Food and Chemical Toxicology xxx (2017) 1–64

Contents lists available at ScienceDirect

Food and Chemical Toxicology

journal homepage: www.elsevier.com/locate/foodchemtox

Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children

Daniele Wikoff [a, *], Brian T. Welsh [b], Rayetta Henderson [c], Gregory P. Brorby [d], Janice Britt [e], Esther Myers [f], Jeffrey Goldberger [g], Harris R. Lieberman [h], Charles O'Brien [i], Jennifer Peck [j], Milton Tenenbein [k], Connie Weaver [l], Seneca Harvey [m], Jonathan Urban [b], Candace Doepker [n]

[a] ToxStrategies, Asheville, NC, USA
[b] ToxStrategies, Austin, TX, USA
[c] ToxStrategies, Research Triangle Park, NC, USA
[d] ToxStrategies, San Francisco Bay Area, CA, USA
[e] ToxStrategies, Tallahassee, FL, USA
[f] EF Myers Consulting, Trenton, IL, USA
[g] University of Miami Miller School of Medicine, Miami, FL, USA
[h] Military Nutrition Division, US Army Research Institute of Environmental Medicine, Natick, MA, USA
[i] Department of Psychiatry, University of Pennsylvania, Philadelphia, PA, USA
[j] College of Public Health, University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA
[k] Department of Pediatrics and Child Health, Department of Community Health Sciences, Children's Hospital, University of Manitoba, Winnipeg, Manitoba R3A 1S1, Canada
[l] Department of Nutrition Science, Purdue University, West Lafayette, IN, USA
[m] ToxStrategies, Houston, TX, USA
[n] ToxStrategies, Cincinnati Area, OH, USA

ARTICLE INFO

Article history:
Received 22 January 2017
Received in revised form
21 March 2017
Accepted 3 April 2017
Available online xxx

Keywords:
Caffeine
Systematic review
Safety
Pregnancy
Coffee
Behavior

ABSTRACT

To date, one of the most heavily cited assessments of caffeine safety in the peer-reviewed literature is that issued by Health Canada (Nawrot et al., 2003). Since then, >10,000 papers have been published related to caffeine, including hundreds of reviews on specific human health effects; however, to date, none have compared the wide range of topics evaluated by Nawrot et al. (2003). Thus, as an update to this foundational publication, we conducted a systematic review of data on potential adverse effects of caffeine published from 2001 to June 2015. Subject matter experts and research team participants developed five PECO (population, exposure, comparator, and outcome) questions to address five types of outcomes (acute toxicity, cardiovascular toxicity, bone and calcium effects, behavior, and development and reproduction) in four healthy populations (adults, pregnant women, adolescents, and children) relative to caffeine intake doses determined not to be associated with adverse effects by Health Canada (comparators: 400 mg/day for adults [10 g for lethality], 300 mg/day for pregnant women, and 2.5 mg/kg/day for children and adolescents). The *a priori* search strategy identified >5000 articles that were screened, with 381 meeting inclusion/exclusion criteria for the five outcomes (pharmacokinetics was

Abbreviations: ADHD, attention deficit hyperactivity disorder; AHRQ, Agency for Healthcare Research and Quality; BMC, bone mineral content; BMD, bone mineral density; BMI, body mass index; CDC, US Centers for Disease Control and Prevention; CDH, chronic daily headache; CI, confidence interval; CNS, central nervous system; COI, conflict of interest; COMT, catechol-*O*-methyl transferase; CVD, cardiovascular disease; CVM, cardiovascular malformation; DGAC, US Dietary Guidelines Advisory Committee; DSM, Diagnostic and Statistical Manual of Mental Disorders; ED, emergency department; ESCALE, Epidemiological Study on Childhood Cancer and Leukemia; FDA, US Food and Drug Administration; HDL, high-density lipoprotein; HF, high frequency; ICD, International Classification of Diseases; ILSI North America, North American Branch of the International Life Sciences Institute; IOM, Institute of Medicine; IUGR, intrauterine growth restriction; LD, limb defect; LDL, low-density lipoprotein; LF, low frequency; LOEL, lowest observed effect level; NOEL, no observed effect level; NTD, neural tube defect; OHAT, National Toxicology Program's Office of Health Assessment and Translation; OR, odds ratio; OTC, over-the-counter; PD, pharmacodynamics; PECO, population, exposure, comparator, and outcome; PK, pharmacokinetics; POMS, Profile of Mood States; RCT, randomized controlled trial; RR, relative risk; SAB, scientific advisory board; SDNN, standard deviation of NN intervals; SGA, small for gestational age; SR, systematic review; SRDR, AHRQ Systematic Review Database Repository; VAS, visual analogue scale.

* Corresponding author. ToxStrategies, Asheville, NC, USA.
E-mail addresses: dwikoff@toxstrategies.com (D. Wikoff), efmyers@efmyersconsulting.com (E. Myers), j-goldberger@miami.edu (J. Goldberger), harris.r.lieberman.civ@mail.mil (H.R. Lieberman), obrien@mail.med.upenn.edu (C. O'Brien), Jennifer-Peck@ouhsc.edu (J. Peck), mtenenbein@exchange.hsc.mb.ca (M. Tenenbein), weavercm@purdue.edu (C. Weaver), cdoepker@toxstrategies.com (C. Doepker).

http://dx.doi.org/10.1016/j.fct.2017.04.002

addressed contextually, adding 46 more studies). Data were extracted by the research team and rated for risk of bias and indirectness (internal and external validity). Selected no- and low-effect intakes were assessed relative to the population-specific comparator. Conclusions were drawn for the body of evidence for each outcome, as well as endpoints within an outcome, using a weight of evidence approach. When the total body of evidence was evaluated and when study quality, consistency, level of adversity, and magnitude of response were considered, the evidence generally supports that consumption of up to 400 mg caffeine/day in healthy adults is not associated with overt, adverse cardiovascular effects, behavioral effects, reproductive and developmental effects, acute effects, or bone status. Evidence also supports consumption of up to 300 mg caffeine/day in healthy pregnant women as an intake that is generally not associated with adverse reproductive and developmental effects. Limited data were identified for child and adolescent populations; the available evidence suggests that 2.5 mg caffeine/kg body weight/day remains an appropriate recommendation. The results of this systematic review support a shift in caffeine research to focus on characterizing effects in sensitive populations and establishing better quantitative characterization of interindividual variability (e.g., epigenetic trends), subpopulations (e.g., unhealthy populations, individuals with preexisting conditions), conditions (e.g., coexposures), and outcomes (e.g., exacerbation of risk-taking behavior) that could render individuals to be at greater risk relative to healthy adults and healthy pregnant women. This review, being one of the first to apply systematic review methodologies to toxicological assessments, also highlights the need for refined guidance and frameworks unique to the conduct of systematic review in this field.

1. Introduction

Caffeine (1,3,7-trimethylxanthine) is a pharmacologically active component of many foods, beverages, dietary supplements, and drugs; it is also used to treat very ill newborns afflicted with apnea (temporary cessation of breathing). Caffeine occurs naturally in some plant leaves, seeds, and fruits, where it serves as an herbicide, insect repellant, and even attractant for pollination (Lee et al., 2009; Wright et al., 2013). This botanically sourced compound is the most commonly consumed stimulant worldwide (Fredholm et al., 1999). Caffeine enters the human food chain through plant-derived foods such as coffee beans, tea leaves, guarana, cocoa beans, and kola nuts (Barone and Roberts, 1996). Coffee is one of the major contributors of caffeine to the diet (Mitchell et al., 2015); since the late 1980s, the energy drink market has emerged as another source of caffeine in the diet (Richards and Smith, 2016). Consumption practices were recently investigated by Mitchell et al. (2014), who reported that 85% of the US population ingests at least one caffeine-containing beverage per day, and by Fulgoni et al. (2015), who reported that 89% of the population uses caffeine in some form. In addition to standard beverages, a number of other caffeinated products, such as maple syrup, beef jerky, and donuts, have entered the market, suggesting substantial consumer interest in diverse sources of caffeine. Because of the variation in caffeine content of beverages due to a wide range in dose and infusion times as well as unexpected sources of caffeine appearing on the market, assessing exposure to caffeine has a great deal of uncertainty. This, in combination with uncertainty in the use of dietary intake surveys, as well as simultaneous exposure to many substances when consuming the various caffeine sources, are well-known limitations in the evidence base.

Caffeine is generally recognized as safe by the US Food and Drug Administration (FDA) at a use level not to exceed 200 ppm (0.02%) in cola-type beverages for the specific intended use of flavor (21CFR§182.1180). Caffeinated beverages, like coffee, have been consumed for centuries. Current estimates suggest that the mean consumption of caffeine (all ages) is 165 mg/day, ~105 mg of which is associated with coffee consumption (Mitchell et al., 2015). Emergence of other products containing caffeine, particularly energy drinks, combined with controversies regarding the potential for increased consumption by nonadult populations (Drewnowski and Rehm, 2016; McGuire, 2014), has been accompanied by concerns regarding the impact of these products on consumer health. Regulatory agencies worldwide, including those in the United States, Europe, Canada, New Zealand, India, and Australia, have evaluated caffeine safety, and several agencies have issued guidance regarding daily intake amounts (DGAC, 2015; EFSA, 2015; Milanez, 2011; Nawrot et al., 2003; for a summary of the 2015 DGAC conclusions, see Millen et al., 2016). The most widely cited of these values is from Health Canada (Nawrot et al., 2003), in which the agency authors conducted a comprehensive (but not systematic) literature search and concluded in a peer-reviewed publication that an intake dose of up to 400 mg caffeine/day was not associated with adverse effects in healthy adults. Nawrot et al. (2003) also concluded that consumption of up to 300 mg/day for pregnant women and 2.5 mg/kg/day for children is not associated with adverse effects.

Since the Nawrot et al. (2003) article was published, >10,000 papers have been published related to caffeine, >5000 of which address effects or exposure in humans. In addition, >800 reviews related to various human health effects and caffeine have also been published (i.e., nearly all are specific to a particular adverse endpoint category), but a robust, transparent, and systematic assessment of the health effects associated with caffeine consumption in humans is not yet available in the peer-reviewed literature. Therefore, we conducted a systematic review (SR) of data published since 2003 and through 2015 to update the review by Nawrot et al. (2003). Specifically, our objective was to determine whether the literature published since the 2003 Health Canada review supports the conclusions that caffeine consumption at amounts up to 400 mg/day for healthy adults, 300 mg/day for healthy pregnant women, and 2.5 mg/kg-day for healthy children is not associated with adverse effects. We also evaluated consumption of 2.5 mg/kg caffeine/day in adolescents, although this was not specifically addressed by Nawrot et al. (2003).

In developing their conclusions, Nawrot et al. (2003) reviewed many outcomes; however, given the voluminous scope, this effort was limited to evaluation of potential effects for five main outcomes: (1) acute toxicity (defined herein as abuse, overdose, and potential death), (2) cardiovascular, (3) bone and calcium, (4) behavior, and (5) development and reproductive toxicity. The areas of genotoxicity, mutagenicity, and carcinogenicity were not included. These endpoints were selected based on relative importance as documented in other comprehensive evaluations (EFSA,

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

2015; IARC, 1991; Milanez, 2011; Loomis et al., 2016; Nawrot et al., 2003) and stakeholder interest. The areas of pharmacokinetics (PK) and pharmacodynamics (PD) were also of interest but this topic area was not considered to be reviewed systematically; rather, information was reviewed to provide contextual evidence (OHAT, 2015a). That is, because the general PK/PD of caffeine is well understood, the specific objective was to summarize any advances in knowledge. We were particularly interested in any new information with respect to differences and similarities between populations of interest, characterization of PK in nonadult populations of interest, and characterization of PK in the context of the five main areas.

Our SR was conducted using the Institute of Medicine's (IOM) *Finding What Works in Health Care—Standards for Systematic Reviews* as guidance (Eden et al., 2011). This document provides standards for (1) initiating a SR, (2) finding and assessing individual studies, (3) synthesizing the body of evidence, and (4) reporting SRs. Per the IOM framework, additional methods are required for individual study assessment and body of evidence assessment. For these aspects, we utilized the National Toxicology Program Office of Health Assessment and Translation (OHAT) framework for evidence integration (OHAT, 2015a), as well as the OHAT Risk of Bias tool (OHAT, 2015b). These references were issued subsequent to project initiation but prior to protocol registration, and they were selected for their specific application to toxicology (versus clinical medicine). It is anticipated that the transparency in reporting the information reviewed, as well as the integrated conclusions, will provide value to scientists and stakeholders interested in this issue of caffeine safety. Since Health Canada's work is so commonly referenced in nearly every discussion of caffeine safety, validating whether or not the Nawrot et al. (2003) conclusions remain current provides a foundation for establishing an acceptable level of protection to the healthy general population. This foundation, in turn, provides confidence in assuring the typical healthy caffeine consumer of a reasonable certainty of no harm, and it also allows scientists to move away from this question and focus more on the sensitive subpopulations that may be at greater risk. It is anticipated that this review will be of utility to a variety of stakeholders, including doctors, dietitians and other health professionals in guiding their patient populations, as well as consumers interested in understanding caffeine safety.

2. Methods

2.1. Establishing team and protocol development

The SR was structured using the IOM publication, *Finding What Works in Health Care—Standards for Systematic Reviews* as guidance (Eden et al., 2011). Consistent with the IOM-recommended standards for initiating a review, the first step involved establishing a team with appropriate expertise and experience. In addition to eight scientists from ToxStrategies, which included a caffeine expert (C.D.) and a SR expert (D.W.), the team also included seven scientific advisory board (SAB) members with expertise in the following areas: SR (E.M.), caffeine (C.O., J.G., J.P., H.R.L., and M.T.), epidemiology (J.P., C.W.), bone and calcium (C.W.), reproduction (J.P.), behavior (H.R.L., C.O.), PK (M.T.), acute toxicity (M.T., C.O.), and clinical medicine (J.G., M.T., C.O.).

Each study team member and each SAB member completed a comprehensive conflict of interest (COI) questionnaire, which documented both financial and nonfinancial COIs via questions regarding investments, employment, consultancies, contracts/grants, patents/royalties/trademarks, expert witness testimony, speaking/writing, past financial interests, other involvements/relationships, personal beliefs, previously published opinions, institutional relationships, career advancement, advocacy/policy positions, other positions, and caffeine consumption. COI declarations were reviewed and the relative likelihood of creating bias in decision making was evaluated using an internal process documented via "Management of Conflicts of Interest within Systematic Review Team and Scientific Advisory Board Members" (internal document, established March 2015). This document included actions for situations in which a team member was determined to have a high degree of COI, defined in this document and based on IOM definitions. Declarations were documented in the registered protocols (PROSPERO protocol nos. CRD42015026704, CRD42015027413, CRD42015026673, CRD42015026609, and CRD42015026736; https://www.crd.york.ac.uk/PROSPERO/). After review of the declared COIs, it was determined, overall, that these would not exert undue influence on the primary interest of the SR. The sponsor, the North American Branch of the International Life Sciences Institute (ILSI North America), was made aware of and given an opportunity to comment on the depth and scope of the SR when there were financial implications; however, the review team had independence in making final decisions about the design, analysis, and reporting of the SR.

A protocol for the SR was developed, and elements included (1) context and rationale for the review, (2) study selection and screening criteria, (3) descriptions of outcome measures, time points, and comparison groups, (4) search strategy, (5) procedures for study selection, (6) data extraction strategy, (7) approach for critically appraising individual studies, and (8) method for evaluating the body of evidence. A protocol for each outcome was registered on PROSPERO (PROSPERO protocol nos. CRD42015026704, CRD42015027413, CRD42015026673, CRD42015026609, and CRD42015026736; https://www.crd.york.ac.uk/PROSPERO/). In the present manuscript, the first element is described in the introduction, and the remaining elements are discussed in the subsequent text. As described in the introduction, the premise for the overall effort was to evaluate the literature published since the Nawrot et al. (2003) work to determine whether the conclusions reached by Health Canada were still supported by the updated literature.

The PECO (population, exposure, comparator, and outcome) question for the SR was as follows: "For [population], is caffeine intake above [dose], compared to intakes [dose] or less, associated with adverse effects on [outcome]?" This SR focused on five outcomes (Fig. 1): acute, cardiovascular, bone and calcium, behavior, and development and reproduction (further descriptions of the endpoints included within each of these outcomes can be found in the results section of each outcome). A sixth outcome, PK, was included as a contextual topic (literature was identified systematically; information was not subject to the individual study assessment and body of evidence evaluation described in the protocol, but rather was reviewed and reported in a narrative format) consistent with practices described by OHAT (2015a). For PK, the objective was to generally characterize the current understanding of caffeine kinetics and critically review any information that advances the science, particularly with respect to differences/similarities between our populations of interest, characterization of kinetics in children and adolescent populations of interest, and characterization of kinetic parameters (particularly fast/slow phenotypes) in the context of the outcomes of interest.

The series of SRs evaluated four populations: healthy adults, healthy pregnant women, healthy adolescents (aged 12–19 years), and healthy children (aged 3–12 years). Only studies in humans were included (i.e., animal studies were excluded). "Healthy" subjects were defined as individuals who were not specifically described as having been hospitalized or diagnosed with disease and/or receiving medical treatment for a disease at the time of the

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

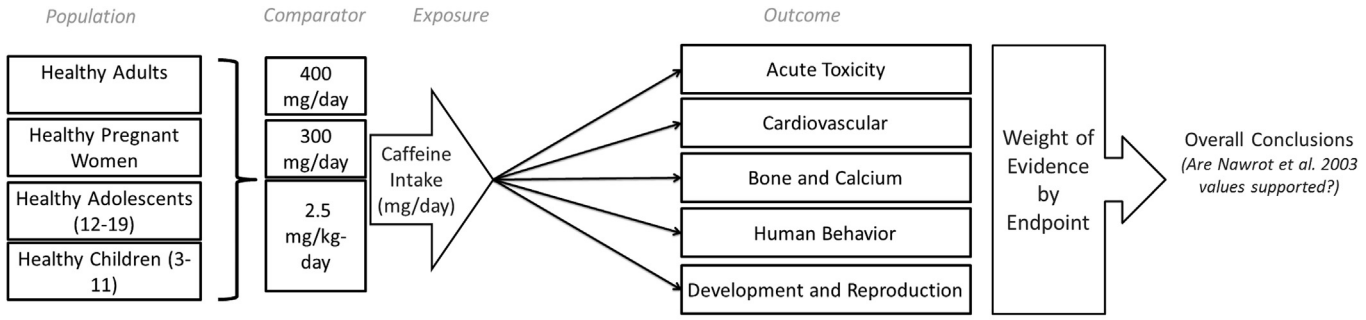


Fig. 1. Diagram illustrating the analytical framework of the systematic review based on the populations (P), exposure (E), comparator (C), and outcome (O).

study. As such, studies evaluating a healthy population (which included athletes, military personnel, and pregnant women, unless otherwise noted as unhealthy) were included. Studies in which healthy individuals were included as a control group (or similar) as part of a study on unhealthy populations (e.g., individuals with asthma) were included; however, only information from the healthy individuals was used in the assessment.

For all outcomes except acute, the exposure values in the PECO were 400 mg/day, 300 mg/day, and 2.5 mg/kg body weight/day for adults, pregnant women, and adolescents and children, respectively; similarly, comparators were ≤400 mg/day for adults, ≤300 mg/day for pregnant women, and ≤2.5 mg/kg body weight/day for adolescents and children. Although Nawrot et al. (2003) discussed other values in context of each of the outcomes, the selected values were based on the overall conclusions they presented. The exposure and comparator dose for lethality (acute outcome) was 10 g for adults but was undefined for the other populations based on the lack of a comparator from Nawrot et al. (2003). Inclusion and exclusion criteria unique to specific outcomes are documented in the PROSPERO registrations.

2.2. Study screening and selection

To be included in the SR, studies had to provide a quantitative estimate or measurement of exposure to a caffeine source associated with an adverse effect. Forms of caffeine included coffee, tea, chocolate, cola-type beverages, energy drinks, supplements, medicines, energy shots, caffeinated chewing gum, caffeinated sport gel, and caffeinated sport bars. Studies evaluating the effects of caffeine alone, in one of the aforementioned forms, or in combination with one or more compounds occurring in the approved sources at amounts designed to match constituents of valid sources (e.g., caffeine and green tea extract) were included. Studies that did not provide a quantitative exposure to an acceptable caffeine source associated with an adverse effect were excluded (e.g., studies that evaluated only decaffeinated coffee/tea and caffeine placebo exposures, exposures where participants were expecting caffeine but did not receive the drug, or studies that evaluated yerba mate, guarana, damiana [caffeine-containing plant], contaminants of caffeine, and/or caffeine metabolites, etc.). Studies that evaluated the effects of caffeine in combination with either another pharmacologically active compound in over-the-counter (OTC) pain relievers (e.g., acetaminophen plus caffeine) or with nicotine, alcohol, or a prescribed drug were excluded.

Studies had to be peer reviewed and available in English for inclusion. Only studies evaluating exposure and response at the individual amount were included (e.g., ecological studies were excluded). Case reports and case series were included only for the acute outcome and were excluded for the other outcomes. Rationale for inclusion of case reports and case series in the acute outcome is based on the unique nature of the outcome (i.e., death, rare events) (CRD, 2009; Fitzpatrick-Lewis et al., 2009) and the lack of data from more reliable study types (which would result in an inability to evaluate an outcome that is important to understanding the safety of caffeine). Letters to the editor that contained original, peer-reviewed data were included. Reviews were excluded from the systematic assessment unless original data, such as meta-analyses, were reported. Although they are not commonly included in a SR, relevant meta-analyses were included to inform the PECO (Eden et al., 2011). Selected reviews were also retained for context, though a critical appraisal and inclusion of reviews was beyond the scope of this assessment. Careful consideration was given to studies evaluating beneficial or therapeutic endpoints (referred to as benefit studies); those that reported parameters or effects associated with adverse effects in an outcome of interest were included, whereas those only reporting on beneficial endpoints were excluded.

The search strategy was developed via an iterative process involving evaluation, validation, and piloting of a variety of databases with syntax unique to each. The process, as well as the final strategy, was informed and reviewed by a librarian with expertise in the conduct of SRs. Three databases were searched: PubMed, EMBASE, and the Cochrane Database of Systematic Reviews (grey literature was not included, primarily due to the volume of primary data available via standard databases). Syntax was developed for each database; terms related to each outcome, as well as caffeine, were run in a concatenated fashion (found at: http://www.crd.york.ac.uk/PROSPEROFILES/26704_STRATEGY_20150829.pdf). Search restrictions included default functions for language (English) and data (restricted to publications between January 1, 2001 and June 8, 2015). EMBASE searches were exclusive of MEDLINE and restricted to selected journals (430 journals were selected based on relevance; expert librarian determined an initial list of journals not indexed by PubMed, the list of journals was refined by project team members using keywords associated with SR topics and expert judgement). The Cochrane library was searched between January 2001 and June 2015 for review articles. All databases were searched on June 8, 2015; articles published after this date were not considered in the SR.

Multiple software tools were considered to facilitate the SR; however, given the complexity and volume of information reviewed, DistillerSR was selected to facilitate and document screening, selection, extraction, and evaluation of data. Results of the search were entered into an EndNote database for identification and removal of duplicates and were then uploaded into a DistillerSR library. Following a series of pilot screening exercises, multiple

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

evidence analysts conducted the full screen of titles and abstracts. In instances where determinations of inclusion/exclusion could not be made due to limited information in the title/abstract, the result was included and carried forward for full text review. The screening process also involved group review of a large number of hits that were deemed by screeners to be ambiguous. A designation of "Needs further discussion—Internal" was made on the screening form for these entries and screeners could optionally provide more detail on the specific issue in the "Notes" field of the form. These titles and abstracts were then carefully reviewed and discussed by two or three other members of the team before assigning inclusion or exclusion status after consensus. The dynamic screening form included categorization of study type and outcome, noting that some studies evaluated multiple outcomes. Following completion of the screen, each SAB member was provided a tabular summary to conduct a second set of reviews of the included/excluded results.

Full text articles of all studies identified for potential inclusion following the screen were obtained. Some articles were not obtainable with reasonable effort, which involved the following hierarchy: search of the National Library of Medicine, direct purchase from the publisher/journal (which included inquires to journals when publications could not be readily identified online), or request sent to the corresponding author. Evidence analysts were assigned to specific outcomes for the subsequent step; extensive piloting exercises were first conducted to refine the individual study assessment process (including the data extraction form), as well as to ensure consistency in form responses across analysts and outcomes. Analysts then reviewed the full text of each article for the assigned outcome; if the article met the inclusion criteria, then the information from the study was extracted and the study was evaluated for quality. Data extraction was facilitated via a second DistillerSR form that included two sets of information: (1) basic information as reported by the author (i.e., direct extraction of information from the text) and (2) customized information (i.e., information to inform the PECO questions, dose/exposure calculations, categorization, or interpretation applied). Basic information fields included the following: outcome(s) and endpoint(s), objective, methods, study design categorization, source of caffeine exposure, exposure metric (measured, self-report), population, reference/comparison, confounders, results, dose-response evaluation, conclusion, pharmacokinetic information, funding sources, and COIs. Customized fields or project-specific fields included the following: standardization of exposure/dose to SR, selection of endpoints and exposure/dose for comparisons, and information related to assessing trends, consistency, and so forth across the body of evidence. During extraction, the level of adversity of the endpoints within the study (Guyatt et al., 2011) was also characterized. All data extracted were placed in a freely available Agency for Healthcare Research and Quality (AHRQ) Systematic Review Database Repository (SRDR) repository (Acute - https://srdr.ahrq.gov/projects/1115; Behavioral - https://srdr.ahrq.gov/projects/1116; Bone and Calcium - https://srdr.ahrq.gov/projects/1062; Cardiovascular - https://srdr.ahrq.gov/projects/1114; Reproductive and Developmental - https://srdr.ahrq.gov/projects/1118). It should be noted that much of the information in the basic information fields is based on text directly from the authors; given the volume and scope of the assessment, significant efforts were not devoted to summarizing such information in the extraction forms. Following extraction, each SAB member was provided a tabular summary to conduct a second set of reviews of the extracted information.

Because the SR involves comparison of caffeine exposures in the literature to values reported by Nawrot et al. (2003), standardization of the exposure metric was a critical step in data extraction. The comparators were in a metric of mass per day (mg/day) or mass per day based on body weight (mg/kg-day). However, studies in the literature often reported a variety of metrics (e.g., milligrams of caffeine, cups of coffee per day, etc.). A decision tree was developed to specify a consistent process of recording and standardizing the results as reported by authors to enable comparison to the PECO. The process for standardizing the exposure metric can be found in Supplementary File S1. When selecting values for comparison to Nawrot et al. (2003), all eligible comparisons from a given study were made (i.e., all relevant populations and subgroups as well as comparisons for multiple endpoints were selected, if sufficient data were available). In order to be characterized as an observed effect (i.e., a lowest observed effect level; non-significant findings were characterized as no observed effect levels), the finding had to be statistically significant; in cases where multiple results were presented, findings from the most sophisticated or refined analysis performed by the authors were selected. Comparators were characterized by the SR authors as no observed effect levels (NOELs) and lowest observed effect levels (LOELs).

2.3. Individual study assessment

Following data extraction, individual studies were assessed for risk of bias (internal validity) using the OHAT (2015b) *Risk of Bias Rating Tool for Human and Animal Studies*. The body of evidence was evaluated and integrated using the OHAT (2015a) *Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review an Evidence Integration*. For ease of readership, the risk of bias evaluation is referred to as study quality, though it was recognized that in doing so, we are referring to internal validity. For graphical purposes, meta-analyses were assigned to the highest-quality end of the Risk of Bias spectrum. Two refinements to the evaluation of RoB was implemented post protocol registration. Question 11, "Other," was included and modified to further addressed confidence in exposure characterization. While Question 8 addressed confidence from the perspective of the type of exposure data (e.g., dietary survey), Question 11 was included to address the purity of caffeine. Responses were as follows: +2 pure caffeine w/ purification; +1 pure caffeine w/o purity; −1 mixture with estimate of caffeine used in analysis; −2 all other. For selected endpoints in reproduction and development (fetal growth, spontaneous abortion, recurrent miscarriage, stillbirth, and preterm birth and small for gestational age), RoB Question 4 (Did the study design or analysis account for important confounding and modifying variables) was refined to place emphasis on the pregnancy signal as a confounder (see the reproductive and development section and the discussion section for biological significance and rationale for placing emphasis on this variable). Because of the known complexities and lack of validity in approaches to evaluate the pregnancy signal (Brent et al., 2011; Lawson et al., 2004; Peck et al., 2010; Stein and Susser, 1991), many studies were assigned a probably high risk of bias based primarily on the methods used to evaluate the pregnancy signal. If authors did not attempt to evaluate the pregnancy signal, the study was scored as high risk of bias for this element (Q4).

Following extraction of the data, it was determined that guidance beyond that provided by OHAT (2015a) or Money et al. (2013) for evaluating individual studies, as well as integrating the body of evidence, were required. Thus, we integrated aspects of the commonly applied GRADE (Grades of Recommendation, Assessment, Development and Evaluation) process presented by Guyatt et al. (2011). GRADE is a well-established process supported by the IOM framework. Specifically, GRADE was used to categorize the level of importance in decision making, which, in context of this review was regarded as the level of adverseness (Guyatt et al., 2011). To do so, endpoints were categorized as (1) physiological or clinical and (2) high/medium/low with respect to the importance

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

of the effect in decision making. These categorizations were used to weight endpoints when developing outcome conclusions. As part of the weight of evidence (described below), analysts applied expert judgement in considering how physiological endpoints were related to clinical outcomes (e.g., known predictors, etc.), as well as how considerations were made regarding the event(s) relative to the progression of the outcome. Several tools were used to facilitate and support the evaluation, including generation of evidence tables (See Section 2.2 for AHRQ links to individual outcomes), risk of bias heat maps, summary plots of selected NOEL/LOEL data from individual studies, and a tabular summary of the confidence in the evidence for each outcome and endpoint.

