants for continuous outcomes), the largest component study reported statistically significant effect (P < 0.05); 95% prediction interval excluded the null; no large heterogeneity ($I^2 < 50\%$), no evidence of small study effects (P > 0.10) and excess significance bias (P > 0.10); class II (highly suggestive): statistical significance with $P < 10^{-6}$, more than 1000 cases (or >20,000 participants for continuous outcomes), the largest component study reported statistically significant effect (P < 0.05); class III (suggestive): statistical significance with $P < 10^{-3}$, more than 1000 cases (or >20,000 participants for continuous outcomes); class IV (weak): the remaining statistically significant associations with P < 0.05.

Table 2
Health outcomes and evidence class reported in included meta-analyses of randomized placebo-controlled studies.

Outcome (reference)	N of studies	N of participants (treated; placebo)	Mean difference (95% CI)	P-value	95% CI prediction intervals	I^2	Publication bias	P-value for excess significance test	Small study effect	Largest study significant
Flow mediated dilatation (90 −150 min) [26]	3	134 (67; 67)	4.452 (1.605−7.299)	0.002	−27.99; 36.89	76.1	No	No	No	Yes
Flow mediated dilatation (2−18 weeks) [26]	3	87 (43; 44)	2.111 (0.735−3.488)	0.003	−6.81; 11.03	0	No	No	No	No
Insulin resistance markers [26]	2	49 (24; 25)	−0.461 (−0.899 to −0.023)	0.039	NA	0	No	NA	NA	No
Fasting insulin [26]	3	88 (44; 44)	−1.940 (−4.863 to 0.983)	0.193	−26.15; 22.27	20.3	No	No	No	No
Fasting plasma glucose [26]	4	127 (63; 64)	−0.07 (−0.282; 0.143)	0.522	−0.54; 0.40	0	No	No	No	No
LDL cholesterol [25]	4	220 (111; 109)	−3.063 (−11.00; 4.875)	0.450	−20.49; 14.36	0	No	No	No	No
HDL cholesterol [25]	4	220 (111; 109)	−1.328 (−7.262; 4.605)	0.661	−25.45; 22.80	61.3	No	No	No	No
Triglycerides [25]	4	220 (111; 109)	−10.09 (−25.50; 5.32)	0.199	−43.92; 23.73	0	No	NA	NA	No
Systolic blood pressure [29]	18	728 (365; 363)	−0.93 (−2.635; 0.769)	0.285	−6.86; 4.99	64.5	No	No	No	No
Diastolic blood pressure [29]	18	728 (365; 363)	−1.117 (−2.592; 0.237)	0.103	−6.37; 4.02	69.8	No	No	No	No
Summary statistics	**Median = 4 (range: 2−18)**	**Median = 177 (89; 88)**								

Abbreviations: CI: confidence intervals; NA: not available; NS: not significant.
Evidence class criteria: class I (convincing): statistical significance with P < 0.01; 95% prediction interval excluded the null; no large heterogeneity ($I^2 < 50\%$), no evidence of small study effects (P > 0.10) and excess significance bias (P > 0.10); class II (highly suggestive): statistical significance with $0.01 \leq P < 0.05$; class III: no significant association (P > 0.05).

globally discourage the use of chocolate for the prevention of these medical conditions. We can try to explain these findings through some hypotheses. First, the kinds of chocolate included were heterogeneous, whilst it is reported that dark chocolate might give its consumers health benefits, the milk variety cannot probably for an higher presence of flavonoids and anti-oxidant components [38]. Unfortunately, the studies included in our syntheses did not differentiate between the consumption of dark and milk chocolate. Second, chocolate is rich added sugars and added sugar consumption seems to be associated with an increased risk for CVD [39]. Moreover, as shown by a large study in European people, persons eating more frequently chocolate had more frequently an unhealthier diet, e.g. they eat less frequently vegetables and fruits and introduce less amounts of fibers [40]. Even if the analyses were adjusted for all these confounders, it is also possible that the beneficial effect of chocolate were counterbalanced by these unhealthy habits. Second, three of the four statistically significant outcomes were affected by publication bias, suggesting that more original works with significant results were published than one would normally expect [41] Finally, the populations included at baseline in these observational studies were heterogeneous, since they included healthy participants or at high risk of a developing a condition (e.g. overweight for diabetes).