2.4. Body of evidence assessment

Consistent with the framework established by the IOM (Eden et al., 2011), the body of evidence was synthesized qualitatively for each outcome using methods recommended within the IOM standards for SR, complemented by those offered by OHAT (2015a) given the specific application to toxicological assessments. In evaluating and conducting a qualitative synthesis the body of evidence for each outcome, findings were summarized relative to the Nawrot et al. (2003) comparators of 400 mg/day for adults (10 g for lethality), 300 mg/day for pregnant women, and 2.5 mg/kg/day for children and adolescents. Data are described based on the volume of data above and below the comparator, as well as the types of effects and quality of evidence of data that are above and below the comparator. Also consistent with the IOM framework (Eden et al., 2011), confidence in the body of evidence was assessed using the approach determined *a priori*. An initial level of confidence was assigned based primarily on the following key features of the study design: controlled exposure, exposure prior to outcome, individual outcome data, and comparison group used (OHAT, 2015a).

Then, using expert judgement, a number of additional factors were considered for the overall body of evidence, rendering increases or decreases in the confidence. These factors included the following: overall risk of bias, indirectness, magnitude of effect, confounding, and overall consistency. With respect to the magnitude of effect, considerations were given both to a) when studies observed an effect below the comparator, and b) the overall weight of the evidence related to the magnitude, which included both considerations of lack of effect (e.g., number of studies without effects, essentially a "0" level of magnitude) as well as magnitude in studies reporting effects below the comparator (as reported in the narrative weight of evidence conclusions and summarized in tabular format). These factors both help characterize the data as well as provide structure for developing a conclusion based on the strength and confidence in the underlying body of evidence with respect to effects relative to the comparator (i.e., effects relative to a specific intake vs. the potential for effects to occur at any intake). Consideration of endpoint importance (Guyatt et al., 2011) or level of adverseness of the endpoints was also important in making weight of the evidence conclusions, both for the individual outcomes and the overall assessment. Typically, this discussion was in the context of classifying endpoints as clinical or physiological, and for the latter, if sufficient information is known to interpret physiological data relative to clinical outcomes.

With respect to evaluation of dose-response, it was ultimately determined not to be a good fit for determining confidence in the body of evidence relative to the research question, and thus was not included as part of the weight of evidence considerations. Because the research question involved evaluation of potential effects at or below a specific intake (the comparator), integration of dose-response as a parameter that would increase confidence in a body of evidence would require evaluation, primarily, of dose-response relationships in studies which both (a) reported effects, and (b) reported effects at or below the comparator. A number of factors precluded a comprehensive integration of dose-response across the body of evidence, including, though not limited to: many studies did not report effects, many studies that reported effects below the comparator were single dose studies, studies which evaluated dose-response did not provide quantitative information that would aid in interpretation of such relative to the comparator (i.e, did not provide information as to effects at or below the comparator). Future evaluations could involve targeted evaluation of the strength of the bodies of evidence in the context of dose-response relationships for studies reporting effects below the comparator.

Similar to the approach and conclusions of Nawrot et al. (2003), the objective in the weight of evidence assessment was not to find the most protective or lowest amount associated with an effect per se, but rather to make a determination based on the body of evidence as a whole, which included considerations for positive and negative findings, quality of data, level of adversity, consistency, and magnitude of effect (for studies with effects below the comparator). Weight of the evidence determinations were made by endpoint and population; specifically, conclusions were developed by categorizing evidence relative to the comparator (an intake value not associated with adverse effects) as follows: comparator is acceptable (i.e., evidence supports the Nawrot et al. (2003) conclusions regarding intake), comparator is too high (i.e., evidence suggests the comparator is too high), or comparator is too low (i.e., evidence suggests the comparator could be higher). Using a similar approach, conclusions were also developed for the outcome, as well as overall conclusions. When developing outcome conclusions, clinical endpoints with a high level of adversity were given the most weight. Conclusions were not developed for endpoints containing fewer than five studies; in these instances, summary thoughts are provided but data were determined to be insufficient to determine a conclusion.

3. Results

Following removal of duplicates, 5706 records of human studies were identified via the multiple databases searched (Fig. 2). Titles and abstracts were screened for potential inclusion. The most common reasons for exclusion during title and abstract review were as follows: outcomes not included in the SR (e.g., cancer), unhealthy populations, coexposures, benefit/therapy studies, and *in vitro* studies. Following committee reviews, internal quality-control efforts, and SAB review of title and abstract screening, 740 records were carried forward to full text review. Based on initial characterizations of the outcome (e.g., cardiovascular, reproductive), analysts conducted full text reviews in which the first step was to confirm inclusion/exclusion. The most common reason for exclusion following full text review was lack of quantitative information required to evaluate the data relative to the comparator (Fig. 1) (e.g., reports of positive or negative associations, but lack of a specific exposure associated with such). Following full text review, a total of 426 studies were included in this SR; a small portion of these studies evaluated multiple outcomes (e.g., cardiovascular and behavior, or reproductive toxicity and PK) and were thus assessed in all appropriate evaluations. Often, a number of endpoints were reported within each study (e.g., heart rate and blood pressure in a cardiovascular study). The number of endpoints ranged from one to six per study and averaged two endpoints per study.

Almost half of the studies (42%) specifically evaluated caffeine as a source; the majority of the remaining studies evaluated coffee (21%), tea (12%), and soda (9%) as a source of caffeine, whereas the

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

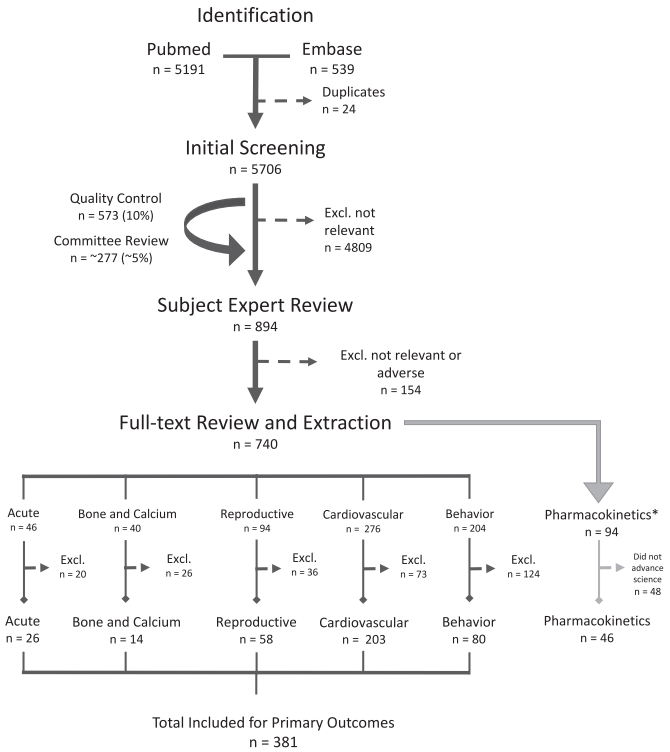


Fig. 2. Summary of literature search and screening process to identify relevant peer-reviewed publications for five primary outcomes (acute, bone and calcium, reproductive, cardiovascular, and behavior) and one contextual outcome (pharmacokinetics).

other studies evaluated caffeine from energy drinks, chocolate, medicine, and other sources. In 77% of the studies, the exposure (dose) of caffeine did not need to be standardized (i.e., the author either evaluated caffeine directly or reported findings based on the amount of caffeine in the given sources; see Supplementary File S1). In ~5% of the studies, multiple metrics were reported—that is, authors reported findings based on both caffeine content/amount as well as source amount (e.g., cups of coffee). In these cases, the caffeine-based data were utilized in this SR. As such, the amount of caffeine was estimated by applying standardized caffeine values by source in <20% of the studies. Exposure was measured in 63% of the studies and was self-reported in 38%.

With respect to study type, more than half of the studies (63%) were controlled trials. The remaining were observational studies as follows: cohort studies (14%), case-control studies (9%), cross-sectional studies (5%), and meta-analyses (2%). Seven percent of the publications were case reports or case series, all of which were associated with the acute outcome (these were excluded for other outcomes). The majority of the literature (79%) identified and reviewed involved adult populations. Literature characterizing the outcomes of interest in other populations was much more limited, including studies that involved pregnant women (14%), adolescents (aged 12–19 years) (4%), or children (aged 3–11 years) (2%).

In the subsequent sections, the findings for each outcome are reported. Each section is structured similarly to provide: (1) an overview of the literature identified for the outcome relative to the PECO, (2) narrative reviews of the data by endpoint, and (3) a qualitative body of evidence assessment for the outcome. The narrative discussions are intended to be succinct summaries, consistent with that provided by Nawrot et al. (2003), characterizing the findings as reported in the literature. In-depth critical assessment of individual studies was beyond the scope of this SR (rather findings reported by authors were relied upon); however, each study was evaluated for internal validity. Results of the risk of bias assessment are provided as figures in each section; color coding was utilized to represent the overall area of the spectrum for the study, with dark green representing an overall lower risk of bias (and thus higher quality) and dark red representing an overall higher risk of bias (and thus lower quality). The internal validity (i.e., study quality) is also considered in the accompanying plots of all of the data by endpoint; the size of the symbol represents where the study falls on the risk of bias spectrum (with larger symbols indicating a lower risk of bias, and so on). These plots are meant to provide a visual display of selected data obtained from the literature that could be compared to Nawrot et al. (2003); specifically, these plots include selected effect levels (i.e., LOELs), or no effect levels (i.e., NOELs), from each study. For succinctness, only one effect level (the LOEL or NOEL) was reported for each endpoint (i.e., all exposure levels not reported in plots); as such, the data shown in the plots are conservative, as higher doses not associated with effects are not shown. Full sets of results are provided via AHRQ (See Section 2.2 for AHRQ links to individual outcomes) and were considered by evidence analysts in making weight of the evidence conclusions. In some cases, levels were reported as a range (e.g., effects observed or not observed in a given quartile of exposure; the range of the quartile would be depicted in the plot rather than an upper or lower end/midpoint, etc.). The majority of the information collected in this SR is displayed in the plots; exceptions include studies reporting on unique endpoints that were not reasonably grouped with others, and, for brevity, cardiovascular endpoints associated with lower importance and/or unknown clinical relevance (as defined by the subject matter expert, J.G.). All data extracted and evaluated, however, are publicly available via the AHRQ repository.

Using a weight of evidence approach similar to that of Nawrot et al. (2003), confidence in the body of evidence was determined (Table 1) and conclusions were drawn for each of the endpoints, outcomes, and populations under investigation when sufficient data were available (Table 2). As noted in the methods (Section 2), the objective in the weight of evidence assessment was not to find the most protective or lowest level associated with an effect per se, but rather to make a determination based on the body of evidence as a whole, which included considerations for positive and negative findings, quality of data, level of adversity, consistency, and magnitude of effect relative to conclusions regarding caffeine safety as determined by Health Canada. Weight of the evidence determinations were made by endpoint and population; specifically, conclusions were developed by categorizing evidence relative to the comparator (an intake value not associated with adverse effects) as follows: comparator is acceptable (i.e., evidence supports the Nawrot et al. (2003) conclusions regarding intake), comparator is too high (i.e., evidences suggests potential for effects below the comparator), or comparator is too low (i.e., evidences suggests a lack of effects above the comparator).

3.1. Bone and calcium

Of the full text papers we reviewed (Fig. 2), 14 studies were included and 26 were excluded. With respect to the PECO, all of the included studies involved adult populations (women and men; often elderly adult populations), although one study also evaluated adolescents. Many of the observational studies were conducted in participants from large cohorts, including, for example, the Nurses' Health Study (Fung et al., 2014). Exposures were typically characterized using self-reported methods (e.g., food frequency questionnaires) and were based on consumption of coffee, soda, tea, and chocolate. Approximately half of the studies evaluated the amount

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

Table 1
Summary of confidence in the body of evidence used to develop conclusions by endpoint and outcome. Initial confidence ratings based on study type and study features (OHAT, 2015a,b). Considerations for risk of bias, indirectness, magnitude, dose response, confounding, and consistency (IOM, 2011) relative to evaluation of the PECO (effects relative to specific intakes [the comparators] rather than if a potential relationship existed or not) were used to up- or down-grade (as indicated by the arrows) the level of confidence in the body of evidence supporting the conclusion.

Endpoint	No. of Studies	Initial Confidence Rating	Final Confidence Rating									
			Risk of Bias					Indirectness	Large Magnitude	Residual Confounding	Consistency	
			Definitely Low (++)	Probably Low (+)	Probably High (−)	Definitely High (−)	Overall					
Factor description		*Based on study type and study features (OHAT, 2015a,b)*	*Domain-based evaluation of risk of bias per the OHAT RoB tool (OHAT, 2015a,b)*				*Was there an overall low risk of bias?*	*Was the study designed to evaluate the PECO?*	*Strength of effect (when effect observed below the comparator)*	*Were plausible confounders that would change the observed effect accounted for?*	*Were findings consistent in demonstrating effects or lack of effects at or below the comparator?*	*What is the overall rating when factors that increase or decrease confidence were considered?*
Bone and calcium	14	Moderate										**Moderate**
Risk of fracture and fall	6	Moderate	–	3	3	–	–	↓/-	↓	↑/-	↑/-	Moderate
Bone mineral density and osteoporosis	7	Moderate	–	6	1	–	↑/-	↓	–	–	↓/—	Moderate to low
Cardiovascular												**Moderate to high**
Adults												
Mortality	13	Moderate	2	11	–	–	↑/-	↑	–	↑	↑/-	Moderate
Morbidity	18	Moderate	–	14	–	–	↑/-	↑	–	↑	↓	Moderate
Blood pressure	115	Moderate to high	53	59	3	–	↑/-	↑	↓/-	↑	–	Moderate
Heart rate	53	Moderate to high	28	23	2	–	↑/-	↑		–		
Cholesterol	24	Moderate to high	3	20	1	–	↑/-	↑/-	↓	–	↑	Moderate to high
Heart rate variability	13	Moderate to high	7	6	–	–	↑/-	↑	–	–	↑/-	Moderate to high
Adolescents and children												
Blood pressure	10	Moderate to high	5	3	2	–	↑/-	↑	↓	–	↑	High
Heart rate	6	High	5	3	–	–	↑	↑	–	–	↑/-	High
Behavior												**Moderate to high**
Adults												
Anxiety	40	High	9	31	–	–	↑	↑	↓	–	↓/-	Moderate to high
Anger/ confusion	12	High	3	9	–	–	–	↑	↓	–	↑	High
Depression	18	Moderate to high	3	15	–	–	↑	–	↓	–	↑	Moderate to high
Headache	15	High	4	10	1	–	↑	↑	↓	↓	↑	Moderate to high
Sleep, subjective	21	High	5	16	–	–	↑	↑	–	–	↑	High
Sleep, objective	18	High	6	11	1	–	↑	↑	–	–	↑	High
Problematic and risk-taking behavior	2	Moderate to high	–	1	1	–	–	↓	–	–	–	Moderate
Adolescents and children												
Mood	2	High	2	–	–	–	–	↑	–	–	–	High
Headache	3	Moderate to high	–	1	2	–	↓	↓	–	↓	–	Low to moderate
Sleep	3	Moderate	1	1	1	–	↓	↓	↓	↓	↑	Low

Reproductive and developmental													**Moderate**
Fertility, fecundability, and male reproductive measures	7	Moderate	0	6	1	0	↑	↑	–	–	↑		Moderate to high
Spontaneous abortion	8	Moderate	0	7	1	0	↑	↑	↓/-	↓	↓		Moderate
Recurrent miscarriage	4	Moderate	0	3	1	0	–	↑	–	–	↑/-		Moderate to high
Stillbirth	4	Moderate	0	4	0	0	↑	↑	–	–	–		Moderate
Preterm birth and gestational age	5	Moderate	0	4	0	0	↑	↑	–	–	↑		Moderate to high
Fetal growth	14	Moderate	1	10	2	0	↑	↑	↓	↓	–		Moderate to high
Birth defects	11	Moderate	0	11	0	0	↑	↑	–	–	↑		Moderate
Acute													**Very low to low**
Lethality	14	Very low to low	0	7	6	1	↓	↑	NE	–	↑/-		Very low to low
Other acute effects	18	Very low to low	0	6	12	0	↓	↑	NE	–	↑/-		Very low to low

Up arrows indicate an increase in confidence; down arrows indicate a decrease in confidence; and dashes indicate no change. NE, not evaluated.

of caffeine in these substances as part of the analysis, and the other half did not (thus, the amount of caffeine was calculated by the SR authors per the methods in Section 2). Most of the studies were observational (including cohort and cross-sectional study types), although experimental studies (randomized controlled trials [RCTs]) were also included. Common variables accounted for in such analyses included age, weight, body mass index (BMI), calcium intake, other nutrient intake, alcohol consumption, smoking habits, and physical activity level. With respect to the comparator of <400 mg/day from Nawrot et al. (2003) (which is equivalent to 4.2 eight-ounce cups of coffee based on our standardization assumptions), the majority of the data points were below this level (Fig. 3). Further, most data points were associated with evaluations that used categorical exposure groupings (e.g., <1 cup/day, 1–3 cups/day, and >3 cups/day). The studies that directly evaluated caffeine (i.e., low level of indirectness) were given more weight in the body of evidence assessment relative to those that evaluated caffeine via consumption of coffee or other substances (e.g., as cups/day) as a determinant in a regression model.

Similar to findings reported by Nawrot et al. (2003), endpoints characterizing the bone and calcium outcome included metabolic impacts on calcium homeostasis (n = 2), bone mineral density (BMD) and osteoporosis (n = 9), and risk of fracture (n = 6). Effects of caffeine on bone are most often associated with increased urinary calcium excretion. However, urinary calcium excretion is affected by calcium intake, so calcium intake needs to be considered in the analysis. Altered calcium balance through perturbing calcium excretion can influence bone mass. The majority of the studies reviewed evaluated associations between caffeine consumption and BMD or bone mineral content (BMC); in some studies, these data were also used to characterize osteopenia and osteoporosis. Results varied by bone site.

3.1.1. Evaluation of individual studies by endpoint

3.1.1.1. Risk of fracture and fall. With respect to fracture and fall, most studies reported a lack of effects, both above and below the comparator of 400 mg/day (Fig. 3) (Albrand et al., 2003; Fung et al., 2014; Hallstrom et al., 2013; Jha et al., 2010; Lee et al., 2014). Hallstrom et al. (2013) reported that consumption of ≥560 mg caffeine (≥8 cups) was not associated with a higher rate of any fracture or of hip fracture in a comprehensive evaluation of long-term coffee consumption in relation to fracture risk and BMD in women. In a recent SR and meta-analysis for coffee consumption and risk of fractures (Lee et al., 2014), an insignificant relative risk (RR) for coffee consumption and risk of fracture (RR, 1.03; 95% confidence interval [CI], 0.91–1.16; I_2, 61.4%; $P = 0.001$) was reported for all studies combined. Results of subgroup analyses indicated contrasting findings by sex; men consuming 760 mg/day had a 24% lower risk of fractures, whereas women consuming 190 mg/day had a 2% higher risk (1.02; 95% CI, 1.01–1.04) of fractures, relative to those who did not drink coffee. Estimates increased based on increased consumption; 8 cups of coffee per day was reported to be associated with a 54% higher risk of fractures (RR, 1.54; 95% CI, 1.19–1.99). This study did not evaluate interactions between caffeine/coffee consumption and calcium intake. Hallstrom et al. (2006) also reported effects below the comparator. The authors reported that a daily intake of ≥330 mg caffeine may be associated with a modestly increased risk of osteoporotic fractures (RR, 1.20; CI, 1.07–1.35), especially in women with a low intake of calcium; when stratified by calcium intake, the increased risk was only significant when calcium intake was low (<700 mg/day). No trend in increased risk was observed with higher caffeine intake in participants with high calcium intake.

The majority, although not all, of the data on risk of fracture or fall demonstrate a lack of effects of caffeine consumption at levels

Table 2
Summary of weight of evidence and conclusions for the PECO question: For a specified population Is Caffeine intake above the (Comparator Dose) compared to other doses associated with adverse effects of specified outcomes/ endpoint.

Outcome/ Endpoint	Comparator	Population (all healthy)	Number of points for comparison	Range of NOELs *(LOELs)* mg/day*	Summary of Spectrum of Data and Endpoint Conclusion			Level of Adversity		Level of Confidence	Outcome Conclusion	
					Comparator is too high	Comparator is acceptable	Comparator is too low	Type of effect	Importance of effect in decision making		Narrative Conclusion	Comparator Conclusion
Parameter definition	*Intake amount data were compared to*	*Data evaluted for each of four populations*	*# of data points from studies compared to PECO (some studies had multiple points for comparison)*	No observed effect levels *(and low observed effect levels identified in literature)**	*"X" if there were any data reporting effects at intakes below (or equal to) the comparator*	*"X" if there were any data reporting a lack of effects below or equal to the comparator*	*"X" if there were any data that showed lack of effects above the comparator*	*Categorization of effect type: clinical or physiological (and reversible)*	*Categorization of endpoint importance (Guyatt et al., 2011)*	*Overall level of confidence in the body of evidence considering RoB, magnitude, consistency, etc. (Table 1)*	*Similar to the approach and conclusions of Health Canada (Nawrot et al. (2003), conclusions were developed using a weight of evidence assessment conducted by evidence analysts and subject matter experts. Determinations were based on considerations for direction of findings, quality of data, level of adversity, and confidence in the body of evidence (including consistency, dose response, and magnitude of effect; Table 1). When developing outcome conclusions, clinical endpoints with a high level of adversity were given more weight than physiological endpoints .*	*Succinct conclusion re: comparator (for healthy populations)*
					Shaded box depicts WOE conclusion about the comparator							
Bone and Calcium												
Risk of fracture and fall	400 mg/day	Adults	6	17-760 *(190-330)*	X	X	X	Clinical	High	Moderate	The SR of 14 studies provided evidence to evaluate potential impacts of the consumption of 400 mg caffeine/day on the bone and calcium outcome in healthy adults; these studies included assessment of the risk of fracture and fall, BMD and osteoporosis, and altered calcium homeostasis. When the weight of evidence was considered (See Section 3.1.2), the comparator, 400 mg/day, was found to be an acceptable intake that is not associated with significant concern regarding overt, adverse effects on bone or calcium endpoints, particularly under conditions of adequate calcium intake. Although effects were observed at exposures below the comparator they were often limited to physiological effects following acute exposure (altered calcium homeostasis), and subgroups in analyses of clinical endpoints, including those with low calcium intake. Such effects were generally of low magnitude, and/or were of overall low/negligible consequence to downstream events. Several studies also reported on a lack of effect on the clinical endpoints following chronic consumption below the comparator, as well as above the comparator. Based on the underlying study type (11 observational, 2 RCTs, 1 meta-analysis) that constitute this evidence base, there is a moderate level of confidence in the research, which supports this conclusion. Key limitations that precluded a higher level of confidence were the inability to fully accommodate for calcium intake, the high level of indirectness, as well as an uncertainty in exposure estimates.	400 mg/day acceptable for adults *No conclusion for pregnant women, adolescents, or children*
	300 mg/day	Pregnant Women	0	N/A	No data					N/A		
	2.5 mg/kg-day	Adolescents	0	N/A	No data					N/A		
	2.5 mg/kg-day	Children	0	N/A	No data					N/A		
Bone mineral density and osteoporosis	400 mg/day	Adults	7	107.5-300 *(237.5-250.7)*	X		X	Clinical	Medium/High	Moderate		
	300 mg/day	Pregnant Women	0	N/A	No data					N/A		
	2.5 mg/kg-day	Adolescents	0	N/A	No data					N/A		
	2.5 mg/kg-day	Children	0	N/A	No data					N/A		
Calcium Homeostasis	400 mg/day	Adults	2	None *(60-285)*	Insufficent data to develop conclusion			Physiological	Low	High		
	300 mg/day	Pregnant Women	0	N/A	No data					N/A		
	2.5 mg/kg-day	Adolescents	0	N/A	No data					N/A		
	2.5 mg/kg-day	Children	0	N/A	No data					N/A		
Cardiovascular												
Mortality	400 mg/day	Adults	13	95-822 *(0.4-459)*	X	X	X	Clinical	High	Moderate	In summary, the SR of 191 studies evaluated the potential for caffeine to be associated with adverse cardiovascular effects, including mortality, morbidity, blood pressure, heart rate, cholesterol, and heart rate variability (and others; see Supplementary File S2). When the weight of evidence was considered (See Section 3.2.2), with particular emphasis on level of adversity, 400 mg caffeine/day was found to be an acceptable intake which is not associated with significant concern regarding adverse cardiovascular effects in healthy adults. For clinical endpoints, some findings suggested that the comparator was too conservative; however, other data, particularly those for physiological endpoints, reported effects below the comparator. For such physiological endpoints (e.g., blood pressure), confidence in determining conclusions relative to the comparator was limited by the inability to ascertain the conditions and magnitude of change that would be considered adverse in a clinical or toxicological context. There is a moderate to high level of confidence (Table 1) in this evidence base.	400 mg/day acceptable for adults *No conclusion for pregnant women, adolescents, or children*
	300 mg/day	Pregnant Women	0	N/A	No data					N/A		
	2.5 mg/kg-day	Adolescents	0	N/A	No data					N/A		
	2.5 mg/kg-day	Children	0	N/A	No data					N/A		
Morbidity	400 mg/day	Adults	18	95-1045 *(0.0- ≥400)*	X	X	X	Clinical	High	Moderate		
	300 mg/day	Pregnant Women	0	N/A	No data					N/A		
	2.5 mg/kg-day	Adolescents	0	N/A	No data					N/A		
	2.5 mg/kg-day	Children	0	N/A	No data					N/A		
Blood Pressure	400 mg/day	Adults	116	50-997 *(0.0-600)*	X		X	Physiological	Low/Medium	Moderate		
	300 mg/day	Pregnant Women	0	N/A	No data					N/A		
	2.5 mg/kg-day	Adolescents	0	N/A	No data					N/A		
	2.5 mg/kg-day	Children	10	0.35-5.0 *(1.0-5.0) mg/kg-day*	X		X			High		
Heart Rate	400 mg/day	Adults	53	80-780 *(39.3-600)*	X	X	X	Physiological	Low/Medium	Moderate to high		
	300 mg/day	Pregnant Women	0	N/A	No data					N/A		
	2.5 mg/kg-day	Adolescents	0	N/A	No data					N/A		
	2.5 mg/kg-day	Children	6	1.0-6.0 *(1.0-5.0)* mg/kg-day	Insufficent data to develop conclusion					High		

Cholesterol	400 mg/day	Adults	24	95-475.2 *(0.0-392)*	X	X	X	Physiological	Low/Medium	Moderate to high	
	300 mg/day	Pregnant Women	0	N/A	No data					N/A	
	2.5 mg/kg-day	Adolescents	0	N/A	No data					N/A	
	2.5 mg/kg-day	Children	0	N/A	No data					N/A	
Heart Rate Variability	400 mg/day	Adults	13	39.3-346.6 *(100-500)*	X	X		Physiological	High	Moderate to high	
	300 mg/day	Pregnant Women	0	N/A	No data					N/A	
	2.5 mg/kg-day	Adolescents	0	N/A	No data					N/A	
	2.5 mg/kg-day	Children	0	N/A	No data					N/A	
Behavior											
Anxiety	400 mg/day	Adults	49	70-480 *(80.6-1200)*	X		X	Clinical	Medium/High	Moderate to high	The SR of 81 studies provided evidence to evaluate potential impacts of the consumption of 400 mg caffeine/day on the behavior outcome, including assessment of mood (comprising anxiety and other mood states), headache, sleep, withdrawal, and risk-taking behavior. When the weight of evidence was considered (See Section 3.3.2), the comparator, 400 mg caffeine/day, was found to be an acceptable intake that is not be associated with significant concern for adverse behavioral effects in adults. However, intake below the comparator may affect some sensitive individuals who are prone to anxiety or sleep disruption. Often, observed effects below the comparator (e.g., anxiety) were limited to subgroups or timing of dose (e.g., sleep), whereas others were complicated by consumer status (e.g., headache and fatigue). For some endpoints (depression, headache, sleep [subjective], and anger/confusion there was largely a lack of effects reported, and in some cases, data suggested that intakes higher than the comparator were without effect. There is a moderate to high level of confidence in the body of evidence supporting this conclusion. Confidence was increased by the overall low risk of bias and low level of indirectness; although the variability introduced by sensitive subpopulations and consumer status were key limitations that precluded a higher level of confidence.	400 mg/day acceptable for adults *No conclusion for pregnant women, adolescents, or children*
	300 mg/day	Pregnant Women	0	N/A	No data					N/A		
	2.5 mg/kg-day	Adolescents	0	N/A	No data					N/A		
	2.5 mg/kg-day	Children	1	2 mg/kg-day *(None)*	Insufficient data to develop conclusion					High		
Anger/Confusion	400 mg/day	Adults	24	70-1200 *(80-1200)*	X	X	X	Clinical	Medium	High		
	300 mg/day	Pregnant Women	0	N/A	No data					N/A		
	2.5 mg/kg-day	Adolescents	0	N/A	Insufficient data					N/A		
	2.5 mg/kg-day	Children	2	2.0-2.32 mg/kg-day *(None)*	Insufficient data					Moderate		
Depression	400 mg/day	Adults	21	80-1200 *(≥100)*	X	X	X	Clinical	Medium/High	Moderate to high		
	300 mg/day	Pregnant Women	0	N/A	No data					N/A		
	2.5 mg/kg-day	Adolescents	0	N/A	Insufficient data					N/A		
	2.5 mg/kg-day	Children	1	2 mg/kg-day *(None)*	Insufficient data					High		
Headache	400 mg/day	Adults	15	70-400 *(100-400)*	X	X		Physiological/ Clinical	Medium/High	Moderate to high		
	300 mg/day	Pregnant Women	0	N/A	No data					N/A		
	2.5 mg/kg-day	Adolescents	0	N/A	Insufficent data to develop conclusion					N/A		
	2.5 mg/kg-day	Children	4	0.0-2.32 *(0.0-0.6)* mg/kg-day	Insufficent data to develop conclusion					Low to moderate		
Sleep - Subjective	400 mg/day	Adults	25	70-444 *(80-1200)*	X	X	X	Clinical	Medium	High		
	300 mg/day	Pregnant Women	0	N/A	No data					N/A		
	2.5 mg/kg-day	Adolescents	0	N/A	No data					N/A		
	2.5 mg/kg-day	Children	3	0.0-2.77 *(0.0-0.6)* mg/kg-day	Insufficent data to develop conclusion					Moderate to high		
Sleep - Objective	400 mg/day	Adults	18	60-595 *(150-1200)*	X	X	x	Physiological/ Clinical	Medium	High		
	300 mg/day	Pregnant Women	0	N/A	No data					N/A		
	2.5 mg/kg-day	Adolescents	0	N/A	No data					N/A		
	2.5 mg/kg-day	Children	1	2.77 mg/kg-day *(None)*	Insufficent data to develop conclusion					Low to moderate		
Problematic and Risk-taking behavior	400 mg/day	Adults	2	None *(80- ≥570)*	Insufficent data to develop conclusion			Clinical	Medium	Moderate		
	300 mg/day	Pregnant Women	0	N/A	No data					N/A		
	2.5 mg/kg-day	Adolescents	0	N/A	No data					N/A		
	2.5 mg/kg-day	Children	0	N/A	No data					N/A		

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

Reproductive and Developmental														
Fecundibility, Fertility and Male Reproductive	400 mg/day	Adults	6	240- ≥800 *(≥308)*	X	X	X	Clinical	Medium/High	Moderate to High	In summary, the SR of 58 studies (55 observational; 3 RCTs) provided evidence to evaluate potential impacts of the consumption of 300 mg caffeine/day in pregnant women, or 400 mg/day in healthy adults, on reproductive and developmental effects. These studies included assessment of fecundability, fertility, male reproductive endpoints, spontaneous abortion, recurrent miscarriage, stillbirth, preterm birth and gestational age, fetal growth, birth defects, childhood cancers, and childhood behavior. When the weight of evidence was considered (See Section 3.4.2), an intake of 300 mg caffeine/day in pregnant women and 400 mg/day in healthy adults are generally without significant concern regarding overt adverse effects on fecundability, fertility, male reproductive endpoints, spontaneous abortion, recurrent miscarriage, stillbirth, preterm birth and gestational age, birth defects, and childhood behavior. For some endpoints, including fetal growth, childhood cancers, isolated congenital malformations, evidence indicated that intake at the level of the comparator was too liberal; however, refined conclusions were difficult to determine given uncertainties in the evidence base, including low attrition rates, lack of ability to account for recall bias particularly for childhood cancers, lack of ability to fully accommodate the impact of the pregnancy signal, and lack of biological relevance of findings (referring primarily to fetal growth changes; the biological significance of such an effect is more robustly evaluated in studies evaluating SGA or IUGR, which, as a whole, did not show effects below the comparator). The relevance of smoking to the analyses in this data set was generally well recognized and, in most studies, the investigators attempted to control for such. There is a moderate level of	400 mg/day acceptable for adults
Spontaneous abortion	300 mg/day	Pregnant Women	14	200- ≥500 *(100-500)*	X	X	X	Clinical	High	Moderate		
Recurrent miscarriage	300 mg/day	Pregnant Women	4	100-300 *(151-301)*	X	X		Clinical	High	Moderate to High		
Stillbirth	300 mg/day	Pregnant Women	5	≥380 *(300-700)*	X	X	X	Clinical	High	Moderate		
Preterm birth and gestational age	300 mg/day	Pregnant Women	5	300-540 *(None)*		X	X	Clinical	Medium/High	Moderate to High		300 mg/day acceptable for pregnant women
Fetal growth	300 mg/day	Pregnant Women	17	300- ≥500 *(25-775)*	X	X	X	Clinical	Medium/High	Moderate to High		
Birth defects	300 mg/day	Pregnant Women	22	190- ≥500 *(10-405)*	X	X	X	Clinical	High	Moderate		
Childhood cancers	300 mg/day	Pregnant Women	4	≥285 *(95- ≥332)*	Insufficient data to develop conclusion			Clinical	High	N/A		
Childhood behavior	300 mg/day	Pregnant Women	3	300- ≥1000 *(None)*	Insufficient data to develop conclusion			Clinical	High	N/A		
Acute Toxicity												
Lethality	10 g	Adults	11	0.56-50 g *(0.24-20 g)*	X	X	X	Clinical	High	Very low to low	The SR of 26 studies provided evidence to evaluate potential acute toxicity. Following a weight of evidence review (see Section 3.5.2), the comparators of 10 g for lethality and 400 mg/day or 2.5 mg/kg/day for other acute effects, were determined to be acceptable for healthy adults and adolescents, respectively. There is very low to low confidence in this evidence base, due primarily to uncertainty in the estimates of exposure and to the high risk of bias. A higher level of confidence is also precluded by the inherent bias introduced by utilizing case reports (i.e., limitations to reporting of effects versus no effects).	Lethality: 10g for adults and adolescents Nonlethality: 400 mg/day for adults and 2.5 mg/kg-day for adolescents *No conclusion for pregnant women or children*
	Undefined	Pregnant Women	0	N/A		No data				N/A		
	Undefined	Adolescents	3	10-12 g *(51.6 g)*		X	X			Very low to low		
	Undefined	Children	0	N/A		No data				N/A		
Other acute events (non-fatal)	400 mg/day	Adults	11	None *(0.167-50 g)*	X	X		Clinical	High	Very low to low		
	Undefined	Pregnant Women	2	None *(0.288-0.9 g)*		Insufficient data				N/A		
	Undefined	Adolescents	5	None *(0.48-12 g)*		X				Very low to low		
	Undefined	Children	0	N/A		No data				N/A		

*Values based on that reported by individual study authors; in many cases, data are based on categorical classifications which represent ranges (e.g., less than two cups of coffee would be represented by 0 to 190 mg caffeine).

both above (up to 760 mg/day) and below 400 mg/day. Evidence of effects below 400 mg/day was of low magnitude (RR, ≤1.20) and was confounded by calcium intake; the potential interaction of calcium intake was not accounted for in the study reporting the lowest effect level (Lee et al., 2014), and the other study (Hallstrom et al., 2006) reporting an effect level below the comparator found in stratified analyses that increased risk was only observed under conditions of low calcium intake. Confidence in these data is moderate (OHAT, 2015a) (Fig. 4; Table 1); findings were generally consistent, and most, although not all, studies controlled for calcium intake. As such, the evidence in this SR supports that an intake of 400 mg caffeine/day in healthy adult populations, particularly those with adequate calcium intake, is not associated with significant concern regarding the risk of fracture and fall.

3.1.1.2. Bone mineral density and osteoporosis. Of the seven studies that evaluated BMD, only one reported on levels of caffeine intake above the comparator. Barbour et al. (2010) reported that higher caffeine intake of 520.7 mg/day was associated with lower cortical and trabecular volumetric BMD in men aged ≥69 years. The remaining studies evaluated consumption levels lower than the comparator, the majority of which found a lack of effects at exposures ranging from 108 to 300 mg/day (El Maghraoui et al., 2010; Hallstrom et al., 2010, 2013; Harter et al., 2013; Rapuri et al., 2001; Wetmore et al., 2008) (Fig. 3).

Four of these studies, however, also reported effects below the comparator (El Maghraoui et al., 2010; Hallstrom et al., 2010; 2013; Rapuri et al., 2001) for some of the endpoints evaluated. Among Moroccan men who consumed >285 mg caffeine/day, El Maghraoui et al. (2010) reported a decreased association of high coffee consumption with osteoporosis at any site (0.82; CI, 0.74–0.91, $P < 0.05$), although there was an increased association with the lumbar spine (1.76; CI, 1.08–2.85; $P < 0.05$) and no association with total hip, the most important site. Study subjects with osteopenia and osteoporosis also reported low calcium intake (62% and 75%, respectively). Rapuri et al. (2001) reported that the rate of bone loss at the spine was higher in a group of high-caffeine consumers compared with low-caffeine consumers (>300 mg caffeine/day versus <300 mg caffeine/day, respectively, with percent change in BMD of −1.90 ± 0.97 versus 1.19 ± 1.08, respectively, at baseline). However, the rate of bone loss at other sites (femoral neck, trochanter, total body, and total femur) was not significantly altered in a longitudinal assessment of data collected from elderly women aged 66–77 years over a 3-year period (Rapuri et al., 2001). These authors also conducted a cross-sectional assessment of data, reporting no significant differences in BMD, no changes in a number of calciotropic hormones, and changes in bone markers in women who consumed >300 mg caffeine/day relative to those who consumed <300 mg/day. Similarly, Hallstrom et al. (2010) reported that consumption of ≥237.5 mg/day was associated with a 4% lower BMD of the proximal femur compared to low or nonconsumers of coffee in a large population of Swedish men aged 72 years. This finding, however, was not observed in women, nor was it modified by calcium intake. Finally, Hallstrom et al. (2013) reported that coffee intake of ≥280 mg was associated with a 2%–4% lower bone density, which did not translate into an increased risk of fracture (discussed below). These authors also reported a lack of association between consumption of ≥280 mg caffeine with an increased incidence of osteoporosis or an incidence of one or two falls in the previous year.

Collectively, the majority of studies reviewed support that the comparator of 400 mg/day in healthy adults is not harmful with respect to BMD and osteoporosis, although more evidence is needed for effects of caffeine intake between 200 and 400 mg/day given the number of studies that reported effects associated with

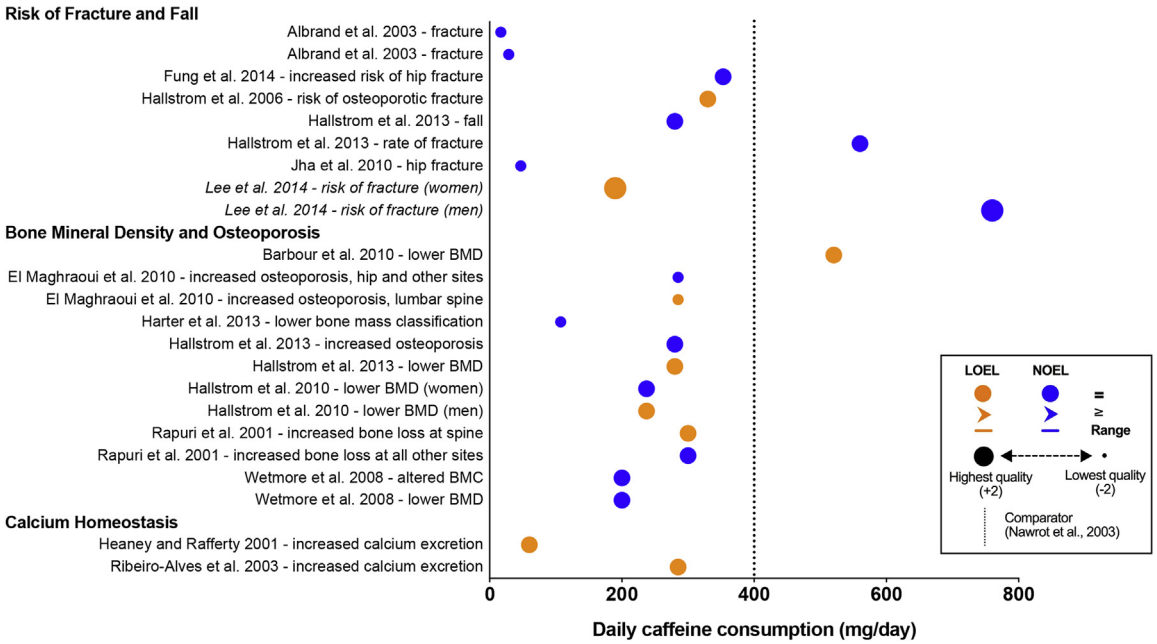


Fig. 3. Summary diagram of exposure-response data relative to the comparator for the bone and calcium outcome (all endpoints). Symbols represent caffeine intake (mg/day) as reported by original study authors. The color of the symbol indicates the type of effect; no effect (NOEL; blue symbols) or the lowest effect level (LOEL; orange symbols). The shape of the symbol represents the type of metric (circles represent a discrete value, arrowheads represent greater than or equal to a value, and a horizontal line represents a range of values; metrics are based on that reported by the original study authors). The size of the symbol indicates the overall risk of bias (larger symbols indicate a lower risk of bias, or higher methodological quality). The dashed vertical line marks the comparator value. Italicized study names indicate a meta-analysis. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

consumption in this range. In the studies reporting effects (both below and above the comparator), such effects were typically associated with subgroup analyses (e.g., limited to females), were associated with single sites (i.e., impacts not observed at all sites evaluated), or were not associated with downstream events (i.e., risk of fracture). In addition, although calcium consumption was integrated into most of the analyses, the method for doing so varied, which thus contributed to uncertainty in findings because some of the studies involved participants with low calcium intake. These factors, along with the use of different consumption groupings by study authors, the uncertainty associated with assessing caffeine exposure (particularly relative to calcium consumption), and the lack of consistently observed effects (above or below the comparator) make it difficult to further refine the conclusion for BMD and osteoporosis. The underlying evidence base is associated with a moderate to low level of confidence (Fig. 4; Table 1).

3.1.1.3. Calcium homeostasis. Two RCTs were reviewed. Heaney and Rafferty (2001) reported that consumption of caffeinated beverages (60 or 92 mg caffeine) produced small increases in calcium excretion, which can be offset by small increases in calcium intake (15–30 mL [1–2 tablespoons] milk; Rafferty and Heaney, 2008). The authors reported that the overall magnitude was sufficiently small such that the observed changes were not meaningful to the calcium economy (Heaney and Rafferty, 2001). Ribeiro-Alves et al. (2003) reported that exposure to 285 mg caffeine resulted in increased excretion of calcium in women (described as ≥1-fold; 0.5 ± 0.5 mmol calcium/mmol creatinine following caffeine exposure, and 0.2 ± 0.1 mmol calcium/mmol creatinine without caffeine); this finding was based on a study population of women who habitually consume a low-calcium diet.

As only two studies were available, a conclusion was not developed; however, data from these two studies suggests that the comparator of 400 mg/day may be too high for physiological impacts on calcium homeostasis; however, when calcium intake is considered in concert and, in particular, the observation that the physiological changes in homeostasis can be offset by small amounts of calcium, the evidence is more supportive that the comparator is acceptable. Furthermore, the amount of change observed was well within typical amounts of calcium excreted, including in those consuming low-calcium diets (Wu, 2006), which thus supports an unlikely impact on calcium economy at the exposure levels evaluated.

3.1.2. Body of evidence assessment

The individual studies were generally associated with low risk of bias ratings, with only three studies at the lower end of the spectrum toward high risk of bias (Figs. 3 and 4; Table 1). The study ratings were most impacted by the confidence in exposure assessment. Few studies involved direct evaluation of caffeine; rather, studies relied on self-reported estimates of consumption of caffeine-containing beverages. The range in the level of indirectness of caffeine intake spans from a low level of indirectness (i.e., direct evaluation of PECO) to a very serious level of indirectness. Thus, studies that directly exposed subjects to a known amount of caffeine or assessed caffeine using validated measures were given more weight when considering the body of evidence for this outcome (Fig. 4); endpoints with higher levels of adversity were also given more weight (Table 2).

The Nawrot et al. (2003) conclusion for the bone and calcium endpoint included both reference to an intake of caffeine (<400 mg/day) as well as calcium intake (800 mg/day). Because of the lack of consistent reporting of calcium intake (i.e., lack of author-reported data or conclusions that directly linked levels of caffeine and levels of calcium), this SR could not make a conclusion about the effect of calcium intake on the relation between caffeine intake and calcium balance or bone outcomes. For example, Hallstrom et al. (2006) reported that an increased risk of fracture

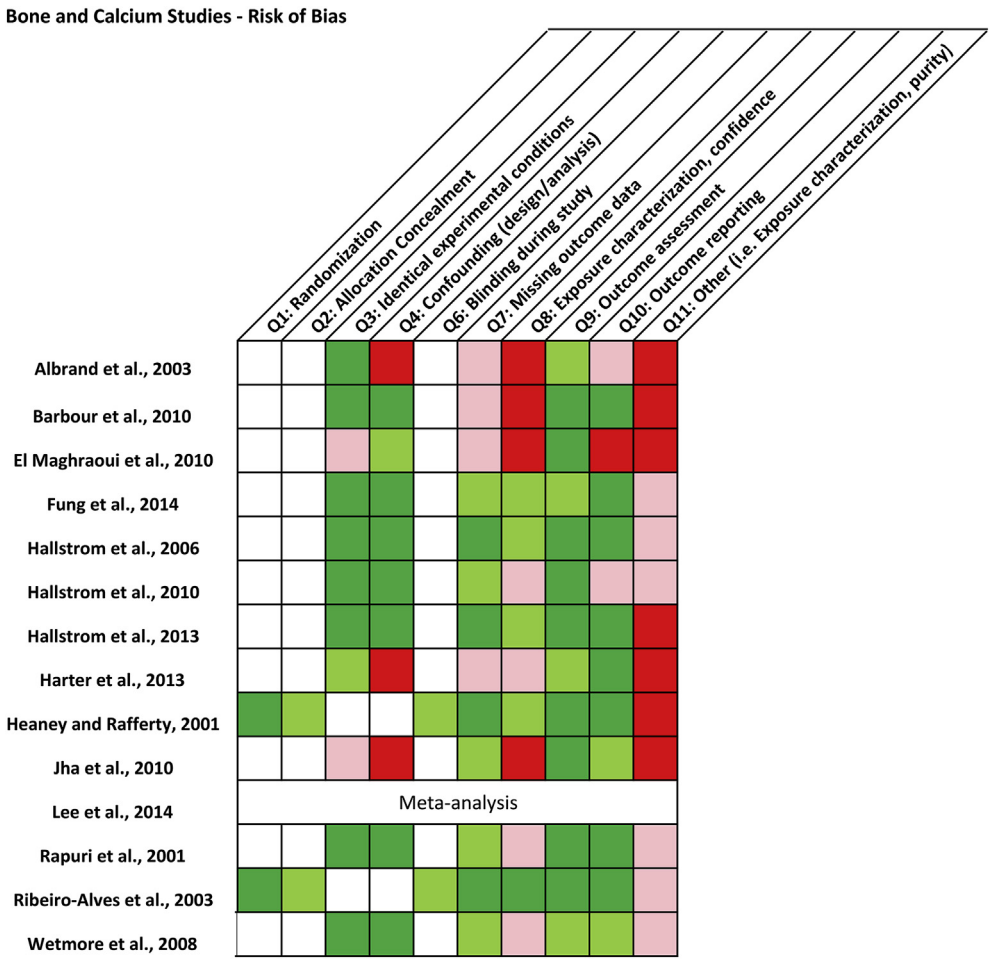


Fig. 4. Risk of bias (RoB) heat map for studies included in the bone and calcium outcome. The domain-based validity was evaluated based on study type per the OHAT (2015b) RoB tool. RoB for each domain is indicated by color: "definitely low risk of bias" (dark green, +2), "probably low risk of bias" (light green, +1), "probably high risk of bias" (light red, −1), and "definitely high risk of bias" (dark red, +2). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

was observed when calcium intake was low (<700 mg/day), but most studies did not do similar stratified analyses.

In studies showing potentially adverse effects of caffeine intake at <400 mg/day, the effect size was generally of low magnitude, effects were only observed in some bone sites, or effects were observed in subsets of the population, such as women who habitually consume a low-calcium diet (El Maghraoui et al., 2010; Hallstrom et al., 2006, 2010; Lee et al., 2014; Rapuri et al., 2001; Ribeiro-Alves et al., 2003). Two controlled trials reported that single exposures of caffeine impacted subsequent measures on urinary calcium excretion (Hallstrom et al., 2013; Heaney and Rafferty, 2001); these changes were not considered to meaningfully impact the calcium economy. It is also notable that when the progression of effects was considered, in some cases, authors indicated that early events were not linked to more critical effects. For example, Heaney and Rafferty (2001) indicated that the portion of the observed excess calciuria that may be due to caffeine can probably be dismissed as being of no consequence to calcium economy, or the 2%–4% lower bone density observed by Hallstrom et al. (2013) following consumption of ≥280 mg caffeine did not translate into an increased risk of fracture.

In summary, the SR of 14 studies provided evidence to evaluate potential impacts of the consumption of 400 mg caffeine/day on the bone and calcium outcome in healthy adults; these studies included assessment of the risk of fracture and fall, BMD and osteoporosis, and altered calcium homeostasis. When the weight of evidence was considered, the comparator, 400 mg/day, was found to be an acceptable intake that is not associated with significant concern regarding overt, adverse effects on bone or calcium endpoints, particularly under conditions of adequate calcium intake. Although effects were observed at exposures below the comparator they were often limited to physiological effects following acute exposure (altered calcium homeostasis), and subgroups in analyses of clinical endpoints, including those with low calcium intake. Such effects were generally of low magnitude, and/or were of overall low/negligible consequence to downstream events. Several studies also reported on a lack of effect on the clinical endpoints following chronic consumption below the comparator, as well as above the comparator. Based on the underlying study type (11 observational, 2 RCTs, 1 meta-analysis) that constitute this evidence base, there is a moderate level of confidence in the research, which supports this conclusion. Key limitations that precluded a higher level of confidence were the inability to fully accommodate for calcium intake, the high level of indirectness, as well as an uncertainty in exposure estimates.

3.2. Cardiovascular

We reviewed 276 full text papers evaluating potential cardiovascular effects of caffeine. A total of 203 of these studies were considered to meet the criteria for inclusion in the SR because they permitted comparison to the Nawrot et al. (2003) conclusions

(Fig. 2). With respect to the PECO, nearly all of the included studies involved adult populations, for which the Nawrot et al. (2003) comparator of ≤400 mg/day was applied. For the 11 studies involving children (aged 3–12 years), adolescents (aged 12–19 years), or both, the Nawrot et al. (2003) comparator of ≤2.5 mg/kg body weight was applied. In several studies, "adults" included individuals who were aged 18 or 19 years (although results could not be separated as such). Although this is in the upper range for adolescents, these populations were treated as adults for purposes of this evaluation.

The majority of the included studies (172 of 203) were controlled trials, most of which were a randomized, double-blinded, crossover design. Of the remaining 31 studies, 26 were observational studies (cohort, case-control, case-crossover, cross-sectional), 4 were meta-analyses of such studies, and 1 was a meta-analysis of RCTs. In all of the controlled trials but one (Christensen et al., 2001), exposures were characterized based on measured values. The reverse was the case for the observational studies and meta-analyses—that is, all but the meta-analysis of controlled trials were based on self-reported exposures (food frequency questionnaires). The majority of the controlled trials were of essentially pure caffeine administered as a pill/capsule or dissolved in a liquid; however, studies involving exposure to caffeine-containing foods or beverages such as chocolate, coffee, tea, soda, or energy drinks or a medical/dietary supplement were also included. For the latter studies, the study authors typically reported the amount of caffeine in the food/beverage/supplement utilized in the evaluation; for studies in which the author did not, the amount of caffeine was calculated by the SR authors per the methods (Supplemental File S1). Exposure in observational studies was based on estimates of total caffeine or consumption of one or more caffeinated foods or beverages; for the latter, caffeine exposure was calculated by the SR authors.

For most of the controlled trials, participants fasted or abstained from caffeine consumption for some number of hours, generally overnight, prior to caffeine exposure; however, in some studies, participants were asked to abstain from caffeine for ≥1 day or not at all (i.e., a satiated state). Most controlled trials were also of a single "acute" exposure, whereas a few evaluated "chronic" exposures over a few days or weeks. In either case, participants may have been caffeine naïve or they may be nonhabitual caffeine or caffeine-containing beverage consumers, whereas other participants (in the same or a different study) were classified by amount of regular caffeine consumption (e.g., light, moderate, or heavy consumers). Another variation in exposure characterization was studies in which participants were "pretreated" for a certain number of days, followed by administration of a "challenge" in which the pretreatment or challenge may have been caffeine and/or a caffeine beverage versus some type of placebo. Measurements from the controlled trials were most commonly collected 30–60 min following exposure (with some studies also collecting measurements before and after this time interval) to capture effects at expected peak plasma concentrations. Finally, most of the controlled trials evaluated few, if any, potential confounders, whereas the majority of the observational studies included analyses accounting for many common risk factors for cardiovascular disease (CVD) (e.g., age, sex, smoking, alcohol consumption, BMI).

With respect to the comparator of ≤400 mg/day from Nawrot et al. (2003), the majority of the data points for adults, regardless of the direction of findings, are below the comparator intake. For studies of children and/or adolescents, about one-half of the data points are below the comparator of ≤2.5 mg/kg body weight, again regardless of findings. Most studies were designed specifically to evaluate caffeine (typically via direct exposure to caffeine in controlled trials or conversion of self-reported consumption of cups of caffeine-containing beverages per day) and thus were considered to have a low level of indirectness.

The majority of the 172 controlled trials evaluated blood pressure (primarily peripheral systolic and diastolic; Table 1) and heart rate and, in the latter case, often during or after exercise. For the 26 observational studies, most evaluated cardiovascular morbidity (e.g., acute myocardial infarctions, atrial fibrillation) and/or mortality (e.g., coronary heart disease, stroke). Nawrot et al. (2003) also evaluated blood pressure, heart rate, and CVD, as well as arrhythmia and cholesterol. However, in this SR, many additional cardiovascular parameters characterizing this outcome were evaluated, including aortic stiffness/wave reflection, cerebral blood flow, plasma or urinary constituents (e.g., catecholamines, homocysteine), endothelial function, heart rate variability, heart rhythm, hemodynamic measurements other than blood pressure and heart rate (e.g., cardiac output, stroke volume), and ventricular function (to note, some studies addressed additional, unique endpoints). Each of these is discussed in more detail below, and in doing so, considerations for the relative importance of the endpoints, or level of adversity, to the outcome are considered (Guyatt et al., 2011) (Table 2). Specifically, the hierarchy considered for cardiovascular endpoints involved clinical effects (e.g., morbidity and mortality) > important physiological endpoints (e.g., heart rate, blood pressure) > other physiological endpoints (e.g., aortic stiffness and hemodynamic measurements other than blood pressure and heart rate). The data for the other physiological endpoints are discussed in Supplementary File S2.

3.2.1. Summary of individual studies by endpoint

3.2.1.1. Cardiovascular mortality. Nine cohort studies were identified that evaluated the association between consumption of caffeine from multiple sources or specifically in coffee or tea and the risk of cardiovascular mortality (or in two cases, combined mortality and morbidity) (Fig. 5A). Six of these studies found no association or no increased risk for caffeine intakes ranging from ~95 to 855 mg/day (Bertoia and Triche, 2013; Gardener et al., 2013; Greenberg et al., 2008; Happonen et al., 2008; Lopez-Garcia et al., 2008; Paganini-Hill, 2011). Of the remaining three studies, one reported an increased risk, but only for a specific genotype (Happonen et al., 2006). Happonen et al. (2006) reported an increased risk in the incidence of "acute coronary events" (defined by the authors as acute myocardial infarction or coronary death) following consumption of >320 mg caffeine/day for those with a low-activity catechol-*O*-methyl transferase (COMT) genotype (odds ratio [OR], >3). In an earlier study, Happonen et al. (2004) reported a j-shaped dose-response curve, with the lowest number of acute coronary events (as defined above) associated with moderate coffee drinkers (150–320 mg caffeine/day). The RR was significantly higher in heavy coffee drinkers (>320 mg caffeine/day) based on 14 years of follow-up (RR, ~1.5), but when limited to 2.5 or 5 years of follow-up, the RRs were higher in both light coffee drinkers (0.4–150 mg caffeine/day) and heavy coffee drinkers (RR, ~2). Finally, Mineharu et al. (2011) reported hazard ratios (HRs) for stroke and total CVD mortality, but not for coronary heart disease mortality (differentiated based on International Classification of Disease [ICD] codes), that were significantly greater than 1 (in the range of 2 or 3) for women who consumed ≥459 mg caffeine/day in coffee as compared to women who consumed <22 mg caffeine/day in coffee. The HRs for men were not significantly greater than 1. Mineharu et al. (2011) also reported that green tea consumption up to ≥180 mg caffeine/day was associated with a decreased risk of total CVD, coronary heart disease, or stroke mortality as compared to those who consumed <4.3 mg caffeine/day in green tea.

Collectively, the majority of evidence support that 400 mg caffeine/day in healthy adult populations is an acceptable intake

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

which is not associated with significant concern for cardiovascular mortality. Even at higher intakes up to ~855 mg/day, there are no consistently reported effects on mortality; further, several studies, reported findings that are suggestive of protective effects. The studies reporting effects, both above and below the comparator, were conditional (e.g., observed in subset of data evaluated). There is a moderate level of confidence in the body of evidence (Fig. 6; Table 1) supporting these conclusions.