We were able to find two systematic reviews regarding the possible role of chocolate on two health outcomes, i.e. depression and cognitive dysfunction, suggesting that chocolate consumption has no impact on these outcomes. However, these findings were unfortunately limited to one longitudinal study (for depression) and to one small RCT (for cognitive outcomes). Therefore, other studies are needed to clarify these associations.

Finally, the findings of the meta-analyses of the RCTs included confirmed the results of observational studies. Of the ten outcomes (all regarding cardiovascular and metabolic health), only three were statistically significant at a P-value<0.05. Using the GRADE assessment for these significant outcomes, the certainty of the evidence was low or very low, mainly due to the limited sample sizes used in these RCTs. The effect of chocolate on cardiometabolic health is attributed to arterial dilatation, a mechanism supported by clinical and animal research. It was hypothesized that NOX-2, the catalytic subunit of NADPH oxidase, has a key role in the formation of reactive oxidant species and is involved in impairment of flow-mediated dilation [42]. Chocolate (particularly dark) exerts artery dilatation via down-regulating NOX2-mediated oxidative stress, and it seems to improve walking ability in patients affected by peripheral artery disease [43]. Moreover, the regulation of nitric oxide (NO) production by the flavonoids present in dark chocolate

Table 3
GRADE assessment of the significant outcomes in meta-analyses of randomized controlled trials.

№ of participants (studies)	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall certainty of evidence
Flow mediated dilatation (90−150 min)						
134 (3 RCTs)	Not serious	Very serious[a]	Not serious	Serious[b]	None	⊕◯◯◯ Very low
Flow mediated dilatation (2−18 weeks)						
87 (3 RCTs)	Not serious	Not serious	Not serious[c]	Very serious[c]	None	⊕⊕◯◯ Low
Insulin resistance markers						
49 (2 RCTs)	Not serious	Not serious	Not serious	Very serious[c]	None	⊕⊕◯◯ Low

Abbreviations: CI: Confidence interval.
[a] $I^2 \geq 75\%$.
[b] Less than 200 participants.
[c] Less than 100 participants.

could explain its effects on insulin resistance [44]. However, the findings of RCTs were small in sample size (median n = 177), and they also used different kinds of chocolate products and usually had a short follow-up period (in median 8 weeks). Moreover, many of the RCTs included in the original meta-analyses were not eligible, since they frequently used as control other kinds of chocolate, e.g. dark vs. milk chocolate. Therefore, we need future interventional trials in order to better understand the role of chocolate in improving these health outcomes.

Whilst this large umbrella review provides important novel insights, a number of limitations should be considered. A credibility assessment criteria was used that was based on established tools for observational evidence [23]. Even if none of the components of these criteria provides definitive proof of lack of reliability, they cumulatively include the possibility that the results are susceptible to bias and uncertainty. Meta-analyses contained studies that differed in their design, population and other characteristics. To account for this, we used an $I^2 < 50\%$ as one of the criteria for class I evidence (convincing) in order to assign the best evidence grade only to robust associations and without any suspect of bias. Second, meta-analyses have limitations [45] and results likely depend on decisions relating to which estimates are selected from each primary study and how to apply them in the meta-analysis (e.g. in the present review a number of meta-analyses failed to report information on types/kinds of chocolate or combined active and placebo groups in the RCTs). This may be a relevant issue when the factor of interest is continuous, and when different comparisons of levels of the risk factor may be selected to express risk [46]. Finally, the meta-analyses of RCTs, as mentioned above, included studies with a low quality, mainly due to a limited sample size. In this regard, we believe that our umbrella review using more stringent criteria for assessing the credibility of the evidence and including systematic reviews that consider many treatment comparisons for the management of the same disease or condition [13] could better approach the issue if chocolate is beneficial or not for health outcomes. Moreover, except for the meta-analyses cited in our work, the other works available for this topic are mainly narrative reviews, consequently limiting the generalization of these findings in daily clinical practice.

5. Conclusions

Our umbrella review of the top tier of evidence from systematic reviews and meta-analyses suggests that there is weak evidence for chocolate consumption on health outcomes such as diabetes and cardiovascular conditions. Therefore, our results did not support routinely use of chocolate for improving or preventing cardiovascular and metabolic diseases. However, we encourage future studies (both observational and intervention) to confirm/refuse our findings.