3.2.1.2. Cardiovascular disease morbidity. Four meta-analyses and 11 cohort, case-control, or case-crossover studies were identified that evaluated the association between consumption of caffeine from multiple sources or specifically in coffee and the risk of cardiovascular morbidity (or as noted above, two studies combined mortality and morbidity), primarily acute or non-fatal myocardial infarction, atrial fibrillation, or stroke (Figs. 5 and 6). Premature ventricular complexes (PVCs) were not evaluated in any of the studies in the time frame of this SR, but previous data have been summarized (Pelchovitz and Goldberger, 2011). The strongest data come from four meta-analyses identified (Caldeira et al., 2013; Cheng et al., 2014; Mostofsky et al., 2012; Sofi et al., 2007), which included cohort and/or case-control studies. The three smallest assessments (five or six studies each) reported no increased risk in heart failure events or atrial fibrillation at ≥500, ≥700, or ≥1050 mg caffeine/day (the latter being in coffee) (Caldeira et al., 2013; Cheng et al., 2014; Mostofsky et al., 2012). Sofi et al. (2007) also reported no increased risk coronary heart disease (primarily acute myocardial infarction) at ≥360 mg caffeine/day in coffee based on their analysis of 10 cohort studies; however, for the 13 case-control studies evaluated, there was an increased risk of coronary heart disease at 275–360 or >360 mg caffeine/day in coffee (ORs, <2).

For the observational studies, five found no association or no increased risk of myocardial infarction, atrial fibrillation, or stroke for caffeine intakes ranging from ~95 to ~1000 mg/day (Conen et al., 2010; Floegel et al., 2012; Frost and Vestergaard, 2005; Greenberg et al., 2008; Larsson et al., 2011) (Fig. 5). Of the remaining six studies, two reported an increased risk, but only for a specific genotype. Cornelis et al. (2006) reported an increased risk of nonfatal acute myocardial infarction following consumption of ≥400 mg caffeine/day in coffee, but only for slow metabolizers of caffeine. Similarly, Happonen et al. (2006) reported an increased risk in the incidence of acute myocardial infarction or CVD mortality (referred to collectively as acute coronary events) following consumption of >320 mg caffeine/day for those with a low-activity COMT genotype. Two other studies reported a significantly increased risk of a nonfatal acute myocardial infarction or ischemic stroke within 1 h of consuming coffee (Baylin et al., 2006; Mostofsky et al., 2010); however, the RR was highest at the lowest exposure category evaluated (≤95 mg caffeine/day), with a significant negative trend such that the RR was lowest at the highest exposure category evaluated (≥360 mg/day and ~>285 mg caffeine/day, respectively). In general, ORs or RRs were ~≤2, although they were as high as 3 or 4 in some cases (Baylin et al., 2006; Happonen et al., 2006).

As noted above, Happonen et al. (2004) reported a j-shaped dose-response curve, with the lowest number of acute coronary events (as defined above) associated with moderate coffee drinkers (150–320 mg caffeine/day). The RR was significantly higher in heavy coffee drinkers (>320 mg caffeine/day) based on 14 years of follow-up (RR, ~1.5), but the RRs were significantly higher in both light coffee drinkers (0.4–150 mg caffeine/day) and heavy coffee drinkers (RR, ~2) when limited to 2.5 or 5 years of follow-up. Finally, Kabagambe et al. (2007) reported a significant association between nonfatal myocardial infarction and consumption of

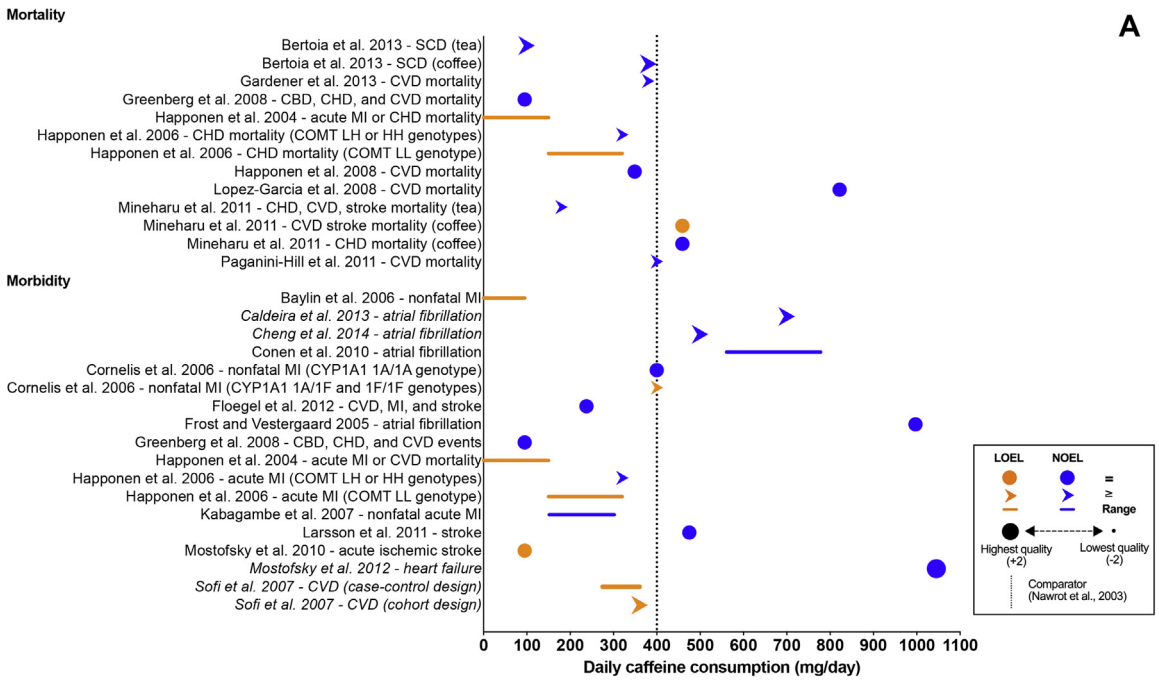


Fig. 5. (A–G) Summary diagram of exposure-response data relative to the comparator for the cardiovascular outcome: (A) morbidity and mortality, (B) blood pressure (adults), (C) blood pressure (children and adolescents), (D) heart rate (adults), (E) heart rate (children and adolescents), (F) cholesterol, and (G) heart rate variability. Symbols represent caffeine intake (mg/day) as reported by original study authors. The color of the symbol indicates the type of effect; no effect (NOEL; blue symbols) or the lowest effect level (LOEL; orange symbols). The shape of the symbol represents the type of metric (circles represent a discrete value, arrowheads represent greater than or equal to a value, and a horizontal line represents a range of values; metrics based on that reported by original study authors). The size of the symbol indicates the overall risk of bias (larger symbols indicate a lower risk of bias, or higher methodological quality). The dashed vertical line marks the comparator value. Italicized study names indicate a meta-analysis. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

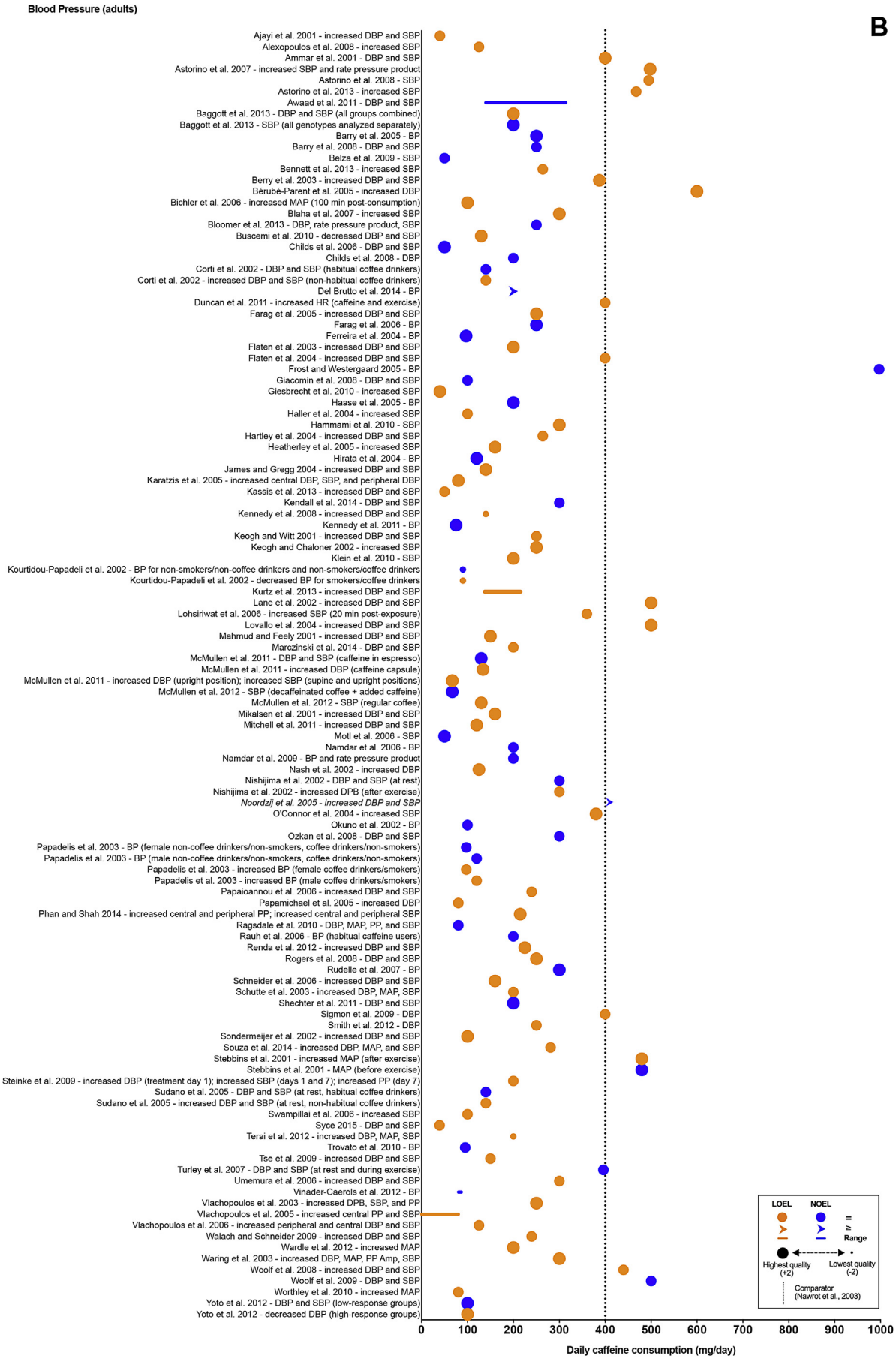
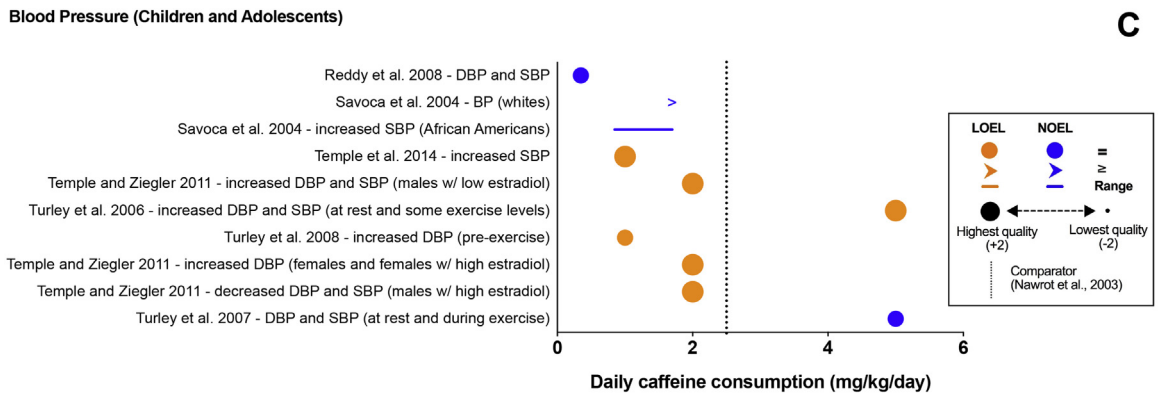
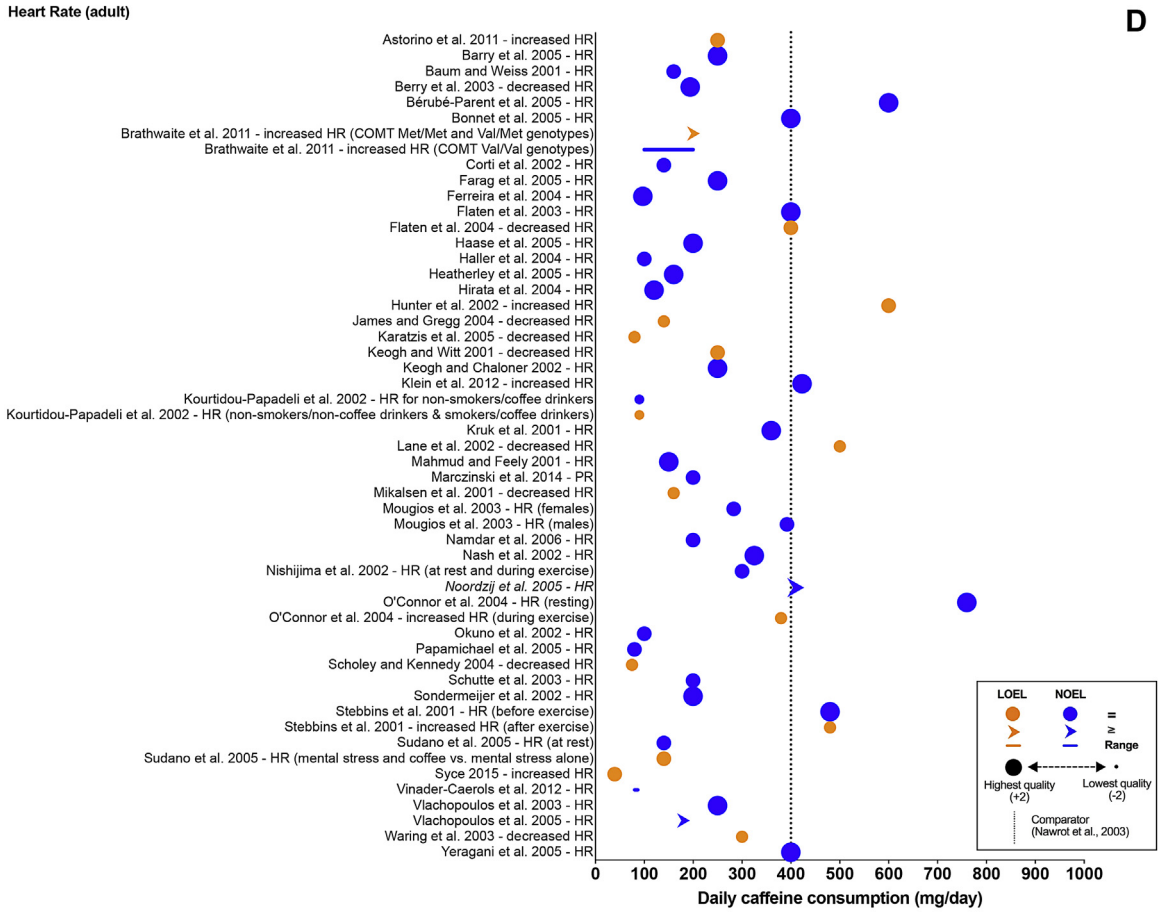


Fig. 5. (*continued*).

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002





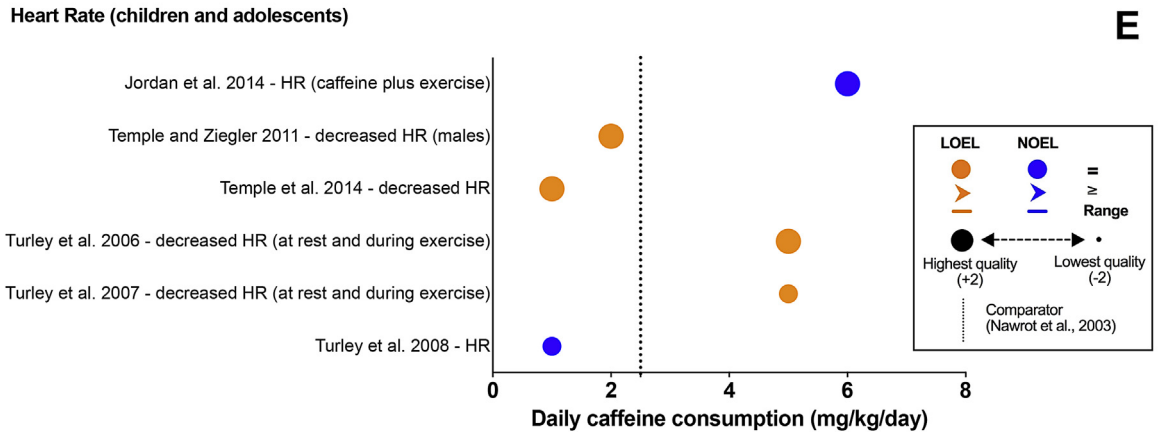


Fig. 5. (*continued*).

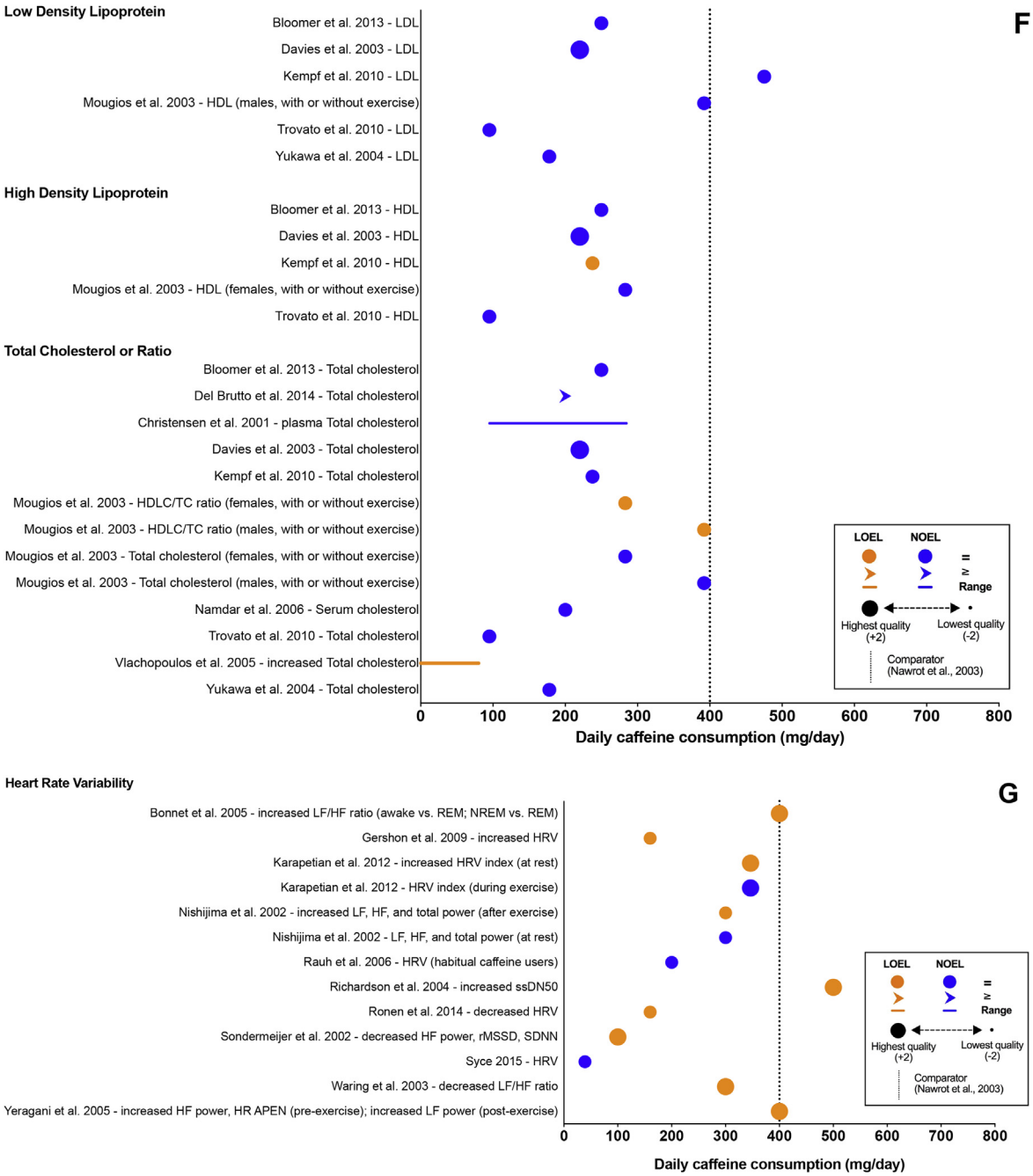


Fig. 5. (*continued*).

303–454 or >454 mg caffeine/day (ORs <2); however, this is the same starting study population as evaluated by Cornelis et al. (2006), in which the elevated RR was limited to slow caffeine metabolizers, and the study authors concede that their exclusion criteria resulted in breakage of the case/control pairs.

When the literature on morbidity are considered collectively, and considering the greater utility of meta-analyses, evidence support that 400 mg caffeine/day in healthy adult populations is an acceptable intake which is not associated with significant concern regarding cardiovascular morbidity. Several studies, including findings from two meta-analyses, suggested that the comparator is too low – that is, several studies reported a lack of effects above 400 mg/day. All comparison points above 400 mg/day were no effect levels; including no effects associated with intakes as high as 1050 mg/day. Some studies, including two meta-analyses, however, reported effect levels below the comparator (suggesting the comparator is too high), adding complexity to the integration of the data. In several cases, associations were observed only in specific genotypes, highlighting the potential role of kinetic influence on PD (discussed elsewhere). There is a moderate level of confidence in the evidence base. Confidence was increased by the low level of indirectness, and the low risk of bias in the individual studies (Table 1).

3.2.1.3. Blood pressure. More than 100 controlled trials were identified that evaluated the effect of <100 to ~1000 mg caffeine/

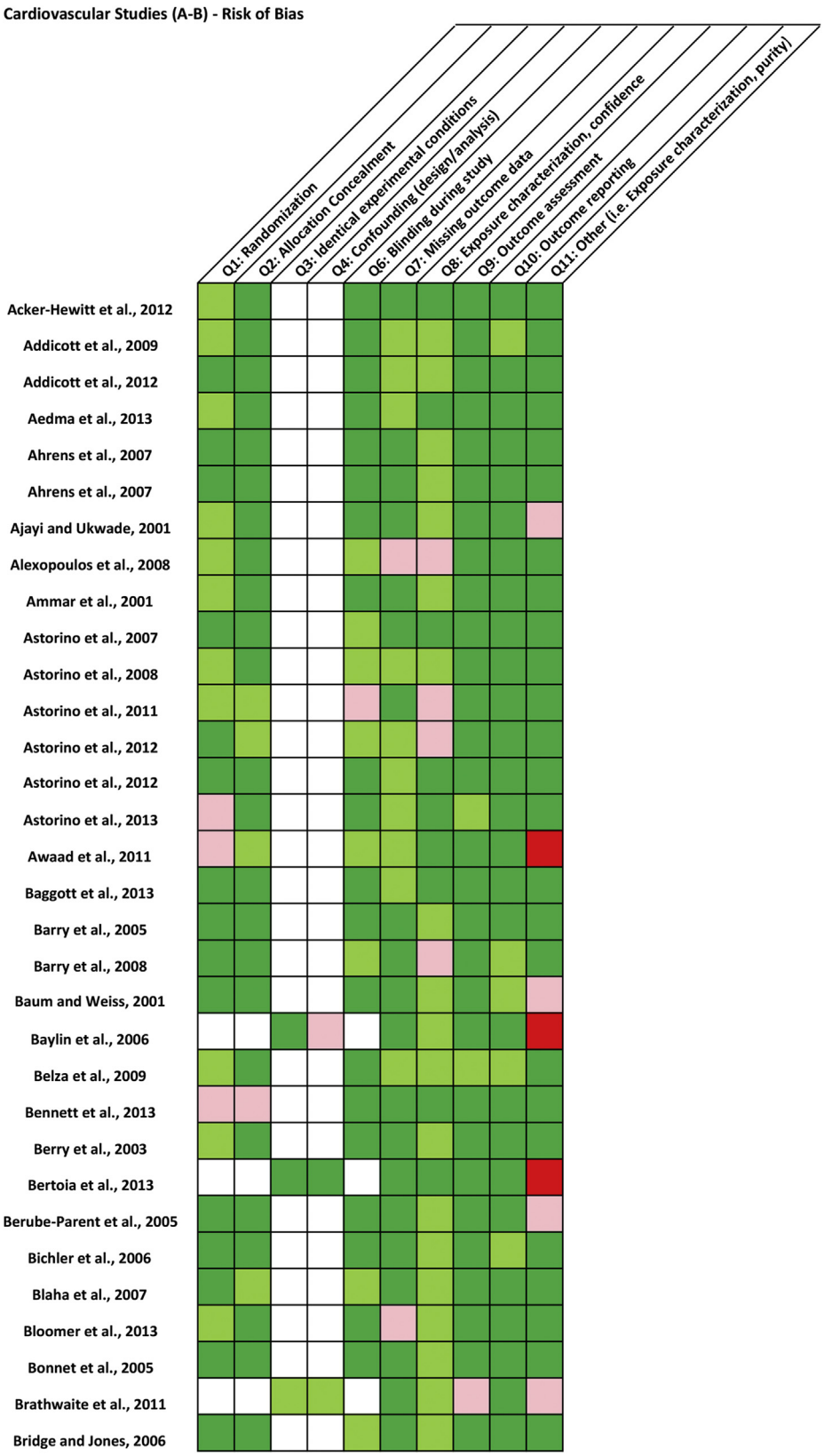


Fig. 6. Risk of bias (RoB) heat map for studies included in the cardiovascular outcome. The domain-based validity was evaluated based on study type per the OHAT (2015b) RoB tool. RoB for each domain is indicated by color: "definitely low risk of bias" (dark green, +2), "probably low risk of bias" (light green, +1), "probably high risk of bias" (light red, −1), and "definitely high risk of bias" (dark red, +2). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

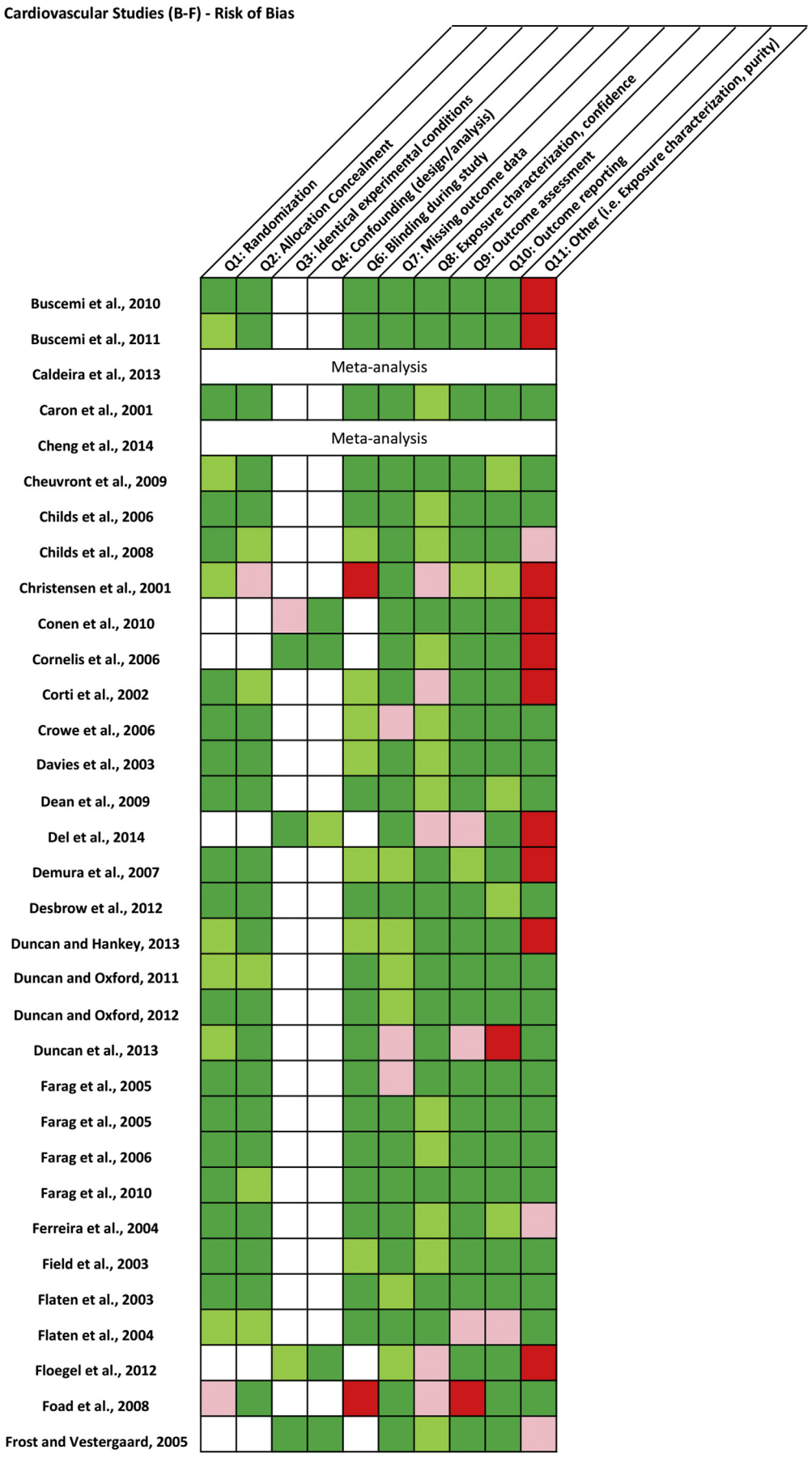


Fig. 6. (*continued*).

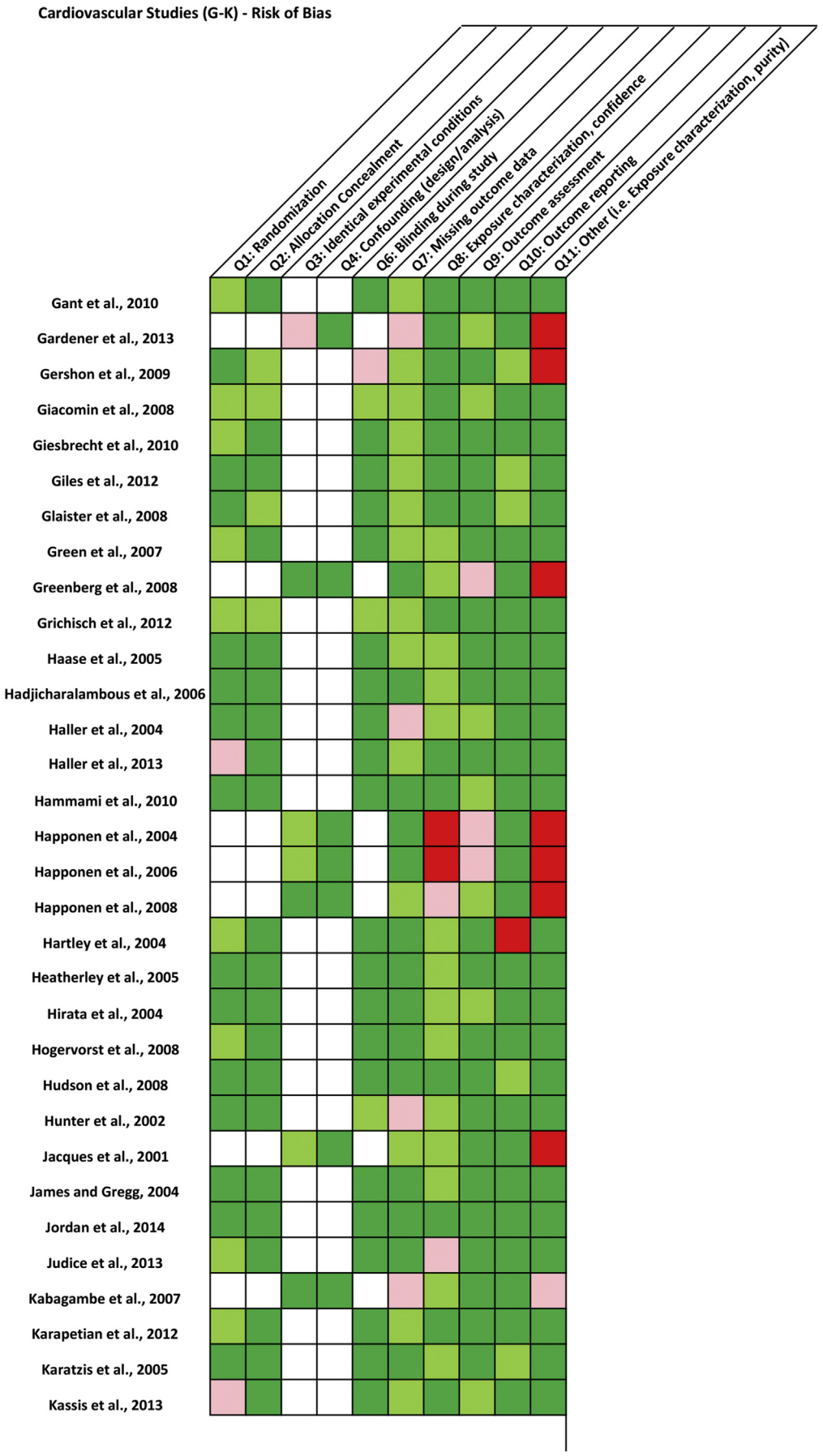


Fig. 6. (*continued*).

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

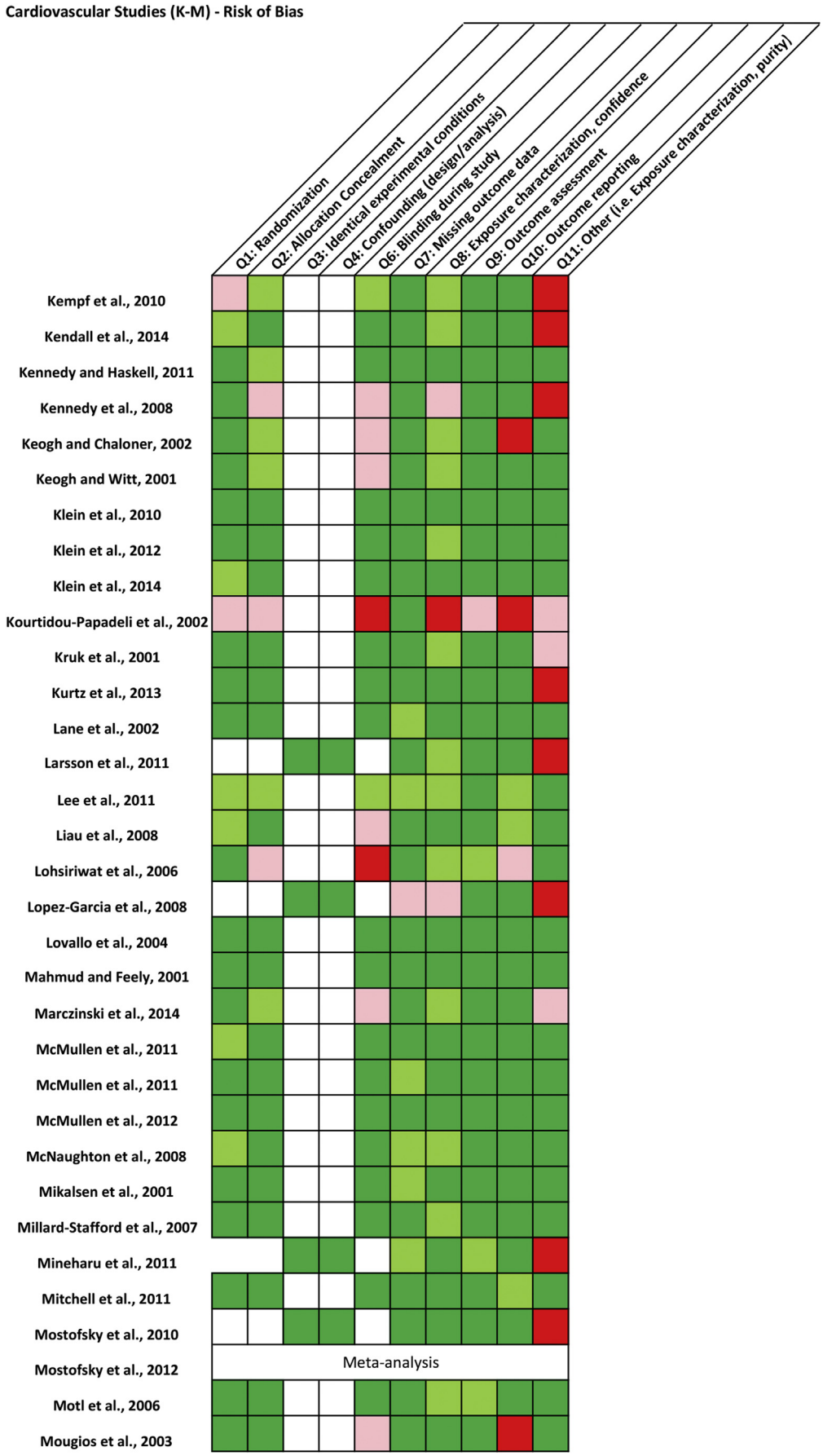


Fig. 6. (*continued*).

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

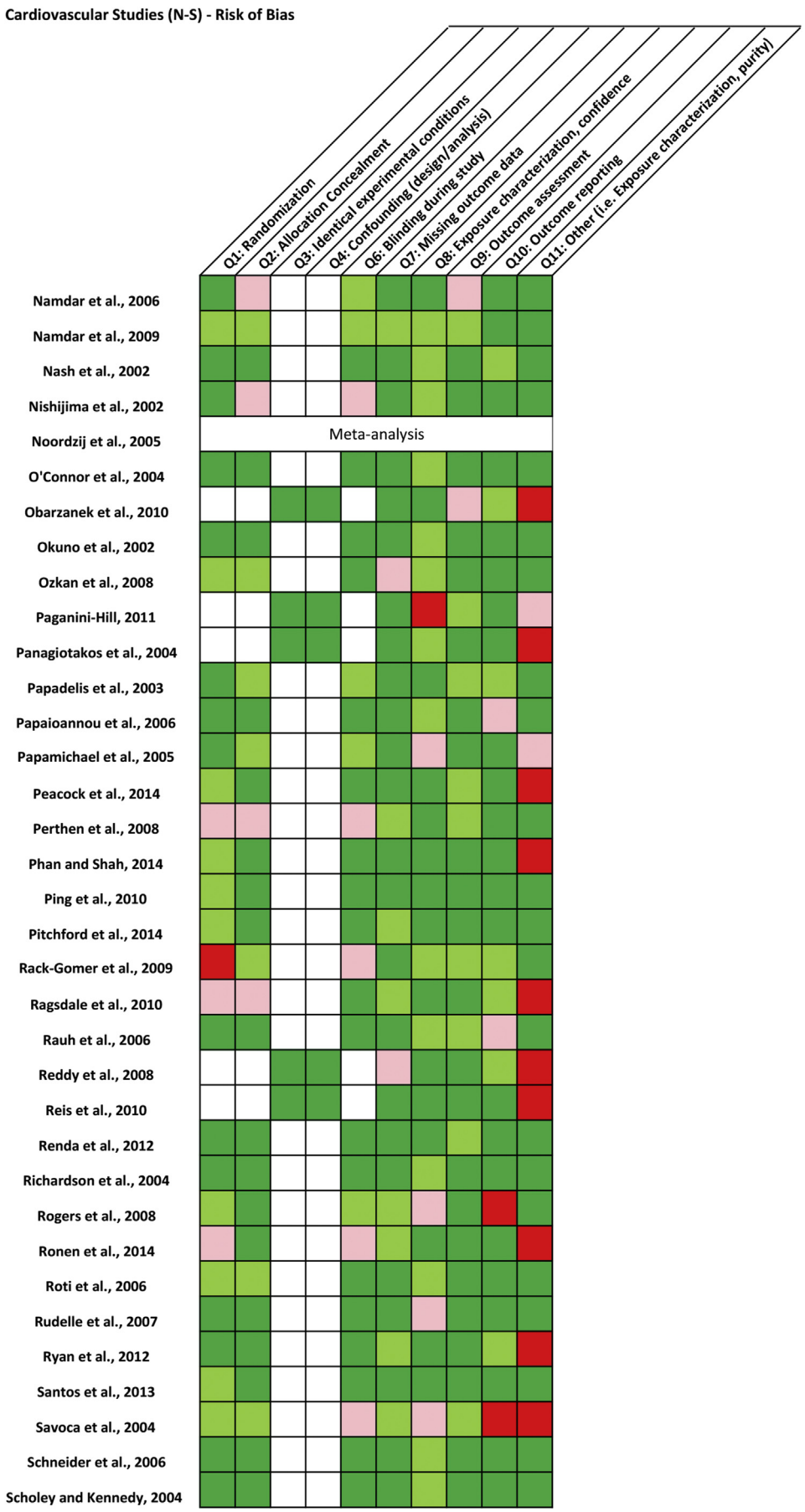


Fig. 6. (*continued*).

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

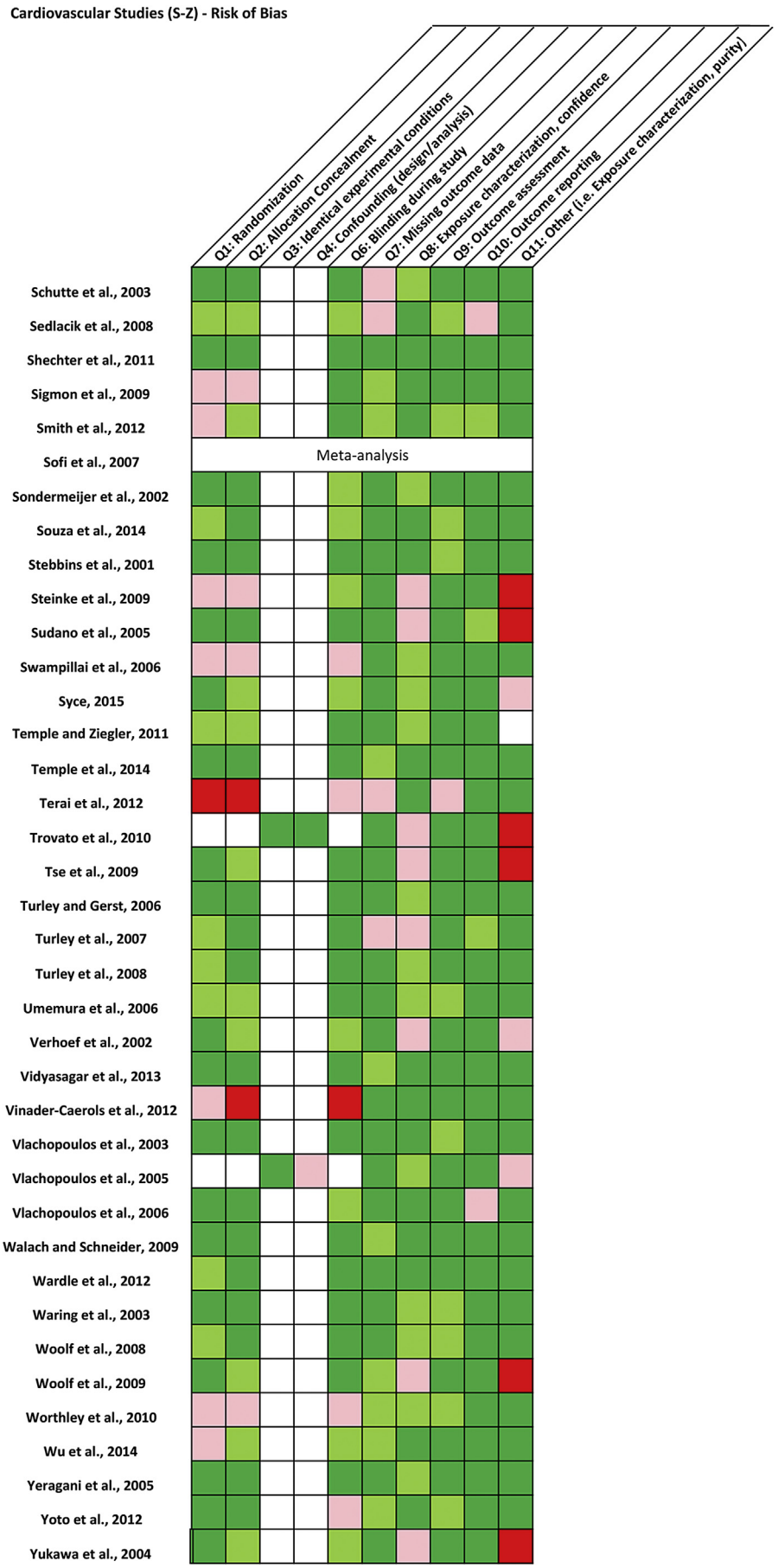


Fig. 6. (*continued*).

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

day on different aspects of blood pressure in adults (see Fig. 5B). Hypertension, a chronically elevated blood pressure, is a known risk factor for CVD (Mozaffarian et al., 2016), whereas intermittent blood pressure elevations, such as those associated with exercise, are not. The majority of the caffeine studies evaluated peripheral systolic and diastolic blood pressure, although some studies evaluated (instead of or in addition to) central (aortic) blood pressure, mean arterial blood pressure, pulse pressure, or pulse pressure amplification. The results of these studies are mixed, with some reporting statistically significant, albeit often small, increases in blood pressure at relatively low caffeine exposures (≤100 mg/day), whereas others reported no effect on blood pressure at much higher caffeine exposures (≥400 mg/day). The magnitude of change was not always reported or was difficult to discern (reported only in figures). However, as an example, Syce (2015) reported an observed increase of 0.5 mmHg in peripheral systolic blood pressure following ingestion of ~39 mg caffeine in black tea, whereas Ajayi and Ukwade (2001) reported observing an increase of ~3–4 mmHg in peripheral systolic blood pressure following ingestion of ~40 mg caffeine in instant coffee. Many studies involved exposures to 100–200 mg caffeine/day to mimic consumption of 1 or 2 cups of coffee, with some studies reporting a significant increase in blood pressure and others not. One meta-analysis of 16 randomized control studies of coffee or caffeine reported a significant increase in blood pressure associated with consumption of ≥410 mg caffeine/day for at least 7 days as compared to those who consumed <410 mg/day (Noordzij et al., 2005). The majority of the authors reporting a significant increase in blood pressure did not comment on whether the magnitude of change represented an adverse effect in healthy adults, except perhaps for individuals who already had elevated blood pressure. Thus, similar to that reported by Nawrot et al. (2003), the data from this SR demonstrated that many controlled studies report statistically significant increases in blood pressure as a result of caffeine exposures below 400 mg/day, although the magnitude of change is often small (or even very small; <1 to a few mmHg) and transient in nature, and many other studies do not report significant changes to blood pressure at exposures up to or exceeding 400 mg/day.

Two cohort studies were also identified that evaluated the effect of caffeine or coffee consumption on different aspects of blood pressure in adults (Del Brutto et al., 2014; Vlachopoulos et al., 2005). Neither reported an association between increased blood pressure and consumption of >180 mg caffeine/day in coffee or >200 mg caffeine/day, respectively. All of the exposure categories evaluated in these two studies were below the comparator of ≤400 mg/day.

Five controlled trials were identified that evaluated the effect of 1–5 mg/kg caffeine on blood pressure in children and/or adolescents (Temple and Ziegler, 2011; Temple et al., 2014; Turley and Gerst, 2006; Turley et al., 2007, 2008). A statistically significant increase in blood pressure was observed in all of these studies except Turley et al. (2007) (on the order of a few mmHg); effects were observed at doses below 2.5 mg/kg in two of the studies and above 2.5 mg/kg in two of the studies. No significant effect was observed in blood pressure following ingestion of 5 mg/kg caffeine in a study of 7- to 9-year-old boys (Turley et al., 2007), although blood pressure was consistently higher (~3–4 mmHg) as compared to controls. In addition, Savoca et al. (2004) observed a significant increase in blood pressure in African American adolescents who consumed a controlled diet containing >100 mg caffeine (>1.7 mg/kg) for 3 days as compared to those who consumed 0–50 mg/day (0.85 mg/kg) or >50–100 mg/day (>0.85–1.7 mg/kg); however, no effect was observed in white adolescents at any exposure level. Finally, one cross-sectional study was also identified, in which Reddy et al. (2008) found no association between an average dietary caffeine exposure of ~0.35 mg/kg and blood pressure in African American girls aged 6–11 years, which is well below the comparator of ≤2.5 mg/kg.

Taken together, studies were relatively consistent in demonstrating that exposures to caffeine, at intakes both below and above the comparator, have the potential to result in an increase in blood pressure (often only a few mmHg) in all populations evaluated. The biological significance of this small magnitude of change is difficult to interpret relative to the determination of adverse effects, because such a determination is likely to be conditional. Several aspects were critical to interpreting the level of adversity. First, the range of normal blood pressure variation during the day exceeds the increase that is associated with caffeine (Mancia, 2012). Second, in some cases, transient increases in blood pressure may not be harmful; for example, blood pressure is increased during exercise, which is associated with decreased cardiovascular risk. The blood pressure increase with exercise is typically substantially greater than that observed with caffeine (Miyai et al., 2002). Third, a decrease in heart rate following caffeine consumption is believed to be in response to an increase in blood pressure; as discussed below, consistent changes in heart rate were not observed following caffeine consumption. Similarly, the long-term effects of transient caffeine-mediated blood pressure increase are unknown relative to the potential impact on known cardiovascular risk factors, such as chronic hypertension. Lastly, some data indicate the potential for unique subgroups of individuals to demonstrate greater blood pressure sensitivity to caffeine than other subgroups. When the evidence is considered collectively, findings suggest that the comparator of 400 mg/day in healthy adults is too high if one is only considering the potential for caffeine to cause a physiological change in blood pressure (which may or may not be adverse). When considering the small magnitude of changes in this physiological parameter, as well as the lack of information demonstrating an association between chronic caffeine-mediated blood pressure increases relative to known cardiovascular risk factors, the comparator of 400 mg/day is likely acceptable. There is a moderate to high level of confidence in the underlying data for this endpoint, primarily driven by the low risk of bias and use of controlled exposures (RCTs). However, confidence in determining conclusions relative to the comparator is limited by the inability to ascertain the conditions and magnitude of change that would be considered adverse in a clinical or toxicological context (which is beyond the scope of this assessment).

Similar to the findings in adults, some data suggest that the comparator of 2.5 mg/kg/day in children is too high if only the potential for caffeine to cause a physiological change in blood pressure (which may or may not be adverse) is being considered; other data suggested that the comparator was too low, as no changes were observed following ingestion of 5 mg/kg. When considering the small magnitude of changes in this physiological parameter, as well as the lack of information demonstrating an association between chronic caffeine-mediated blood pressure increases relative to known cardiovascular risk factors, evidence shifts to support the comparator of 2.5 mg/kg/day. There is a moderate to high level of confidence in this body of evidence; confidence is limited by inconsistency of findings. Thus, results indicate that it would be prudent to evaluate blood pressure in children and/or adolescents with significant caffeine intake and consider limiting this for those with significant caffeine-mediated blood pressure rise.

3.2.1.4. Heart rate. More than 20 controlled trials were identified that evaluated the effect of <100 to ~750 mg caffeine/day on heart rate in adults, often during or after exercise (Fig. 5D). The results of these studies are mixed, although the majority of studies reported a

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

lack of effect. Some studies reported decreases in heart rate at caffeine exposures of ≤100 mg/day, whereas many others reported no change in heart rate at caffeine exposures of ≥400 mg/day. The magnitude of change was often not reported or only reported in figures; however, as an example of a study that addresses such, Scholey and Kennedy (2004) reported observing a decrease of ~5 bpm following consumption of 75 mg caffeine. Heart rate was often, but not always, significantly increased during or after exercise at a wide range of caffeine exposures (Fig. 5D), with the reported increase in these studies considered to be a beneficial (i.e., performance-enhancing) effect (heart rate increase during exercise is a key mechanism to improve cardiac output).

Two observational studies were also identified that evaluated the effect of caffeine or coffee consumption on heart rate (Brathwaite et al., 2011; Vlachopoulos et al., 2005). One study reported no association between heart rate and consumption of >180 mg caffeine/day in coffee, whereas the other study reported an increased likelihood of self-reporting an increased heart rate within 12 h of consuming one caffeine-containing beverage only in individuals with the *COMT* Met/Met polymorphism (slower breakdown of catecholamines) who consume >200 mg caffeine/day.

One meta-analysis of 16 randomized control studies of coffee or caffeine reported no effect on heart rate associated with consumption of ≥410 mg caffeine/day for at least 7 days as compared to those who consumed <410 mg/day (Noordzij et al., 2005).

Six controlled trials were identified that evaluated the effect of 1–6 mg/kg caffeine on heart rate in children and/or adolescents (Jordan et al., 2014; Temple et al., 2014; Temple and Ziegler, 2011; Turley et al., 2007, 2008; Turley and Gerst, 2006) (Fig. 5E). A significant decrease in heart rate (~4–6 bpm change) was observed in all of these studies but one, with two studies testing dose levels below 2.5 mg/kg and three testing dose levels above 2.5 mg/kg. The magnitude of change was again difficult to consistently discern (e.g., information in graphical format); however, one study reported a decrease of ~5 bpm (Turley et al., 2008). The one exception was a study by Jordan et al. (2014) of elite youth soccer players, in which no change in heart rate was observed following consumption of 6 mg/kg caffeine after a standard warm-up or subsequent reactive agility tests.

When the evidence for potential changes to heart rate is considered collectively, data support that the comparator of 400 mg caffeine/day in healthy adults is acceptable as an intake which is not associated with meaningful concern regarding adverse effects on heart rate. There is a moderate to high level of confidence in this evidence base (Fig. 6; Table 1). Confidence in determining conclusions relative to the comparator is limited by the inability to ascertain the conditions and magnitude of change that would be considered adverse in a clinical or toxicological context (which is beyond the scope of this assessment). For children and adolescents, data support a relationship between caffeine exposure and decreased heart rate; however, further characterization of exposures associated with such were difficult, given that changes were observed in studies both below and above the Nawrot et al. (2003) comparator of 2.5 mg/kg— yet no changes were observed in a study involving exposure to 6 mg/kg. Thus, it was determined that the evidence base was insufficient to render a conclusion regarding appropriateness of the comparator for potential impacts of caffeine consumption on heart rate in children and adolescents.

3.2.1.5. Cholesterol. Seven controlled trials were identified that evaluated the effects of 180–475 mg caffeine/day on serum cholesterol (Bloomer et al., 2013; Christensen et al., 2001; Davies et al., 2003; Kempf et al., 2010; Mougios et al., 2003; Namdar et al., 2006; Yukawa et al., 2004). Increased total serum or low-density lipoprotein (LDL) cholesterol is a well-recognized risk factor for CVD (Mozaffarian et al., 2016). Three studies were of caffeine, and the remaining studies were of caffeinated coffee or tea. A significant increase in total cholesterol was observed following consumption of ≥380 mg/day caffeine in filtered coffee for 4–6 weeks (no effects at 95–285 mg/day) (Christensen et al., 2001). In contrast, Kempf et al. (2010) reported no significant effects of consumption of 238 or 475 mg caffeine on total cholesterol or LDL cholesterol, respectively, and reported significant increases in high-density lipoprotein (HDL) cholesterol (considered a beneficial effect) at consumption of 238 mg/day (Kempf et al., 2010). A significant increase in the HDL/total cholesterol ratio, which is also considered a beneficial effect, was observed following consumption of 283 mg or 392 mg caffeine (in females and males, respectively) (Mougios et al., 2003). The magnitude of observed change in these studies was on the order of 10–20 mg/dL. No changes or significant decreases in cholesterol (in the latter case, total or LDL cholesterol) were observed in the remaining four studies of caffeine, coffee, or tea associated with lower caffeine intakes of 180–250 mg/day for a single day or up to 12 weeks.

The results of three cohort studies are inconsistent. Vlachopoulos et al. (2005) observed a significant increase in total cholesterol in participants with self-reported consumption of <80, 80–180, and >180 mg caffeine/day in coffee; however, no dose-response was observed. LDL cholesterol was significantly higher only in the highest exposure category (>180 mg caffeine/day in coffee). In contrast, Trovato et al. (2010) did not observe changes in total, HDL, or LDL cholesterol in participants with self-reported consumption averaging 95 mg caffeine/day in espresso. Del Brutto et al. (2014) also did not observe changes in total cholesterol in participants with self-reported caffeine consumption of up to >200 mg/day. Thus, for the controlled trials, a significant increase in total cholesterol was observed only in the two studies of relatively high caffeine consumption (≥380–475 mg/day), and one of the three cohort studies reported a statistically significant increase in total cholesterol following self-reported exposures to <80, 80–180, and >180 mg caffeine.

More than other endpoints evaluated, data are relatively consistent in showing a lack of effect of caffeine consumption on cholesterol at intakes below and above the comparator (Fig. 5F), thus supporting that for cholesterol, 400 mg/kg is an acceptable comparator in healthy adults (Table 2). There is a moderate to high level of confidence in the evidence base supporting this conclusion (Fig. 6; Table 1).

3.2.1.6. Heart rate variability. Twelve controlled trials were identified that evaluated the effect of ~40–500 mg caffeine/day on heart rate variability in adults (Bonnet et al., 2005; Gershon et al., 2009; Karapetian et al., 2012; Nishijima et al., 2002; Ragsdale et al., 2010; Rauh et al., 2006; Richardson et al., 2004; Ronen et al., 2014; Sondermeijer et al., 2002; Syce, 2015; Waring et al., 2003; Yeragani et al., 2005). Most subjects were habitual consumers of caffeine or coffee, whereas others were relatively caffeine naïve or not specified. The results from these studies were not consistent (Fig. 5G). Five studies did not observe an effect on resting heart rate variability at exposures ranging from ~40 to 300 mg caffeine/day (Nishijima et al., 2002; Ragsdale et al., 2010; Rauh et al., 2006; Syce, 2015; Waring et al., 2003), although two of these studies (Nishijima et al., 2002; Waring et al., 2003) did report effects during exercise following exposure to 300 mg caffeine/day (cycling or hand grip exercises, respectively).

The remaining seven studies all reported significant changes in one or more measures of heart rate variability, with lower exposures (100–200 mg/day) resulting in significant decreases (Gershon et al., 2009; Ronen et al., 2014; Sondermeijer et al., 2002)

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

and higher exposures (~350–500 mg/day) resulting in significant increases (Bonnet et al., 2005; Karapetian et al., 2012; Richardson et al., 2004; Yeragani et al., 2005). Thus, caffeine exposures in the range of 400 mg/day generally resulted in increases in low (LF) and high frequency (HF) power, LF/HF ratio, or total power, whereas lower exposures appear to result in decreases in HF or the standard deviation of NN intervals (SDNN). The exception was Bonnet et al. (2005), who observed an increase in LF/HF ratio following consumption of 400 mg/day. The magnitude of observed changes is also difficult to compare across studies because of the variety of metrics used to evaluate heart rate variability, and it is difficult to discern in some cases because the data are only reported in figures. However, as an example, Sondermeijer et al. (2002) observed an average decrease in SDNN of approximately 16 or 17 ms (~23–24%) following consumption of 100 or 200 mg caffeine, respectively.