Authors' contribution

Screening: Stubbs, Veronese. Data extraction: Demurtas, Celotto; Drafting of the paper: Veronese, Bolzetta, Stubbs, Yang, Celotto; Statistical analysis:Veronese, Solmi, Koyanagi; critical revision: Caruso, Maggi, Firth, Smith, Schofield. All the authors approved the final version to be submitted.

Sources of support

Ai Koyanagi's work was supported by the Miguel Servet contract financed by the CP13/00150 and PI15/00862 projects, integrated into the National R + D + I and funded by the ISCIII − General Branch Evaluation and Promotion of Health Research − and the European Regional Development Fund (ERDF-FEDER).

Conflicts of interest

None.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.clnu.2018.05.019.

References

[1] ICCO. Challenges and opportunities of the global cocoa and chocolate sector & position of the Russian market. 2015. https://www.statista.com/statistics/238849/global-chocolate-consumption/.
[2] Latif R. Chocolate/cocoa and human health: a review. Neth J Med 2013;71(2): 63–8.
[3] De Araujo QR, Gattward JN, Almoosawi S, Silva M, Dantas PA, De Araujo Junior QR. Cocoa and human health: from head to foot—a review (1549−7852 (electronic)). Crit Rev Food Sci Nutr 2016;56(1):1−12.
[4] Messerli FH. Chocolate consumption, cognitive function, and nobel laureates. N Engl J Med 2012;367(16):1562−4.
[5] Larsson SC. Coffee, tea, and cocoa and risk of stroke. Stroke 2014;45(1): 309−14.
[6] Corti R, Flammer AJ, Hollenberg NK, Luscher TF. Cocoa and cardiovascular health. Circulation 2009;119(10):1433−41.
[7] Latham LS, Hensen ZK, Minor DS. Chocolate−guilty pleasure or healthy supplement? J Clin Hypertens 2014;16(2):101−6.
[8] Silva NR. Chocolate consumption and effects on serotonin synthesis. Arch Intern Med 2010;170(17):1608. author reply 1608−1609.
[9] Rose N, Koperski S, Golomb BA. Mood food: chocolate and depressive symptoms in a cross-sectional analysis. Arch Intern Med 2010;170(8):699−703.
[10] Bellou V, Belbasis L, Tzoulaki I, Evangelou E, Ioannidis JP. Environmental risk factors and Parkinson's disease: an umbrella review of meta-analyses. Park Relat Disord 2016;23:1−9.
[11] Dinu M, Pagliai G, Casini A, Sofi F. Mediterranean diet and multiple health outcomes: an umbrella review of meta-analyses of observational studies and randomized trials. Nutr Metabol Cardiovasc Dis 2017;27(1):e21.
[12] Theodoratou E, Tzoulaki I, Zgaga L, Ioannidis JPA. Vitamin D and multiple health outcomes: umbrella review of systematic reviews and meta-analyses of observational studies and randomised trials. BMJ Br Med J 2014;348.
[13] Ioannidis JP. Integration of evidence from multiple meta-analyses: a primer on umbrella reviews, treatment networks and multiple treatments meta-analyses. CMAJ 2009;181(8):488−93.
[14] Aromataris E, Fernandez R, Godfrey CM, Holly C, Khalil H, Tungpunkom P. Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. Int J Evid Base Healthc 2015;13(3): 132−40.
[15] Lau J, Ioannidis JP, Schmid CH. Quantitative synthesis in systematic reviews. Ann Intern Med 1997;127(9):820−6.
[16] Higgins JP, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. J R Stat Soc Ser A Stat Soc 2009;172(1):137−59.
[17] Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002;21(11):1539−58.
[18] Ioannidis JP, Patsopoulos NA, Evangelou E. Uncertainty in heterogeneity estimates in meta-analyses. BMJ 2007;335(7626):914−6.
[19] Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997;315(7109):629−34.
[20] Carvalho AF, Köhler CA, Brunoni AR, Miskowiak KW, Herrmann N, Lanctôt KL. Bias in peripheral depression biomarkers. Psychother Psychosom 2016;85(2): 81−90.
[21] Ioannidis JP, Trikalinos TA. An exploratory test for an excess of significant findings. Clin Trials 2007;4(3):245−53.
[22] Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Meth 2007;7(1):10.
[23] Li X, Meng X, Timofeeva M, Tzoulaki I, Tsilidis KK, Ioannidis JP. Serum uric acid levels and multiple health outcomes: umbrella review of evidence from observational studies, randomised controlled trials, and Mendelian randomisation studies. BMJ 2017;357.
[24] Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008;336(7650):924.
[25] Scholey A, Owen L. Effects of chocolate on cognitive function and mood: a systematic review. Nutr Rev 2013;71(10):665−81.
[26] Garcia-Blanco T, Davalos A, Visioli F. Tea, cocoa, coffee, and affective disorders: vicious or virtuous cycle? J Affect Disord 2017;224:61−8.