When the evidence is considered collectively, there was no consistent effect of caffeine on HRV at intakes below or above the comparator, thus supporting that 400 mg caffeine/day in healthy adults is an acceptable intake which is not associated with significant change in heart rate variability. There is a moderate to high level of confidence in the data (Fig. 6; Table 1).

3.2.2. Body of evidence assessment

Overall, the initial confidence in the body of evidence is high (OHAT, 2015a), as studies involved controlled exposures or exposure prior to the outcome, data were reported based on individual outcomes, and comparison groups were used in the studies evaluated. The low risk of bias scores and low level of indirectness increase confidence in the overall body of evidence (see Fig. 6). The majority of the 203 studies addressing the cardiovascular effects of caffeine reviewed in this SR were associated with a "definitely low" or "probably low" risk of bias, with only 5 studies associated with a "probably high" risk of bias (Fig. 6; Table 1). Similarly, most studies were associated with a low level of indirectness. This is primarily due to the fact that most studies were RCTs specifically designed to assess cardiovascular effects of pure caffeine. All of the observational studies but one were associated with a "probably low" risk of bias (rather than "definitely low," primarily because of uncertainty in the exposure level), which were all based on self-reported consumption. The magnitude of the effects, when observed at all, was often small. It is likely that some of the inconsistency observed with the observational studies was due to classification of exposure based on self-reported information, although most authors relied on validated questionnaires and controlled for the most common confounders (e.g., age, weight, smoking status).

Based on their review of data published prior to this SR, which focused on five endpoints (blood pressure, heart rate, cholesterol, arrhythmia (unspecified), and CVD), Nawrot et al. (2003) concluded "that moderate caffeine intake ($\leq$400 mg caffeine day^{-1}) does not adversely affect cardiovascular health. There are insufficient epidemiological data to draw any conclusions about the risk for coronary heart disease or mortality associated with consumption of 10 or more cups of coffee per day ($\geq$1000 mg caffeine/day)." The current body of evidence characterizing this outcome, which consisted of 203 studies, demonstrates that caffeine can cause a variety of physiological effects on the cardiovascular system at consumption levels below 400 mg/day for adults and 2.5 mg/kg for children and adolescents. However, these effects are often very small (although statistically significant), are transient in nature, and may affect only specific subsets of the population (specific genotypes), and at least some habitual caffeine consumers develop tolerance over time. For some endpoints, the results were fairly consistent across studies (e.g., blood pressure, aortic stiffness, cerebral blood flow), whereas for other endpoints, the results are mixed (e.g., heart rate, catecholamines). For two endpoints (i.e., endothelial function and heart rate variability), the results, in some cases, were even counterintuitive (e.g., changes associated with beneficial effects, or detrimental effects, are observed at low exposures but not at higher exposures). For some endpoints (i.e., endothelial function and heart rate variability), the results, in some cases, suggested that there might be a more complex relationship between dose and response, such that directional changes at low exposures differed from those at higher exposures. Such complexities underscore the limitations of characterizing potential long-term effects of caffeine exposure on cardiovascular health based on short-term (often single-exposure) controlled trials.

The majority of observational studies of clinical endpoints suggest that consumption of caffeine was not associated with an increased risk of cardiovascular morbidity or mortality. When statistically significant effects were observed below the comparator, the magnitude was low (ORs, RRs, or HRs were generally below 2 but ranged as high as 3 or 4 in some studies), and most showed either a j-/u-shaped or negative dose-response curve.

In summary, the SR of 191 studies evaluated the potential for caffeine to be associated with adverse cardiovascular effects, including mortality, morbidity, blood pressure, heart rate, cholesterol, and heart rate variability (and others; see Supplementary File S2). When the weight of evidence was considered, with particular emphasis on level of adversity, 400 mg caffeine/day was found to be an acceptable intake which is not associated with significant concern regarding adverse cardiovascular effects in healthy adults. For clinical endpoints, some findings suggested that the comparator was too low; however, other data, particularly those for physiological endpoints, reported effects below the comparator. For such physiological endpoints (e.g., blood pressure), confidence in determining conclusions relative to the comparator was limited by the inability to ascertain the conditions and magnitude of change that would be considered adverse in a clinical or toxicological context. There is a moderate to high level of confidence (Table 1) in this evidence base.

Data in children and adolescents were limited to 11 studies that evaluated physiological endpoints. As such, it was determined that the evidence base was insufficient to render a conclusion regarding appropriateness of the comparator for potential impacts of caffeine consumption on cardiovascular outcomes in these populations. The available data for blood pressure and heart rate are inconsistent; several studies report physiological changes below the comparator, although some studies reported a lack of effect on these parameters following consumption of $\geq$5 mg/kg/day.

3.3. Behavior

The full text review of the behavior literature began with 204 published studies, of which 80 were ultimately included and 124 were excluded (Fig. 2). Of the excluded papers, 66 did not provide a quantitative finding that could be used for comparison. Of the remaining studies, 58 met some other exclusion criteria (*e.g.*, unhealthy study population, no data on adverse effects, etc.) or could not be retrieved (n = 3 papers). The majority (approximately 77%) of the included papers were controlled trials using healthy adult populations and only five of the included studies specifically investigated the adverse effects of caffeine in child or adolescent populations (although >40 studies were identified, but not included, as having qualitative information). Effects in the younger populations will be discussed separately from the effects in adults. Of the controlled studies, 51 administered some form of pure caffeine, whereas the rest provided caffeine in the form of coffee, energy drinks, or some other source. The remainder of the included studies were cross-sectional (n = 12) and cohort (n = 6) studies in addition to a single case-control study, and caffeine exposure was

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

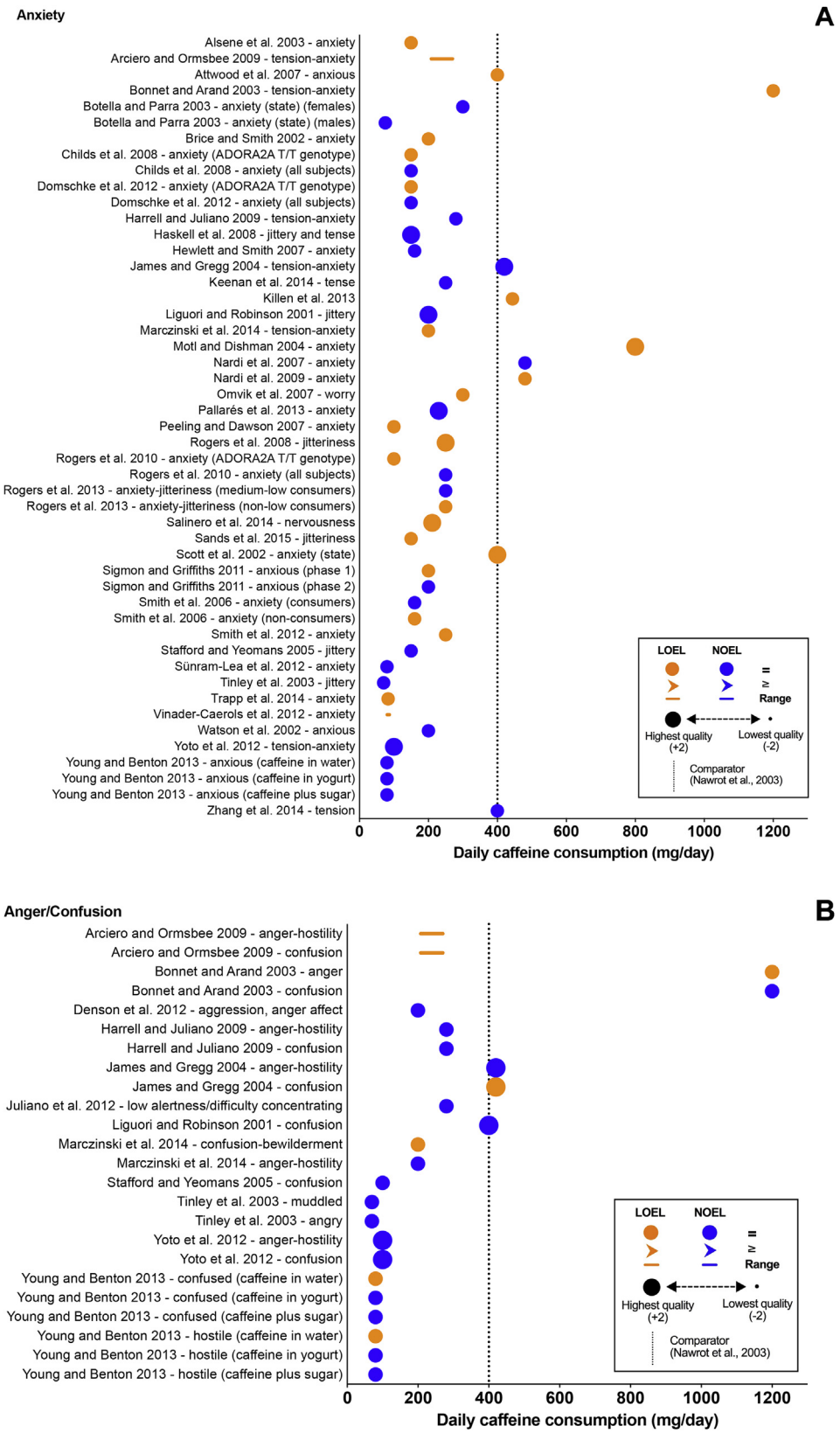
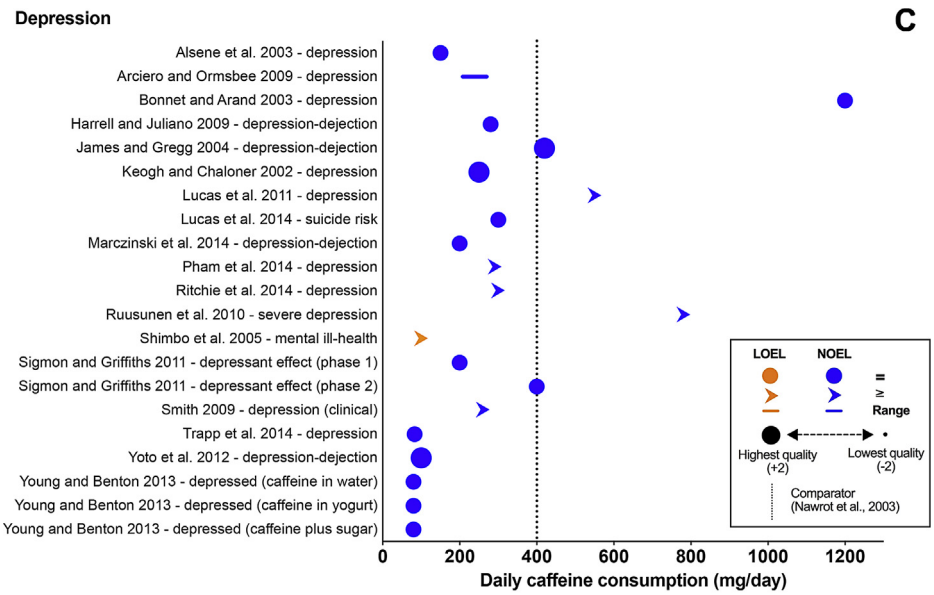
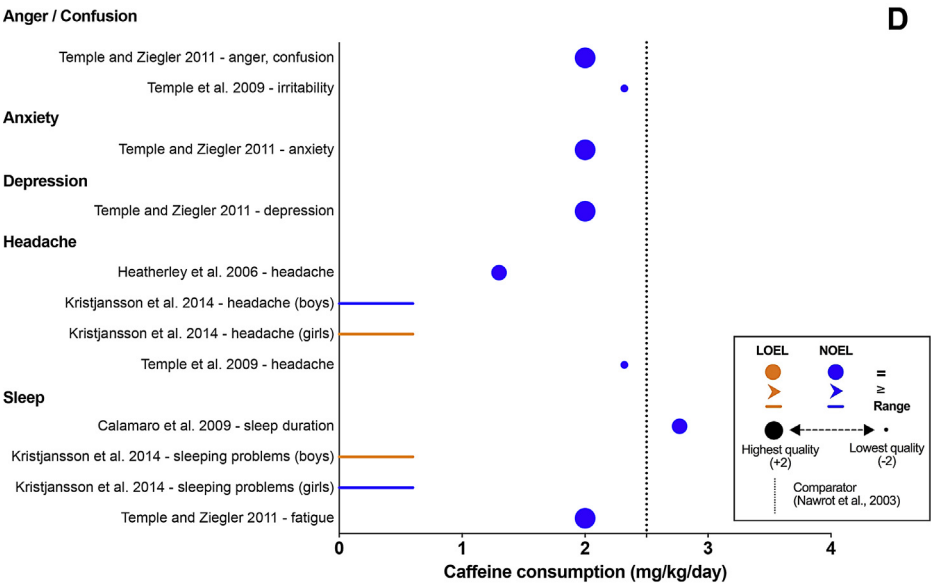
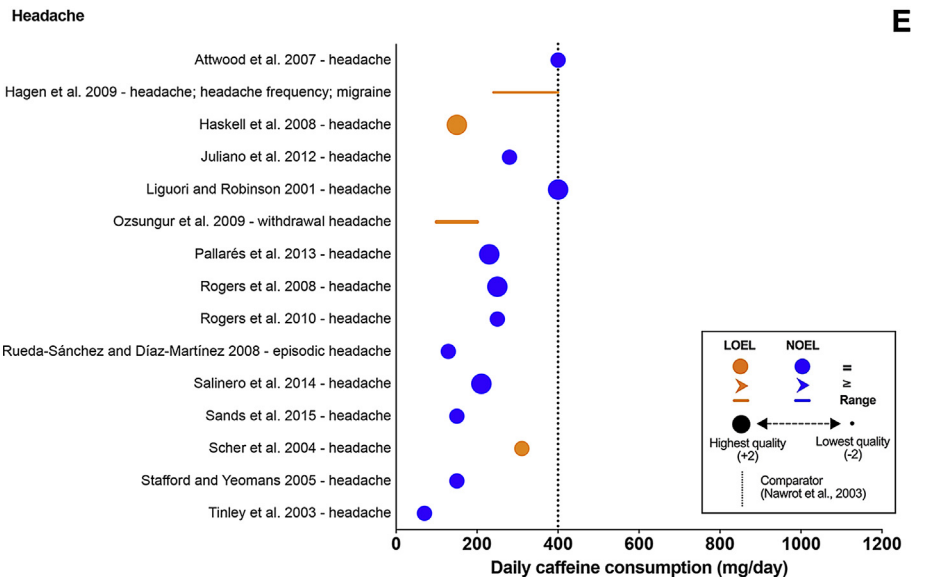


Fig. 7. (A–F) Summary diagram of exposure-response data relative to the comparator for the behavior outcome: (A) anxiety, (B) anger/confusion, (C) depression, (D) all endpoints (children and adolescents), (E) headache, (F) sleep. Symbols represent caffeine intake (mg/day) as reported by original study authors. The color of the symbol indicates the type of effect; no effect (NOEL; blue symbols) or the lowest effect level (LOEL; orange symbols). The shape of the symbol represents the type of metric (circles represent a discrete value, arrowheads represent greater than or equal to a value, and a horizontal line represents a range of values; metrics based on that reported by original study authors). The size of the symbol indicates the overall risk of bias (larger symbols indicate a lower risk of bias, or higher methodological quality). The dashed vertical line marks the comparator value. Italicized study names indicate a meta-analysis. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002







Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

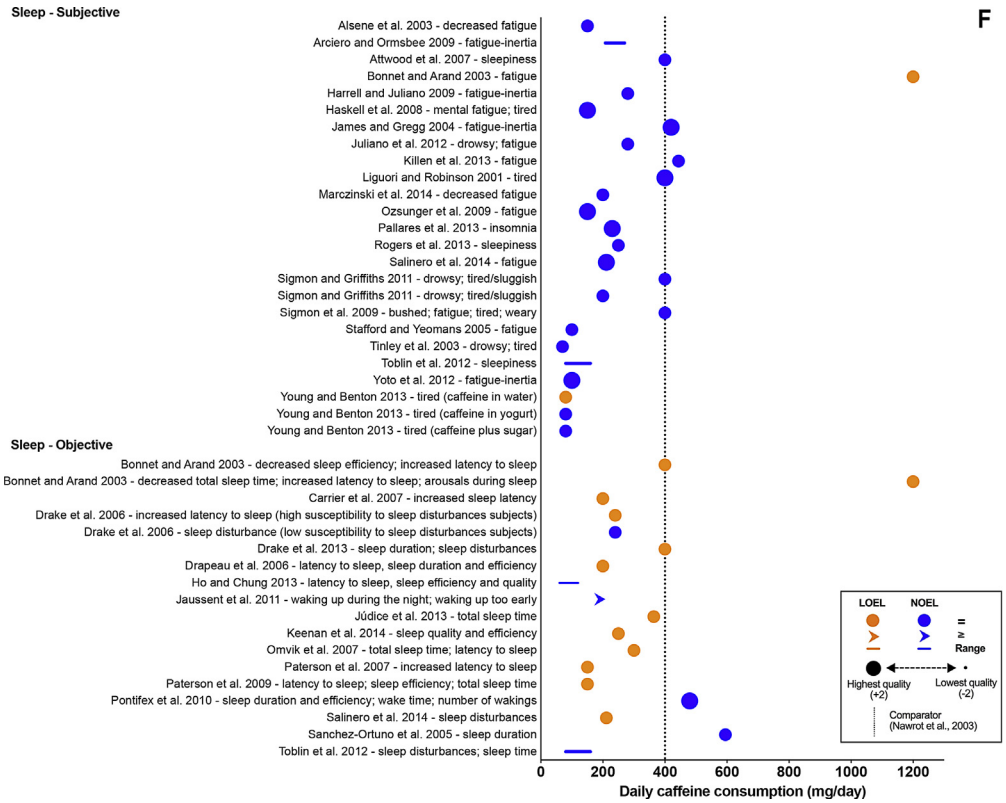


Fig. 7. (*continued*).

self-reported in all of them. Most of the analyses were conducted using categorical groupings of caffeine exposure such as high versus low consumption or a cups-per-day metric. Overall, two-thirds of the studies reviewed directly evaluated the amount of caffeine as part of the analysis and association with the endpoints, whereas for the remaining third, the amount of caffeine for comparison was calculated per the methods in this paper (Section 2).

In studies in which multiple variables or confounders were accounted for, smoking, age, and gender were often covariates. Other more specific variables such as anxiety sensitivity or sleep behavior were also sometimes considered, depending on the endpoint objective. By design, the clinical studies that comprise the majority of the results herein, regardless of the findings, evaluated doses at or below the 400 mg/day intake put forward for adults in Nawrot et al. (2003) (Fig. 7). Because most of the studies were clinical trials and/or evaluated the effects of caffeine specifically, the overall level of indirectness for the body of evidence is low; as such, these controlled studies provide more weight in the assessment relative to others.

The endpoints characterized in the behavior outcome fall into several major categories: mood, sleep, withdrawal, and headache, generally mirroring the endpoints in Nawrot et al. (2003). One exception is risk-taking behavior (*e.g.*, licit or illicit drug use or behavioral problems), which is discussed herein and has become particularly more prominent as an area of interest in adolescents and young adult populations, resulting from the rise in popularity of energy drink consumption in these cohorts (Skewes et al., 2013). Notably, the majority of the studies identified on risk taking did not provide a quantitative caffeine value for a comparison to the intake level in Nawrot et al. (2003) and thus were excluded. For brevity, endpoints considered to be less adverse, such as effects on hunger, self-reported bruxism, or others with low information, are not discussed but were reviewed and can be found in the supplementary material via AHRQ (https://srdr.ahrq.gov/projects/1116/).

3.3.1. Summary of individual studies by endpoint

3.3.1.1. Mood. The category of "mood" was subdivided, consisting primarily of studies on the relationship between caffeine intake and anxiety, but also includes effects on other general mood states. Collectively, these are usually measured by questionnaires, such as the Profile of Mood States (POMS), that gauge items such as vigor, depression, fatigue, anger, and confusion, along with anxiety. These dimensions represent normal, nonclinical mood states and changes in them do not necessarily indicate negative effects. Furthermore, although the POMS (for example) was developed to measure multiple dimensions of baseline and transient moods, an intercorrelation between scores for some of the factors has been observed, indicating that these do not change independently (McNair et al., 1981; Norcross et al., 1984). Mood states such as alertness and vigor have repeatedly been shown to be positively affected by caffeine at a wide range of doses but these have not been reviewed here because they are not adverse effects (Lieberman, 2001; Ruxton, 2008).

3.3.1.1.1. Anxiety in adults. Forty controlled trials were identified that evaluated the effects of 70–1200 mg caffeine/day on some aspect of anxiety (Fig. 7A) in adults. Most of these studies used questionnaires, such as the POMS or visual analogue scales (VASs), to measure subjective effects on anxiety, including effects on state measures such as "tension," "jitteriness," "nervousness," and "worry," depending on the study design and type of questionnaire. Consideration must also be given to the possibility that some of these subjective effects categorized as anxiety may also be related to caffeine's ability to increase alertness and arousal (Nawrot et al., 2003).

The results of the five studies evaluating effects of caffeine at the lower end of the range (<100 mg) largely indicate no or little effect on anxiety. One study reported a small but statistically significant increase in anxiety from doses of caffeine as low as 80–87 mg (Vinader-Caerols et al., 2012); however, the others that administered doses just below this range (70–80 mg) in the form of coffee, tea, or other caffeine sources did not report any significant change (Botella and Parra, 2003; Sünram-Lea et al., 2012; Tinley et al., 2003; Young and Benton, 2013). The majority of the remaining controlled trials administered single doses of caffeine between 100 and 400 mg/day and reported mixed results. Although nine investigations reported no anxiogenic effect of single caffeine doses ranging from 100 to 400 mg (Harrell and Juliano, 2009; Haskell et al., 2008; Hewlett and Smith, 2007; Keenan et al., 2014; Pallarés et al., 2013; Stafford and Yeomans, 2005; Watson et al., 2002; Yoto et al., 2012; Zhang et al., 2014), 12 other studies have shown statistically significant increases, of varying magnitude, in measures of anxiety following ingestion of caffeine in this dose range (Arciero and Ormsbee, 2009; Attwood et al., 2007; Brice and Smith, 2002; Marczinski et al., 2014; Omvik et al., 2007; Peeling and Dawson, 2007; Rogers et al., 2008; Salinero et al., 2014; Sands et al., 2015; Scott et al., 2002; Sigmon and Griffiths, 2011; Smith et al., 2012). Of the seven controlled studies using doses above 400 mg/day, just two found no effect of caffeine on anxiety measures (James and Gregg, 2004a,b; Nardi et al., 2007), although it should be noted that James and Gregg administered the caffeine dose as 1.75 mg/kg body weight three times per day. The remaining five high-dose caffeine (i.e., >400 mg/day) studies demonstrated that caffeine significantly increased some aspect of anxiety (Bonnet and Arand, 2003a; Killen et al., 2013; Motl and Dishman, 2004; Nardi et al., 2009; Pallarés et al., 2013). A single observational study was identified that characterized the effects of caffeine on anxiety. Using a logistic regression analysis, Trapp et al. (2014) found that consumption of energy drinks in young adult males, equivalent to approximately 83 mg caffeine/day (100 mL/day), was associated with an increase in anxiety (OR, 1.23; 95% CI, 1.03–1.48) and stress scores (OR, 1.21; 95% CI, 1.01, 1.45); no such effects were observed in females.

Taken together, some but not all evidence, primarily from RCTs involving single/short term caffeine exposure and subjective measures of anxiety, suggests that the comparator of 400 mg/day can lead to increases, albeit small, in measures of anxiety in adults. Evidence suggests that some of the variability in findings may be due to individual variation in the response to caffeine. Much of the variation in response, including the magnitude and sensitivity of its effects, may be attributed to other factors such as genotype (e.g., ADORA2a receptor TT genotype polymorphisms), consumer status (e.g., studies that involve participants with genetic predispositions or participants who do not regularly consume caffeine tend to report effects below the comparator), and subjectivity inherent to the evaluation of anxiety (e.g., POMS). This later point being notable when considering the level of adversity of the reported changes in anxiety –the often-small changes observed were considered to be of low magnitude. These findings highlight the need for additional research to further characterize population-based sensitivities based on genotype and consumption status.

3.3.1.1.2. Anger and confusion in adults. As mentioned above, the POMS questionnaire and the VAS can be used to measure anger/hostility and confusion/bewilderment among other mood states (McNair et al., 1981). Nine controlled studies reported the effects of 70–1200 mg caffeine/day on anger, seven of which assessed doses less than 400 mg. Similar to the findings in Nawrot et al. (2003), none of these seven studies identified a negative effect of <400 mg caffeine on levels of anger/hostility (Arciero and Ormsbee, 2009; Denson et al., 2012; Harrell and Juliano, 2009; Marczinski et al., 2014; Tinley et al., 2003; Yoto et al., 2012; Young and Benton, 2013). With respect to the two studies that administered >400 mg/day, findings were mixed. James and Gregg (2004a,b) administered caffeine three times daily for 7 days and found that in well-rested individuals, a cumulative dose of 420 mg caffeine/day resulted in no significant effect on negative mood states, including anger/hostility. However, at a much higher dose level, Bonnet and Arand (2003a,b) reported that 1200 mg caffeine/day (administered as 400 mg, three times daily) did produce a statistically significant increase in POMS anger scores following 7 days of caffeine treatment.

Mood states related to confusion (including muddled, difficulty concentrating, and bewilderment) were measured in 11 controlled trials. Nine of these studies used caffeine doses ranging from 70 to 400 mg. Six of these studies using ≤400 mg determined that there was no negative effect of caffeine on measures of confusion/bewilderment (Arciero and Ormsbee, 2009; Harrell and Juliano, 2009; Liguori and Robinson, 2001; Stafford and Yeomans, 2005; Tinley et al., 2003; Yoto et al., 2012) and one study (Juliano et al., 2012) indicated an improvement in this endpoint. One exception was the study by Marczinski et al. (2014), who found that consumption of a 2-ounce, 5-h energy shot containing 200 mg caffeine resulted in an increase in subjective ratings of confusion/bewilderment 40 min after dosing, which was evident when compared to the "no beverage" group but was not significant when compared to the placebo beverage condition. The second exception was Young and Benton (2013), who found that subjects reported an increase in confused mood following consumption of 80 mg caffeine administered with water, but researchers observed no such effect, or an improvement, when caffeine was delivered with other vehicles such as glucose or yogurt, respectively. The authors speculated that there is a possible influence of caffeine vehicle on mood, which is related to the glycemic index (i.e., a lower glycemic index is more likely to be associated with observed benefits). Findings from the two studies that evaluated doses of caffeine above 400 mg (also discussed above) differed. In the first study, James and Gregg (2004a,b) reported that 420 mg caffeine/day resulted in a significant increase in confusion/bewilderment scores and an overall decrement in mood after caffeine consumption. In the other study, Bonnet and Arand (2003a,b) reported increasing dysphoria (general unhappiness) but no significant effect on confusion scores in subjects over the course of a 7-day, 1200-mg/day (administered as 400 mg, three times daily) caffeine protocol designed to produce insomnia by increasing physiological arousal in moderate caffeine consumers. Here, the authors suggested that the adverse changes in mood are related to the insomnia induced by the caffeine consumption (see more on this study in Section 3.3.1.3 on sleep). Following a 2-night withdrawal phase (after the 7-day caffeine protocol), however, scores for anger, depression, and confusion were, in fact, increased (Bonnet and Arand, 2003a).

Overall, the evidence, which is primarily based on data obtained from short term/single exposure trials, suggests that the comparator of 400 mg/day is an acceptable intake that is not associated with significant concern regarding anger and confusion mood states in adults (Fig. 7B). The majority of data showed a lack of effects following exposure; some studies reported effects at intakes both above and below the comparator, particularly during a withdrawal phase – such effects were generally considered to be of low magnitude. For the later, most studies reported a combination of effects/no effects for different endpoints (e.g, confusion or hostility) or different vehicles (e.g., caffeine in water vs. caffeine in yogurt).

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

The confidence in this evidence base is high.

3.3.1.1.3. Depression in adults. Depression or other related endpoints was measured in 10 controlled trials; in these studies, caffeine consumption at all doses tested (ranging from 80 to 1200 mg/day) had no negative impact on this endpoint. This conclusion is similar to that of Nawrot et al. (2003). Specifically, in seven of the clinical studies in which caffeine was administered acutely, caffeine had no effect on scores of depression (Bonnet and Arand, 2003a; Harrell and Juliano, 2009; James and Gregg, 2004a; Marczinski et al., 2014; Sigmon et al., 2009; Yoto et al., 2012; Young and Benton, 2013). In the other three studies, acute administration of caffeine improved scores for this endpoint. For instance, Alsene et al. (2003) found that 150 mg caffeine decreased POMS scores measuring depression. Moreover, Keogh and Chaloner (2002) found that consumption of 250 mg caffeine (in coffee) significantly improved the depression/elation scores of women who had low anxiety-sensitivity scores, although it had no effect on the scores of women who had medium or high sensitivity to anxiety. Lastly, following the ingestion of 208–270 mg caffeine, depression ratings were significantly lower after caffeine ingestion in older women but not younger women (Arciero and Ormsbee, 2009).

There were also eight observational studies that evaluated the effect of caffeine on depression, mental health, or suicide risk, only one of which identified any increased risk. Shimbo et al. (2005) found that higher green tea consumption (assumed by the authors to contain 30 mg caffeine per 150-mL cup) was associated with a small but significant increased risk of "ill mental health" in females, but not males (OR for females, 1.26; 95% CI, 1.01–1.56), at 100 mg caffeine. In contrast, four of the observational studies indicated a protective effect of caffeine at doses both above and below 400 mg/day. In a cross-sectional study of non-working adults, Smith (2009) found that all caffeine consumption was associated with a decreased risk of clinical depression, with the largest effect observed at levels greater than 260 mg/day (OR, 0.12; 95% CI, 0.1, 0.2). During 10 years of follow-up in a cohort of adult women who were free from depressive symptoms at baseline, no increased risk of clinical depression was observed in those with an average consumption of ≥550 mg caffeine/day compared to the lowest consumer group (<100 mg/day) (Lucas et al., 2011). In fact, Lucas et al. (2011) reported that depression risk decreased with increased caffeinated coffee consumption (≥550 mg/day compared to <100 mg/day). Pham et al. (2014) found that caffeine consumption up to >291 mg/day was associated with a decreased risk of depressive symptoms (OR, 0.57; 95% CI, 0.30–1.05) in an adult Japanese working population. Similar results were obtained when analyzing the effect of either coffee or green tea consumption individually. Ritchie et al. (2014) reported that consumption of ≥3 cups of caffeine a day (≥300 mg caffeine) in elderly individuals had no significant effect on prevalent or incident depression in adults; however, a nonsignificant trend for lower depression was noted in women. Lastly, caffeine has also been reported as a potential protective factor against suicide. Lucas et al. (2014) found that caffeine intake of 300 mg/day was associated with a pooled multivariate RR for suicide of 0.77 (95% CI, 0.63, 0.93). Similar results were obtained when analyzing increased coffee consumption, in which 2 cups of coffee per day resulted in a RR of suicide of 0.75 (95% CI, 0.63–0.90) (Lucas et al., 2014). The remaining three observational studies found no effect of caffeine consumption (ranging from 83 to >781 mg/day) on the risk of depression (Ritchie et al., 2014; Ruusunen et al., 2010; Trapp et al., 2014).

Taken together, the WoE suggests that the comparator of 400 mg/day of caffeine is acceptable intake which is not associated with significant concern regarding depression in adults. Of the 17 studies evaluating such, only one study reported the potential for adverse effects (at an intake below the comparator) – findings were very low in magnitude. In contrast, four studies suggested that the comparator is too conservative, as studies reported a lack of effects at intakes up to 1200 mg of caffeine (Bonnet and Arand, 2003a) (Fig. 7C). In addition, several observational studies (Lucas et al., 2011, 2014; Pham et al., 2014; Smith, 2009) indicate that consumption of caffeine may be protective for moods and behavior related to this endpoint (i.e., beneficial effect). There is a moderate to high level of confidence in the underlying data, primarily driven by the consistency in findings and low risk of bias of the individual studies.

3.3.1.1.4. Mood (anxiety, anger and confusion, and depression) in children and adolescents. Two controlled trials that studied mood in adolescents were identified (Fig. 7D). In the first trial, Temple et al. (2009) found that 2.32 mg/kg caffeine/day did not significantly increase irritability in adolescents aged 12–17 years compared to the placebo group. The second trial used a double-blind, placebo controlled design to directly investigate the effects of caffeine and levels of circulating steroid hormones in males and females, in which participants (aged 15–16 years) were administered 2 mg/kg on two separate occasions (Temple and Ziegler, 2011). One hour after administration, mood states were assessed using the POMS questionnaire and compared to baseline. Following Bonferroni correction, no significant changes were observed for any of the measurements of anger, confusion, anxiety, or depression (Temple and Ziegler, 2011).

Given the limited data, for all mood endpoints (anger, confusion, anxiety, depression) measured in children and adolescents, it was determined that there were insufficient data to develop refined conclusions regarding the potential effects of caffeine in populations other than healthy adults. However, the two studies discussed above indicated no effect of caffeine on mood parameters in adolescents.

3.3.1.2. Headache

3.3.1.2.1. Headache in adults. Studies reviewed for this endpoint included both those that measured headache directly following administration, as well as those that looked at the effects of caffeine consumption over time so that the data include both "acute" effects and potential "withdrawal" effects of caffeine. Ratings of headaches (pain or severity), which are often captured via customized questionnaires or a VAS, were not significantly increased in any of the 11 controlled trials that evaluated the effect of acute caffeine ingestion at doses ranging from 70 to 400 mg (Attwood et al., 2007; Haskell et al., 2008; Juliano et al., 2012; Liguori and Robinson, 2001; Pallarés et al., 2013; Rogers et al., 2008, 2010; Salinero et al., 2014; Sands et al., 2015; Stafford and Yeomans, 2005; Tinley et al., 2003). In two of these studies, in which caffeine was administered following a period of abstinence in regular consumers, caffeine significantly improved headache ratings (Juliano et al., 2012; Tinley et al., 2003) (see below for further discussion). Slightly higher doses of caffeine, however, may increase headache ratings, but only after some period. For example, Pallarés et al. (2013) studied adults participating in a weight-lifting protocol and assessed headache ratings the following day after administration of three different caffeine doses. Compared to a placebo, a slight increase in headache was seen with a 230-mg caffeine dose; higher single doses of approximately 459 mg and 689 mg caffeine/day increased the reports of headaches 24 h later, although it should be noted that this study did not fully describe the statistical significance (Pallarés et al., 2013).

Four observational studies (one cohort, one case-control, and two cross-sectional) that investigated the association between caffeine intake and headache were also identified. At the lower dose range of these studies, Rueda-Sanchez and Diaz-Martinez

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

(2008) conducted a cross-sectional survey and found no association between caffeine from daily coffee consumption and either episodic headache (NOEL of 129.2 mg/day) or chronic daily headache (CDH) (NOEL of 158.7 mg/day). The remaining three studies support a potential association between headache and moderate to high caffeine consumption. In a case-control analysis of subjects with and without CDH, Scher et al. (2004) found that prior consumption (i.e., consumption over the past year prior to development of CDH) of 311 mg caffeine/day, primarily from coffee, was associated with CDH (OR, 1.50; $P = 0.05$), but that current high caffeine consumption was not significantly associated with such. The authors concluded that these results both supported the "biological model of caffeine withdrawal" with respect to caffeine's relationship to headache yet indicated that this may not be true for all CDH sufferers (Scher et al., 2004). Furthermore, in a large-scale cross-sectional study of the effects of caffeine consumption and headaches in the general adult population, a significant increase in the prevalence of total headache (migraine plus nonmigrainous headache) was reported in the high-consumption group (>540 mg/day) compared to the low-consumption group (0–240 mg) (Hagen et al., 2009). When analyzing the prevalence of nonmigrainous headaches only, Hagen et al. (2009) reported a significant increase in headaches for those individuals consuming 241–400 mg caffeine/day compared to the group with a low consumption level. Headache frequency (either <7 days/month or 7–14 days/month) was also greater in the high-consumption group compared to the low-consumption group, although the authors cautioned that reverse causation may play a role (i.e., that some may consume caffeine for headache relief). In contrast, chronic headaches (i.e., >14 days/month) were not significantly associated with higher caffeine consumption (Hagen et al., 2009). Lastly, Ozsungur et al. (2009) included headache in one of three clusters describing caffeine withdrawal symptoms; the most sensitive of these was the "fatigue and headache," factor, which was significantly increased in the 100–200 mg/day group (OR, 1.97; 95% CI, 1.21–3.21).

For adults, the WoE support that consumption of ≤400 mg caffeine is not associated with an increase in headaches (Fig. 7E). However, like the evidence presented in Nawrot et al. (2003), observational studies do indicate a potential link between caffeine use and headache prevalence in some individuals, although some of this effect is likely due to withdrawal-related symptoms. In this regard, timing of the dose is important, since increases in reports of headache may only occur some significant time after ingestion (e.g., 12–24 h for habitual users) (Cappelletti et al., 2015). Although these studies were relatively consistent among themselves, withdrawal-related effects may be a factor in the outcomes of these observational studies, in addition to some residual confounding in these data due to reverse causation – factors making integration and interpretation of the findings quite complex. When effects were observed, the overall strength of association, however, was generally small (i.e., small magnitude). The confidence in the body of evidence is moderate to high.

3.3.1.2.2. Headache in children and adolescents. The effect of caffeine on headache in children and adolescents was assessed in two controlled trials, both of which suggest an effect of consumer status on this endpoint (Fig. 7D). In the first study, Heatherley et al. (2006) found that 1.3 mg/kg caffeine administered to children (aged 9–11 years) had no effect on headache ratings among those who were typically non- or low consumers (mean consumption of 12 mg/day); however, in regular consumers (mean consumption of 109 mg/day), caffeine reduced headache ratings compared to placebo. The authors suggest that these results indicate a reversal of the adverse effects, which may occur following overnight abstinence. In the second study, Temple et al. (2009) analyzed the effects of caffeine intake on irritability, hunger, and headaches in adolescents aged 12–17 years. Compared to the placebo group, changes in headache ratings did not reach statistical significance after consumption of 2.32 mg/kg caffeine/day; however, both male and female participants who were regular high-caffeine consumers (considered by the authors to be ≥ 50 mg/day) self-reported significantly more headaches than low consumers (Temple et al., 2009).

One observational study was also identified that evaluated the relationship between caffeine consumption and headache in children. Kristjansson et al. (2014) reported on the physical complaints (e.g., headache, problems sleeping, and low appetite) that were associated with the daily intake of cola and energy drinks in Icelandic children (aged 10–12 years). Girls appeared to be more sensitive than boys to the caffeine-related headaches and <0.6 mg/kg caffeine/day (less than one cola drink day) was associated with an increase in headaches (OR, 1.21; 95% CI, 1.05–1.41). For boys, significant increases in headaches were linked with consumption of more than one cola (OR, 1.29; 95% CI, 1.03–1.62) and less than one energy drink (OR, 1.87; 95% CI, 1.42–2.46) per day (i.e., >0.6 mg/kg to <1.4 mg/kg caffeine/day).

For children and adolescents, there was insufficient evidence to make conclusions regarding the appropriateness of the comparator. The limited evidence available, however, suggests that the comparator may be acceptable for headache; however, the data show that the relationship between headache and caffeine in children and adolescents is likely dependent on the timing of the dose and the subject's typical consumption.

3.3.1.3. Sleep. For purposes of discussion, the effects of caffeine on sleep have been divided by population as well as into subjective and objective categories because the types of endpoints evaluated by each metric vary (i.e., different endpoints of sleep). The subjective effects are those that looked at perceptions of "sleepiness" mood states such as fatigue, tiredness, drowsiness, or weariness. These endpoints are often measured with POMS or VAS questionnaires shortly following caffeine administration. In addition to these endpoints, we also discuss objective measures of sleep, such as sleep latency, duration, and efficiency, which are quantified the night(s) following caffeine intake.

3.3.1.3.1. Sleep in adults

3.3.1.3.1.1. Subjective effects

In total, there were 20 controlled trials that evaluated the subjective effect of caffeine on "sleepiness" mood states. Sixteen of these studies examined doses of caffeine ≤400 mg and found that caffeine either had no effect (Arciero and Ormsbee, 2009; Attwood et al., 2007; Harrell and Juliano, 2009; Liguori and Robinson, 2001; Sigmon et al., 2009; Sigmon and Griffiths, 2011; Tinley et al., 2003; Yoto et al., 2012) or may lead to improvements (i.e., decreases) in the ratings of sleepiness or fatigue (Alsene et al., 2003; Haskell et al., 2008; Juliano et al., 2012; Killen et al., 2013; Marczinski et al., 2014; Rogers et al., 2013; Salinero et al., 2014; Stafford and Yeomans, 2005) when measured either immediately following consumption or the day afterward (Fig. 7F). Furthermore, a single study with a slightly higher cumulative daily dose, approximately 420 mg caffeine (1.75 mg/kg body weight, three times daily), showed no effect on POMS ratings of fatigue (James and Gregg, 2004b). In contrast to the aforementioned studies, there were just three controlled trials that found that caffeine intake was associated with increased ratings of "sleepiness." One was a lower dose study, investigating the effect of caffeine in vehicle on mood. Like the effects on other mood endpoints discussed above, Young and Benton (2013) found that although 80 mg caffeine delivered with water led to a short-term boost in energetic feelings, tiredness ratings over the course of a proceeding test session were significantly increased (along with other adverse mood ratings) when

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

caffeine was given with water alone, compared to the same amount of caffeine in yogurt or with glucose. The second study specifically associated fatigue with withdrawal side effects, stating that one of the defining elements of this state, the "fatigue and headache" factor, was significantly increased in the 100–200 mg/day group (OR, 1.97; CI, 1.21, 3.21) (Ozsungur et al., 2009). The remaining clinical trial indicated that prolonged high-dose consumption may also negatively affect ratings of fatigue over time. Bonnet and Arand (2003a,b) conducted a 7-day protocol in which adults who were moderate caffeine consumers were administered 1200 mg caffeine/day (400 mg, three times daily). After the week was complete, subjects reported significantly higher levels of fatigue than at baseline, likely due to a negative impact of caffeine on nighttime sleep observed throughout the study (Bonnet and Arand, 2003a). One observational study was also identified that examined the effect of caffeine on sleepiness. The US Centers for Disease Control and Prevention (CDC) found that adult US service members who consumed ≥240 mg caffeine/day (three or more energy drinks) were more likely to experience daytime sleepiness than individuals who drank zero or one to two energy drinks per day (CDC, 2012).

Based on the studies reviewed, the majority demonstrate that the comparator of 400 mg caffeine/day is acceptable as an intake generally not associated with concern regarding adverse effects on sleep. There were a few cases in which prolonged dosing was associated with increased fatigue; the magnitude of these changes was difficult to assess. Caffeine's mode of action in the central nervous system (CNS) helps, in part, explain why most caffeine doses tested in these studies may indeed provide some benefit on this endpoint by reducing perceived fatigue; however, higher doses may disrupt sleep and lead to an increase in fatigue when consumed over the course of several days (see Bonnet and Arand, 2003a, and Section 3.3.1.3.1.2 on objective effects).

3.3.1.3.1.2. Objective effects

Fifteen controlled studies evaluated objective aspects of sleep itself, often measured using polysomnography and/or actigraphy. Many of these were designed such that the dose of caffeine was administered within several hours prior to bedtime, sometimes with the deliberate intention of altering sleep (*e.g.*, Keenan et al., 2014). As such, 11 of the 15 controlled trials using caffeine doses ≤400 mg reported that caffeine interfered with some aspect of sleep quality, including nocturnal activity, sleep efficiency (percentage of total time in bed spent in sleep), and sleep-onset latency (Bonnet and Arand, 2003b; Carrier et al., 2007; Drake et al., 2006, 2013; Drapeau et al., 2006; Judice et al., 2013; Keenan et al., 2014; Omvik et al., 2007; Paterson et al., 2009, 2007; Salinero et al., 2014). Only 1 of the 15 studies demonstrated no effect of a lower dose of caffeine on sleep when taken in the evening. Ho and Chung (2013) found that consumption of at least 1 cup of coffee (60–120 mg caffeine) within 6 h of bed time had no significant effect on sleep latency, efficiency, or quality. Drake et al. (2006) reported that 240 mg caffeine/day administered to adults 1 h prior to bedtime resulted in an increase in the "latency to persistent sleep" in individuals who were previously identified as being highly susceptible to sleep disturbance; however, no significant adverse effect on sleep was seen in individuals with a low susceptibility to sleep disturbance.

The remaining three controlled studies investigated the effects of doses of caffeine >400 mg/day on objective measures of sleep and the results were split. After (on average) 6 h following the administration of 480 mg caffeine to a sample of male athletes with low caffeine consumption, Pontifex et al. (2010) reported no change from baseline levels for sleep duration, wake time, number of awakenings, or sleep efficiency. Alternatively, Pallarés et al. (2013) found that a similar dose of caffeine (6 mg/kg or approximately 459 mg/day) administered prior to a weight-lifting protocol increased reports of sleep problems in subjects the following day, compared to placebo, although no effects were seen at the 230-mg/day dose. Furthermore, (Bonnet and Arand, 2003b) found that a cumulative dose of 1200 mg caffeine/day (400 mg, three times daily) given for 7 days can induce symptoms of insomnia, including reduced sleep quality and increased latency of sleep onset, in healthy subjects. The authors expected that adverse effects would occur, given the protocol and that the decrement in mood could be related to the sleep disruption caused by physiological arousal.

Three observational studies were identified that provided inconsistent conclusions relative to the results from clinical studies above. Jaussent et al. (2011) found no observable effect on insomnia symptoms in male and female participants who self-reported a daily consumption of ≥190 mg caffeine/day. As described above, however, the CDC (2012) noted that adult US service members who consumed ≥240 mg caffeine/day (three or more energy drinks) experienced more sleep disruptions and were more likely to have reduced sleep than individuals who drank zero or one to two energy drinks per day. This study was conducted with soldiers who, due to the demands of a combat environment, had limited opportunity for sleep and may have purposely been using caffeine to remain awake and alert (IOM, 2001; McLellan et al., 2016). Interestingly, the observational study by the CDC (2012) indicated that sleep disruption may occur at much higher levels in habitual caffeine consumers. A cohort study of French workers found that 680 mg caffeine/day was associated with a decreased sleep duration, but the investigators found no effect following consumption of 595 mg caffeine/day (Sanchez-Ortuno et al., 2005).

With respect to the data obtained via objective measures of sleep in adults, results indicate that 400 mg caffeine/day is likely too high as an intake, in that it would be expected to disrupt sleep when administered with the intention to do so. Specifically, ingestion of caffeine even at doses below the comparator can lead to delayed sleep onset and decreases in sleep quality and efficiency, but this is particularly the case when caffeine is consumed near bedtime. Overall, caffeine at doses both above and below the comparator may provide short-term benefits to improve perceived fatigue but, depending on the dose and timing, may also disrupt sleep, leading to increased fatigue the following day.

3.3.1.3.2. Sleep in children and adolescents. The adverse effects of caffeine on sleep in children and adolescents have been frequently investigated as associations between the two (see below for further discussion of association studies), yet we identified only one clinical study and two observational studies that contained quantitative data and were thus included (Fig. 7D). Temple and Ziegler (2011) (described above) found no change in ratings of fatigue in adolescents following a dose of 2 mg/kg caffeine. As for the observational studies, Kristjansson et al. (2014) (also described above) found that sleeping problems in boys (aged 10–12 years) were associated with caffeine intake as low as <0.6 mg/kg/day (OR, 1.21; 95% CI, 1.03–1.42), whereas girls (aged 10–12 years) seemed less sensitive to the sleep-disrupting effects that occurred between 0.6 mg/kg caffeine/day (OR, 1.55; 95% CI, 1.21–1.98) and 1.4 mg/kg caffeine/day (OR, 1.56; 95% CI, 1.07–2.25). In the other observational study, Calamaro et al. (2009) sought a link between total length of sleep on school nights and caffeine intake among adolescents aged 12–18 years. Those who slept 8–10 h consumed a median of 54.1 mg caffeine/day, whereas those with the least sleep (3–5 h) consumed a median of 157.6 mg/day (2.77 mg/kg) (Calamaro et al., 2009). Calamaro et al. (2009) noted the increased caffeine consumption, but it was not statistically significant ($P = 0.067$). Like the other endpoints, there are insufficient data to evaluate with confidence the effect of caffeine dose on sleep in children and adolescent populations. However, based on the limited data, it appears that similar to adults, considerations such as timing and

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

duration of dose are likely to be important for these populations as well.

3.3.1.4. Problematic and risk-taking behavior in adults. Endpoints in this category include substance abuse, alcohol consumption, and cigarette smoking, in addition to violent or disorderly behavior. Two studies, including one controlled trial and one observational study, were identified that addressed risk-taking behavior in (mostly young) adults, but there were no studies in younger populations. Using a computer game that simulates risk taking (Balloon Analogue Risk Task), Peacock et al. (2013) found that consumption of one 250-mL energy drink (containing 80 mg caffeine) significantly increased risk-taking behavior in young adults, although the magnitude of effect was small. In the observational study, Krall et al. (2002) found that consumption of >6 cups of coffee per day (>570 mg caffeine) was a predictor of an increased risk of smoking relapse in a multivariable analysis of men who were former smokers. The available data for adults were considered to be too limited to make a refined conclusion relative to the comparators for this endpoint.

3.3.2. Body of evidence assessment

The initial confidence in the body of evidence (OHAT, 2015a) is moderate to high because most included studies involved measured and/or controlled exposures with adequate data presented on the endpoints of interest (Table 1). The low risk of bias scores (Fig. 8) and low level of indirectness increase confidence in the overall body of evidence, although this is much more so the case for the data on adults than for children and adolescents. Of the 80 studies reviewed and graded for the SR, 96% (77 of 80) were associated with a definitely or probably low risk of bias. The magnitudes of effects, when apparent, were typically small except for those related to the impact of caffeine on sleep, which could be larger. Across the body of evidence, most of the endpoints were consistent in observations, apart from anxiety, which was associated with both effects and a lack of effects below the comparator (a trend that was in line with previous literature, according to many of the original study authors). Only 12 of the included studies evaluated dose response (half of which showed evidence of a gradient effect on the endpoint), whereas most studies were controlled trial designs that administered only a single dose, thus limiting the ability to evaluate dose-response relationships.

Overall, the Nawrot et al. (2003) conclusions for the safe levels of daily caffeine intake for healthy adults are generally supported by this body of evidence. There are sufficient data from RCTs that lower doses of caffeine may negatively affect some aspects of behavior (particularly anxiety) and sleep for adults; however, these changes are often low in magnitude and/or are more apparent in sensitive subpopulations.

As Nawrot et al. (2003) predicted, more evidence has become available addressing how individual differences (namely, polymorphisms of the ADORA2A receptor and their relationship to anxiety) may explain some interindividual variability in sensitivity to caffeine's effects. Furthermore, caffeine's ability to disrupt objective measures of sleep when administered later in the evening, or shortly before bedtime, may not reflect common consumer behavior or is often self-limiting (Nawrot et al., 2003; Penolazzi et al., 2012). Otherwise, there was little to no evidence identified to suggest that <400 mg caffeine/day has any negative effects on mood states, outside of effects on anxiety in sensitive subpopulations. In contrast, caffeine consumption appears to provide some benefit with regard to fatigue- and depression-related endpoints.

The body of evidence is inconsistent with regard to the effect of caffeine consumption below the comparator and its effect on headache. Some of the mixed results may be linked to symptoms of withdrawal and consumer status. The confidence in this body of evidence for this endpoint is moderately high, because most of the data are from RCTs that directly assessed the relationship between caffeine and headache (Table 1) and all studies but one were associated with a low risk of bias (Fig. 8). This factors in some inconsistency in the dataset, driven by the difference in results between the observational studies and the RCTs.

Lastly, there is sparse evidence that caffeine is associated with an increase in risk-taking behavior in adults. This latter effect is a research area that has seemingly attracted more attention since the work by Nawrot et al. (2003) was published, particularly for younger consumers (see below).

For adolescents and children, Nawrot et al. (2003) concluded "… it is unknown if long-term daily consumption of caffeine would produce effects similar to those observed in the studies reviewed above. However, it is known that the human nervous system (including the brain) continues to develop and mature throughout childhood. It is possible that the protracted development of the nervous system may render children more sensitive to any adverse effects of caffeine." The available literature for children and adolescents included in this SR was scant, but the higher-quality studies suggest no major adverse effects on the observed endpoints at doses near or less than 2.5 mg/kg. Overall, the body of literature reviewed for children and adolescents was generally of lower quality compared to the data for adults, due to issues of study design, indirectness, and potential for reverse causation (Table 1). For children and adolescent populations, there was not enough information, high quality or otherwise, to fully evaluate the appropriateness of the comparator. More targeted research is required to identify sensitive subpopulations in these younger groups, to better quantify the levels at which adverse behavioral effects are observed as well as to better understand the link between caffeine consumption and adverse effects (e.g., sleep and risk-taking behavior).

The SR of 80 studies provided evidence to evaluate potential impacts of the consumption of 400 mg caffeine/day on the behavior outcome, including assessment of mood (comprising anxiety and other mood states), headache, sleep, withdrawal, and risk-taking behavior. When the weight of evidence was considered, the comparator, 400 mg caffeine/day, was found to be an acceptable intake that is not be associated with significant concern for adverse behavioral effects in adults. However, intake below the comparator may affect some sensitive individuals who are prone to anxiety or sleep disruption. Often, observed effects below the comparator (e.g., anxiety) were limited to subgroups or timing of dose (e.g., sleep), whereas others were complicated by consumer status (e.g., headache and fatigue). For some endpoints depression, headache, sleep [subjective], and anger/confusion there was largely a lack of effects reported, and in some cases, data suggested that intakes higher than the comparator were without effect. There is a moderate to high level of confidence in the body of evidence supporting this conclusion. Confidence was increased by the overall low risk of bias and low level of indirectness; although the variability introduced by sensitive subpopulations and consumer status were key limitations that precluded a higher level of confidence.

Of the 80 included studies, the data in children and adolescents were limited to just 5 studies, which together evaluated mood, headache, and sleep. As such, it was determined that the evidence base was insufficient to render a conclusion regarding appropriateness of the comparator (2.5 mg caffeine/day) for potential impacts of caffeine consumption on behavior outcomes in these populations.

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

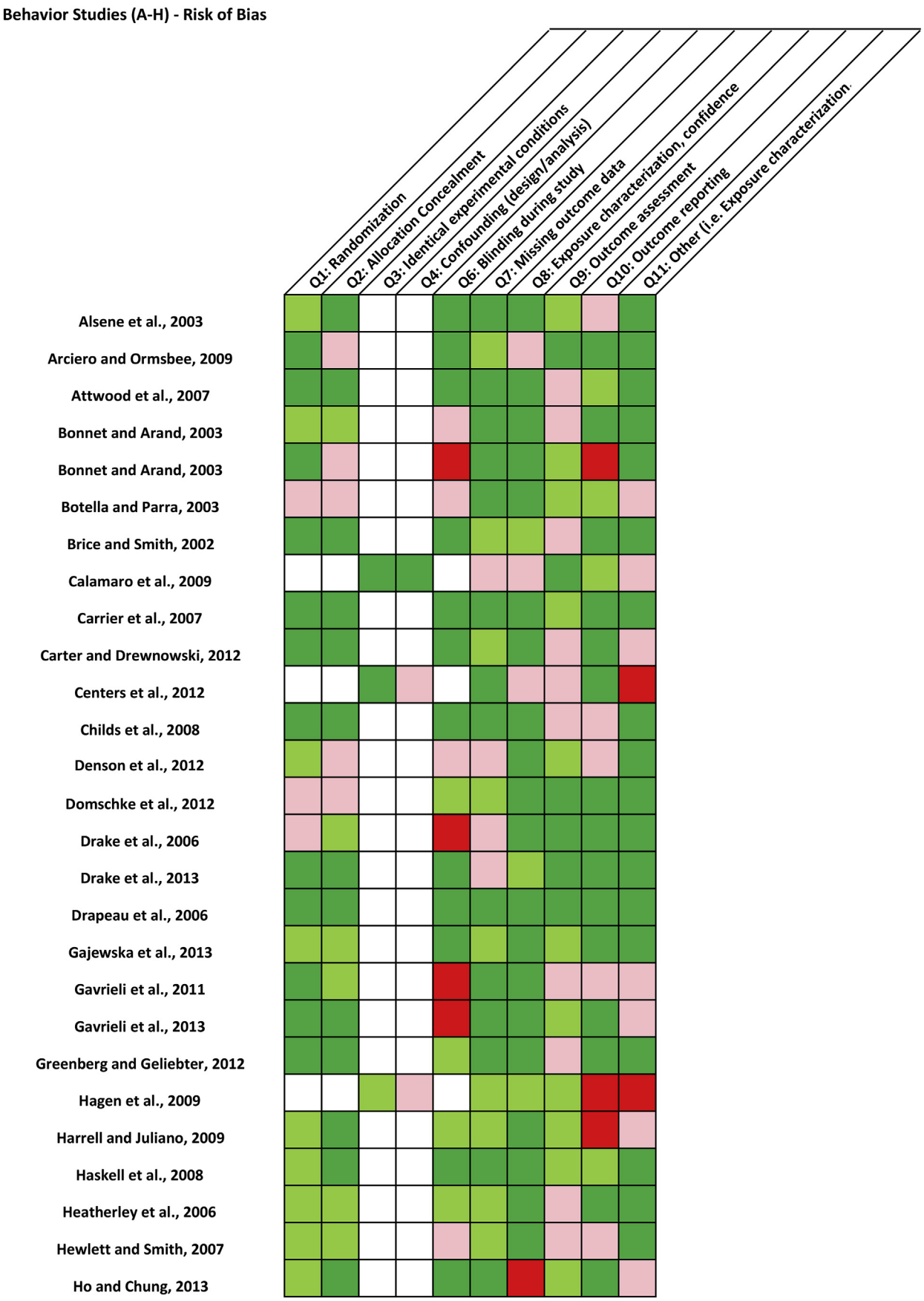


Fig. 8. Risk of bias (RoB) heat map for studies included in the behavior outcome. The domain-based validity was evaluated based on study type per the OHAT (2015b) RoB tool. RoB for each domain is indicated by color: "definitely low risk of bias" (dark green, +2), "probably low risk of bias" (light green, +1), "probably high risk of bias" (light red, −1), and "definitely high risk of bias" (dark red, +2). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

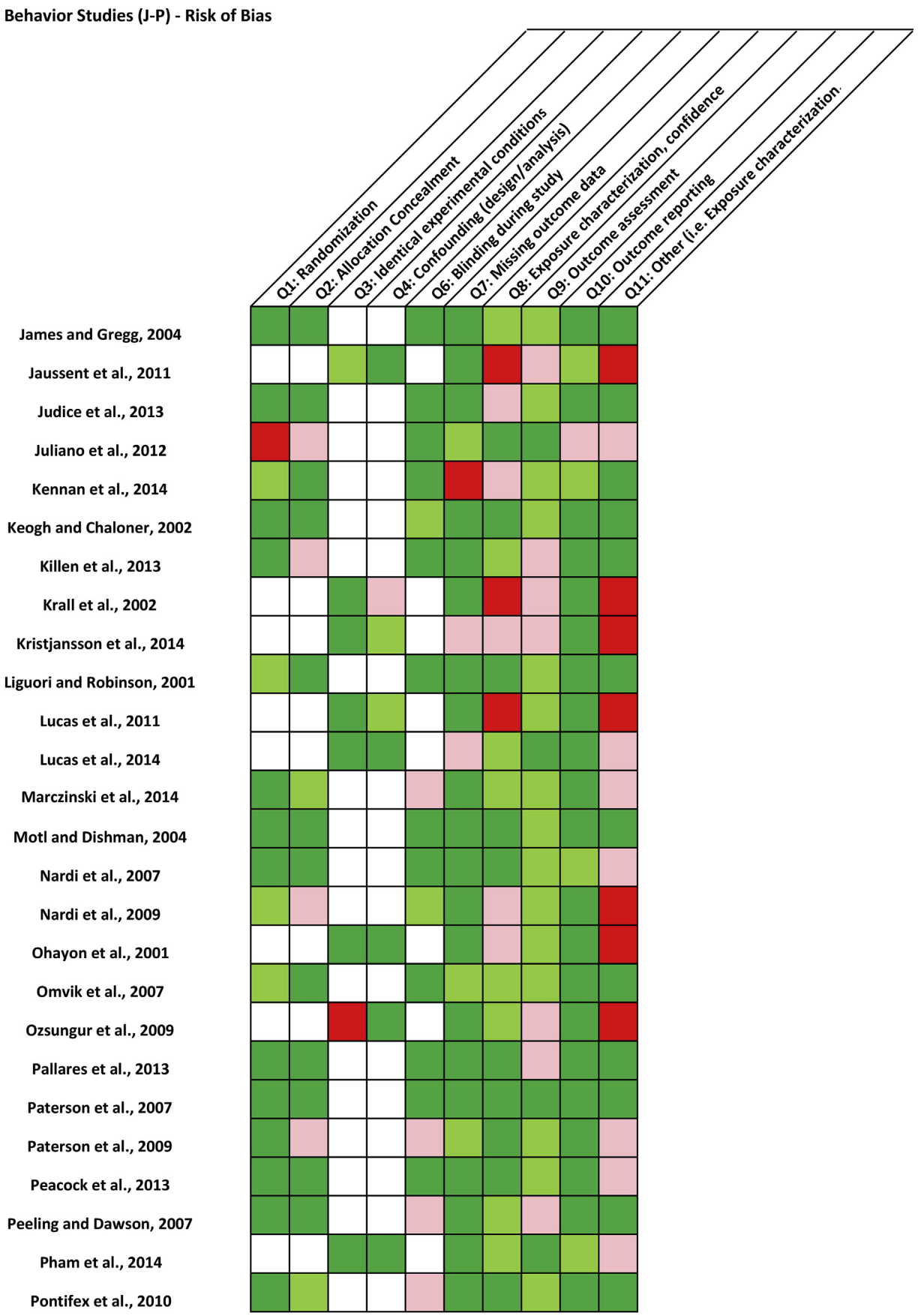


Fig. 8. (*continued*).