[27] Tokede OA, Gaziano JM, Djousse L. Effects of cocoa products/dark chocolate on serum lipids: a meta-analysis. Eur J Clin Nutr 2011;65(8):879–86.

[28] Hooper L, Kay C, Abdelhamid A, Kroon PA, Cohn JS, Rimm EB. Effects of chocolate, cocoa, and flavan-3-ols on cardiovascular health: a systematic review and meta-analysis of randomized trials. Am J Clin Nutr 2012;95(3):740–51.

[29] Larsson SC, Akesson A, Gigante B, Wolk A. Chocolate consumption and risk of myocardial infarction: a prospective study and meta-analysis. Heart (British Cardiac Society) 2016;102(13):1017–22.

[30] Gianfredi V, Salvatori T, Nucci D, Villarini M, Moretti M. Can chocolate consumption reduce cardio-cerebrovascular risk? A systematic review and meta-analysis. Nutrition 2018;46:103–14.

[31] Ried K, Fakler P, Stocks NP. Effect of cocoa on blood pressure. Cochrane Database Syst Rev 2017;(4), CD008893.

[32] Gong F, Yao S, Wan J, Gan X. Chocolate consumption and risk of heart failure: a meta-analysis of prospective studies. Nutrients 2017;9(4).

[33] Larsson SC, Drca N, Jensen-Urstad M, Wolk A. Chocolate consumption and risk of atrial fibrillation: two cohort studies and a meta-analysis. Am Heart J 2018;195:86–90.

[34] Yuan S, Li X, Jin Y, Lu J. Chocolate consumption and risk of coronary heart disease, stroke, and diabetes: a meta-analysis of prospective studies. Nutrients 2017;9(7).

[35] Belbasis L, Savvidou MD, Kanu C, Evangelou E, Tzoulaki I. Birth weight in relation to health and disease in later life: an umbrella review of systematic reviews and meta-analyses. BMC Med 2016;14(1):147.

[36] Veronese N, Solmi M, Caruso MG, Giannelli G, Osella AR, Evangelou E. Dietary fiber and health outcomes: an umbrella review of systematic reviews and meta-analyses. Am J Clin Nutr 2018;107:436–44.

[37] Johnson VE. Revised standards for statistical evidence. Proc Natl Acad Sci Unit States Am 2013;110(48):19313–7.

[38] Serafini M, Bugianesi R, Maiani G, Valtuena S, De Santis S, Crozier A. Plasma antioxidants from chocolate. Nature 2003;424(6952). 1013–1013.

[39] Yang Q, Zhang Z, Gregg EW, Flanders WD, Merritt R, Hu FB. Added sugar intake and cardiovascular diseases mortality among US adults. JAMA internal medicine. 2014;174(4):516–24.

[40] Buijsse B, Weikert C, Drogan D, Bergmann M, Boeing H. Chocolate consumption in relation to blood pressure and risk of cardiovascular disease in German adults. Eur Heart J 2010;31(13):1616–23.

[41] Dalton JE, Bolen SD, Mascha EJ. Publication bias: the elephant in the review. Anesth Analg 2016;123(4).

[42] Ray R, Murdoch CE, Wang M, Santos CX, Zhang M, Alom-Ruiz S. Endothelial Nox4 NADPH oxidase enhances vasodilatation and reduces blood pressure in vivo. Arterioscler Thromb Vasc Biol 2011;31(6):1368–76.

[43] Loffredo L, Perri L, Catasca E, Pignatelli P, Brancorsini M, Nocella C. Dark chocolate acutely improves walking autonomy in patients with peripheral artery disease. J Am Heart Assoc 2014;3(4).

[44] Fraga CG. Cocoa, diabetes, and hypertension: should we eat more chocolate? Am J Clin Nutr 2005;81(3):541–2.

[45] Ioannidis JP. The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. Milbank Q 2016;94(3): 485–514.

[46] Kavvoura FK, Liberopoulos G, Ioannidis JP. Selection in reported epidemiological risks: an empirical assessment. PLoS Med 2007;4(3):e79.