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

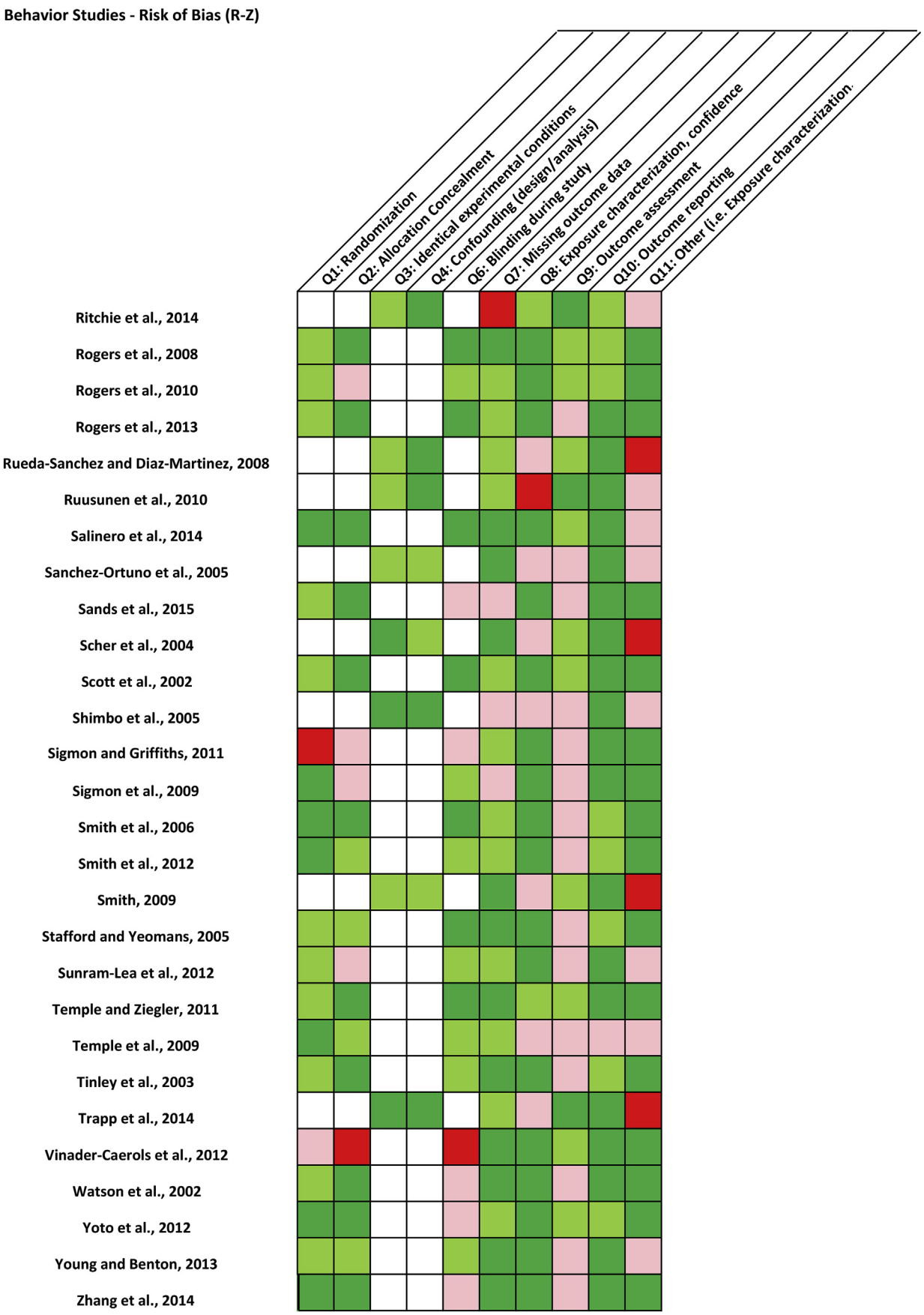


Fig. 8. (*continued*).

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

3.4. Reproduction and development

We reviewed 94 full text papers evaluating reproductive and developmental effects (Fig. 2). A total of 58 these studies were considered to meet the criteria for inclusion in the SR as they permitted comparison to Nawrot et al. (2003) conclusions. Of the 36 studies that were excluded, 26 studies contained information regarding potential associations between caffeine consumption and reproductive or developmental endpoints but did not provide quantitative data that could be compared. With respect to the PECO, all of the studies included involved adult populations. The majority of studies involved exposures in pregnant women, for which the Nawrot et al. (2003) comparator of <300 mg/day was applied. For the few studies evaluating nonpregnant women (e.g., studies evaluating fecundity or age at menopause) or men (e.g., studies evaluating sperm quality), the Nawrot et al. (2003) comparator of <400 mg/day for a healthy adult was applied.

Only 3 of the 58 included studies were randomized clinical trials; the remaining 55 were observational studies, primarily cohort and case-control studies. Exposures in the observational studies were characterized using self-reported methods (e.g., food frequency questionnaires) in all of the clinical studies but one and were based on consumption of coffee, soda, and tea in most studies. Chocolate was also included in a number of studies, whereas caffeine-containing medications and energy drinks were evaluated only in a few. Most studies evaluated the amount of caffeine in these substances as part of the analysis; in the few studies that did not, the amount of caffeine was calculated by the SR authors per the methods (Section 2).

Many of these studies were conducted in participants from large cohorts such as the Danish National Birth Cohort (Bech et al., 2005, 2006, 2015) and the National Birth Defects Prevention Study (Browne et al., 2007, 2011; Chen et al., 2012; Collier et al., 2009; Hoyt et al., 2014; Kancherla et al., 2014; Miller et al., 2009; Schmidt et al., 2009, 2010). Common variables accounted for in such analyses included maternal characteristics such as race, age, weight, BMI, smoking (some using cotinine as a marker), and alcohol consumption. Other factors more specific to endpoints of concern were also considered, such as history of previous pregnancy or miscarriage, partner characteristics, family history of condition, gestational age at birth, and maternal nutrient and supplement intake. Some studies included changes in caffeine consumption during pregnancy as a variable, although most studies did not. However, nausea was evaluated as a confounder in most studies included in the SR that investigated fetal/neonatal demise (spontaneous abortion, recurrent miscarriage, stillbirth), although the extent to which information was collected and incorporated varied. Controlling for symptoms of the so-called “pregnancy signal,” such as nausea, aversion to smells or tastes, and vomiting, has been considered critical for robustly assessing these endpoints, since these correlates of pregnancy health status may also influence caffeine intake (Lawson et al., 2004; Stein and Susser, 1991). Without specific analyses of caffeine aversion, it is difficult to ascertain whether an increased incidence of spontaneous abortion in a study is due to higher caffeine consumption, or if reduced caffeine consumption is being observed in healthier pregnancies due to the pregnancy signal (i.e., reverse causation).

With respect to the comparator of <300 or <400 mg/day from Nawrot et al. (2003), the majority of the data points, regardless of findings, were below these levels (Fig. 9). The majority of analyses were conducted using categorical exposure groupings (e.g., <1 cup/day, 1–3 cups/day, and >3 cups/day); such data were standardized either as part of the authors' analysis or, in a few cases, per the SR methods. About half of the studies evaluated 300 mg/day (or just above) as the highest intake level, whereas <10% evaluated upper intake levels below the 300 mg/day comparator. The remaining third of the included studies evaluated intake levels higher than 300 mg/day, with the highest evaluated in a single study was >1000 mg/day. Most studies were designed specifically to evaluate caffeine (typically via conversion of self-reported consumption of cups of a caffeine-containing substance per day) and thus were considered to have a low level of indirectness.

Endpoints characterizing the reproductive and developmental toxicity outcome included fecundability and infertility, spontaneous abortion, recurrent miscarriage, stillbirth (including late spontaneous abortion), preterm birth, fetal growth (including small for gestational age [SGA]/intrauterine growth restriction [IUGR]), birth defects, childhood behavior, childhood cancer, markers of maternal stress, pregnancy-induced hypertension and/or pre-eclampsia, and age at menopause (Fig. 9; Table 1). The endpoints with the largest number of studies (e.g., spontaneous abortion, fetal growth) were also in Nawrot et al. (2003). However, some of the endpoints reviewed herein, such as childhood cancers or childhood behavior associated with prenatal caffeine exposure, were not reviewed by Nawrot et al. (2003). Conversely, this effort did not identify studies evaluating the risk of sudden infant death syndrome.

3.4.1. Summary of individual studies by endpoint

Data for two general endpoints, fecundability and reproductive measures (male and female), were evaluated in healthy adults using the Nawrot et al. (2003) comparator of 400 mg/day (Fig. 9A). The remaining endpoints within this outcome were associated with caffeine exposures in pregnant women and thus were evaluated relative to the Nawrot et al. (2003) comparator of 300 mg/day (Fig. 9, B and C).

3.4.1.1. Fecundability, fertility, and male reproductive measures. Fecundability (the ability to conceive during a given menstrual cycle) was evaluated in two studies meeting the criteria for inclusion in the SR; in these studies, no association between increased time to pregnancy and female caffeine intakes at levels of ≥300 mg/day was found (Hatch et al., 2012; Taylor et al., 2011). Related to fertility, one study reported a lack of effects for ovulatory disorder infertility at caffeine consumption of ≥333 mg/day (Chavarro et al., 2009).

Four studies were included that evaluated male reproductive parameters. Two of these studies reported a lack of effects at levels above the 400 mg/day comparator. Jensen et al. (2010) reported a lack of effects observed on reduced sperm concentration or total sperm count at intake levels of >570 mg/day. Sobreiro et al. (2005) reported a lack of effects on semen quality (as measured by semen volume, sperm concentration, total sperm count, and percent motile sperm and morphologically normal forms) following consumption of >800 mg/day. With regard to the other two studies that evaluated exposures below the 400 mg/day comparator, Paton et al. (2010) reported no effects on salivary testosterone following up to 240 mg caffeine via chewing gum during repeated sprint athletic performance. Schmid et al. (2007) reported that men consuming >308 mg/day had significantly higher (~20%) frequencies of sperm with DNA damage, as measured under neutral, but not alkaline, conditions compared to men with less caffeine consumption.

Taken together, there is a moderate to high level of confidence that the comparator of 400 mg/day in healthy adults is acceptable for fertility, fecundability, and male reproductive endpoints (Table 2). Confidence is increased by the consistency of findings (lack of effects above and below the comparator), low risk of bias, and low level of indirectness.

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

3.4.1.2. Spontaneous abortion. Spontaneous abortion during early pregnancy was evaluated in eight observational studies reviewed in this SR; the gestational ages included in the studies varied but generally were included in the first trimester and up to 20 weeks gestation. In four of these studies, no increased risk of spontaneous abortion was identified at maternal caffeine intake levels at or below the comparator of 300 mg/day (Giannelli et al., 2003; Karypidis et al., 2006; Maconochie et al., 2007; Rasche, 2003). One of these studies (Maconochie et al., 2007) further noted that no effects were observed at levels of >500 mg/day (Fig. 9B). In a fifth study, maternal caffeine consumption was found to not be associated with risk of miscarriage when intakes among caffeine consumers were compared above the 75th percentiles of >463.1 mg/day (approximately 4 weeks gestation) and >273.2 mg/day (approximately 16 weeks gestation) versus nonusers (Savitz et al., 2008). In two case-control studies, daily maternal caffeine consumption above the comparator (>300 and ≥ 375 mg/day) was found to be significantly associated with risk of spontaneous abortion (Giannelli et al., 2003; Rasche, 2003). Of note, these studies attempted to account for the pregnancy signal. Of the eight studies evaluating spontaneous abortion, all but Rasche (2003) and Wen et al. (2001) controlled for nausea and/or vomiting.

A significant association was observed between maternal caffeine intake at levels >500 mg/day during early pregnancy and spontaneous abortion in women with the CYP1B1 432 Val/Val genotype (Karypidis et al., 2006). These effects were not seen in women with the Leu/Leu or Leu/Val genotypes, and analysis of all genotypes combined was not performed.

Two prospective studies reviewed in the SR reported that maternal caffeine consumption during the first trimester increased

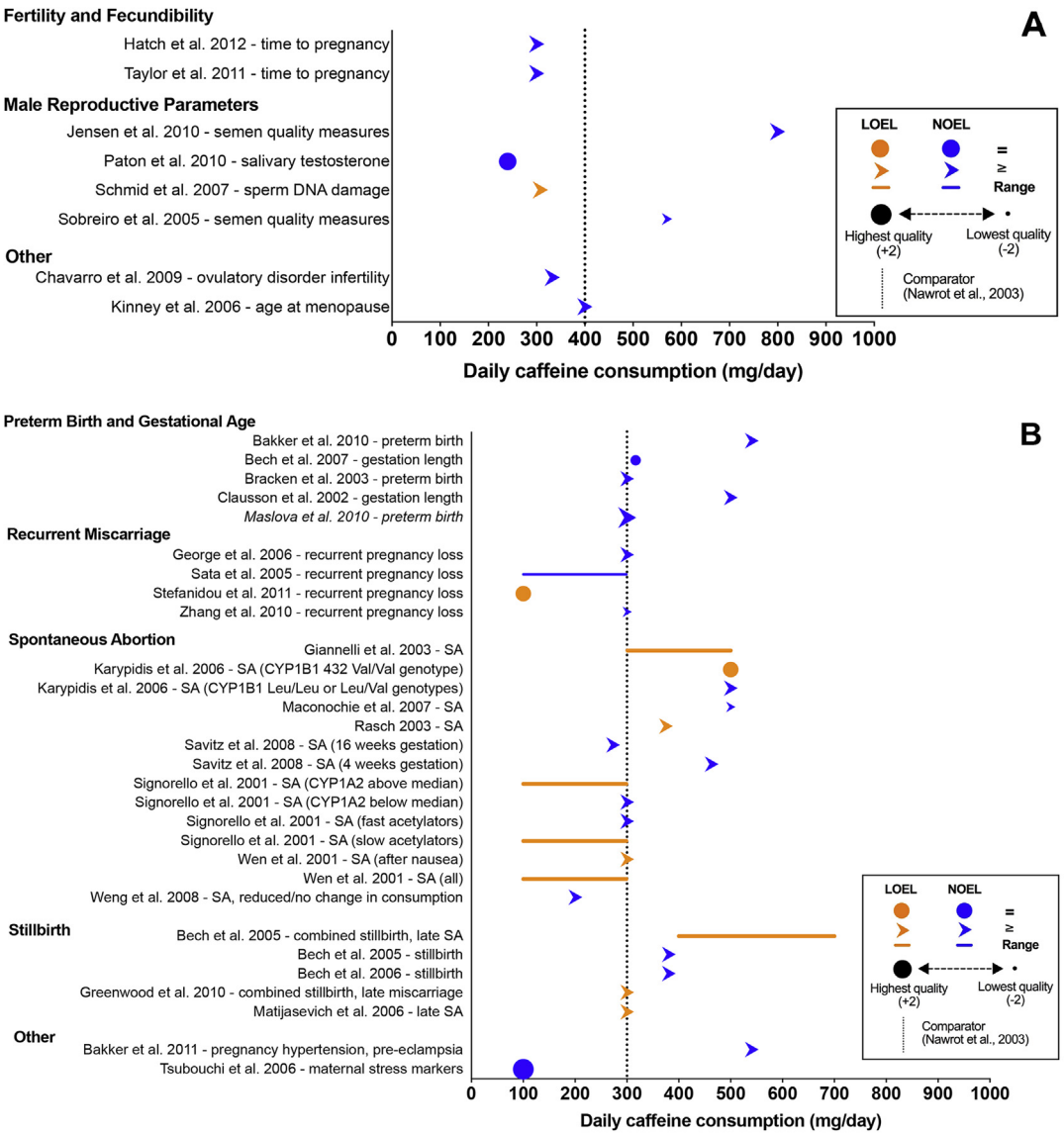


Fig. 9. (A–C) Summary diagram of exposure-response data relative to the comparator for the reproductive and developmental outcome: (A) fertility and fecundability, male reproductive parameters, other; (B) other, preterm birth, recurrent miscarriage, spontaneous abortion, stillbirth; and (C) birth defects, childhood behavior, childhood cancer, fetal growth. Symbols represent caffeine intake (mg/day) as reported by original study authors. The color of the symbol indicates the type of effect; no effect (NOEL; blue symbols) or the lowest effect level (LOEL; orange symbols). The shape of the symbol represents the type of metric (circles represent a discrete value, arrowheads represent greater than or equal to a value, and a horizontal line represents a range of values; metrics based on that reported by original study authors). The size of the symbol indicates the overall risk of bias (larger symbols indicate a lower risk of bias, or higher methodological quality). The dashed vertical line marks the comparator value. Italicized study names indicate a meta-analysis. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

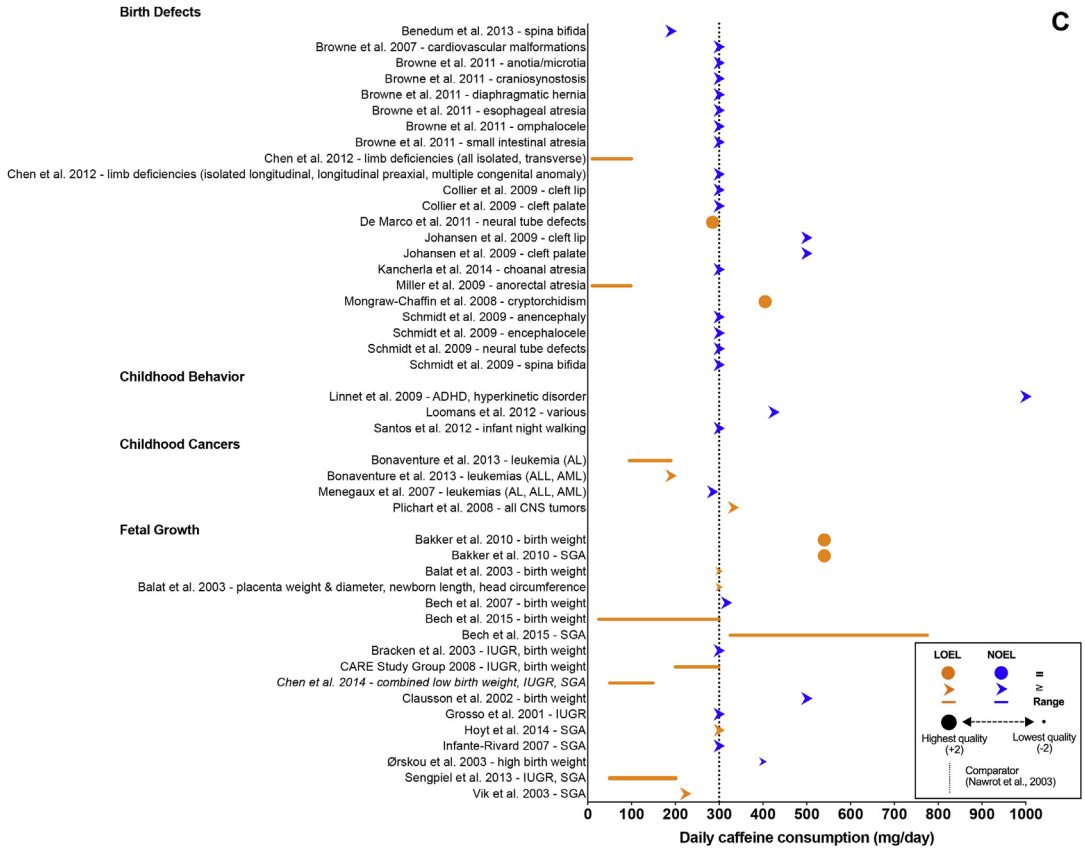


Fig. 9. (*continued*).

the risk for spontaneous abortion for intakes below the comparator (Wen et al., 2001; Weng et al., 2008). In the study by Wen et al. (2001), consumption of 100–299 mg/day was found to be associated with an increased risk (RR, 2.0; 95% CI, 1.0–4.1) compared to <20 mg/day; however, for mothers reporting nausea, an association was found only after nausea had started and only at levels ≥300 mg/day (RR, 5.4; 95% CI, 2.0–14.6). Weng et al. (2008) found a significant association at intakes ≥200 mg/day, the highest consumption category evaluated (HR, 2.23; 95% CI, 1.34, 3.69), compared to women with no caffeine intake. However, when analysis was restricted to women who reduced caffeine intake during pregnancy, this association was no longer significant (≥200 mg/day; HR, 1.47; 95% CI, 0.87–2.51). In a separate study, Signorello et al. (2001) did not conduct an analysis on the study population overall, but rather only analyzed data stratified by indices of caffeine metabolism. Mean daily caffeine intake was found to be significantly associated with an increased risk of spontaneous abortion in pregnant women with CYP1A2 activity above the median (OR, 2.42; 95% CI, 1.01, 5.80). Conversely, pregnant women with CYP1A2 activity below the median had reduced risk, with ORs of 0.32 and 0.46 for these same caffeine intake levels (100–299 mg/day). Slow acetylators had a significant increase in risk of spontaneous abortion at intakes of 100–299 mg/day (OR, 2.38; CI, 1.04–5.49); this effect was not observed at the higher dose level (OR, 1.65; CI, 0.67–4.06).

Considering the totality of the evidence, there is a moderate level of confidence in the body of evidence that the comparator of 300 mg/day is acceptable as an intake that would not be associated with significant concern of spontaneous abortion. Several studies documented effects at lower levels of consumption; thus, some data indicate that the comparator may be too high for certain subpopulations/phenotypes, although other studies reported lack of effect following consumption >500 mg/day. Confidence in the overall body of evidence on spontaneous abortion is significantly limited by the inability to fully accommodate for the pregnancy signal. Confidence is also decreased by the lack of consistency in the direction of findings both above and below the comparator. An overall low risk of bias (Fig. 10) and a low level of indirectness increase the overall confidence of the studies included on spontaneous abortion. Consideration of the magnitude of effect neither strengthened nor lessened confidence in the body of evidence given that in the two studies reporting effects below the comparator, the magnitude was low (OR or RR between 2.0 and 2.4), and for one study effects were only found in subgroups. The overall level of confidence in the body of evidence which supports the conclusion remained moderate, primarily due to the strong potential confounding effects of the pregnancy signal and inconsistency of findings.

3.4.1.3. Recurrent miscarriage. Four case-control studies that evaluated recurrent miscarriage (defined as at least two or more pregnancy losses) were included in the SR. Three of these studies reported that maternal caffeine consumption was not associated with overall incidence of recurrent miscarriage at intake levels at or below the comparator of 300 mg/day (George et al., 2006; Sata et al., 2005; Zhang et al., 2010) (Fig. 9B). Although no significant findings were observed overall, some effects were noted in subgroup analyses from these three studies. For example, Sata et al. (2005) noted that, while no association was found in subjects with other genotypes of CYP1A2 studied at levels ≥300 mg/day,

women who were homozygous for CYP1A2*1F were at higher risk (OR, 5.23; 95% CI, 1.05–25.9). In the fourth study, Stefanidou et al. (2011) reported that maternal consumption of 151–300.9 mg/day was associated with recurrent miscarriage based on an unadjusted analysis (OR, 3.05; 95% CI, 1.28–7.29). Adjusted analyses evaluated caffeine intake as a continuous measure rather than within consumption categories and indicated that the odds of unexplained recurrent miscarriage were 2.72 (95% CI 2.72–2.73, rounded) times greater for cases compared to controls for each 100 mg/day of caffeine consumed. However, the reporting of caffeine consumption late after miscarriage greatly increased the potential for recall bias, and the analyses were not adjusted for the pregnancy signal (which was evaluated as presence of nausea/vomiting by the authors but not controlled in these analyses).

The evidence from these four studies provides a moderate level of confidence that ≤300 mg/day is an acceptable intake that would not be associated with significant concerns for endpoints related to recurrent miscarriage in healthy pregnant women. Confidence is increased by the overall low level of indirectness (Table 1), low risk of bias (Fig. 10), and reasonable consistency in the lack of findings.

3.4.1.4. Stillbirth. Four studies included in the SR evaluated stillbirth, which in some cases combined stillbirth with late spontaneous abortion. No effects were reported in these studies below the 300 mg/day comparator; however, two studies reported increased risk at ≥300 mg/day (Fig. 9B). In a large study of prospective data from the Danish National Birth Cohort, no increased risk of fetal death (stillbirth and late spontaneous abortion combined) or stillbirth was found with maternal caffeine intake levels equivalent to the comparator (Bech et al., 2005). Above the comparator, an effect was observed at 400–700 mg/day based on analysis of fetal death (all deaths after 20 weeks gestation) (Bech et al., 2005). When stillbirth (all causes) was analyzed separately in this same study, no effects were observed following consumption of ≥380 mg/day. When stillbirths were further analyzed by attributed cause, only stillbirth due to placental dysfunction was significantly associated with consumption ≥380 mg/day (Bech et al., 2005). A nested case-control study of this same cohort reported similar findings, with no effects for stillbirth following consumption of ≥380 mg/day (Bech et al., 2006). Mean caffeine intake up to 299 mg/day throughout pregnancy was not significantly associated with fetal death (>20 weeks gestation) in a case-control study of an Uruguayan population; intake of ≥300 mg/day showed a statistically significant increased risk (OR, 2.33; 95% CI, 1.23, 4.41) (Matijasevich et al., 2006). Greenwood et al. (2010) reported that maternal caffeine intake during the first trimester was significantly associated with increased stillbirth and late miscarriage (defined as loss between 12 and 24 weeks) at ≥300 mg/day; the adjusted OR was 5.1 (95% CI, 1.6–16.4). The wide CIs are likely attributed to the small number of events in this study (28 fetal deaths out of a population of 2635); this is in contrast to the larger study by Bech et al. (2005) (no increased risk of fetal death below the comparator), which consisted of 1102 fetal deaths and a total population of 88,842.

These four studies provide a moderate level of confidence to support a conclusion that comparator of ≤300 mg/day is an acceptable intake that would not be associated with significant concern for endpoints related to stillbirth in healthy pregnant women. The overall low risk of bias and high level of indirectness increase the overall confidence in the evidence base. When effects were observed (intakes of 300 mg/day or more), the magnitude of effects ranged from <2 to >5 (Table 1).

3.4.1.5. Preterm birth and gestational age. Effects on preterm birth and gestational age were considered together for the purposes of this SR; five studies were identified to characterize this endpoint (Fig. 9B). No association between preterm birth and maternal caffeine consumption at or above the comparator was found in the four observational studies included (Bakker et al., 2010; Bracken et al., 2003; Clausson et al., 2002; Maslova et al., 2010). Bakker et al. (2010) and Clausson et al. (2002) evaluated higher intake levels, and no effects were observed following a maternal caffeine intake ≥540 and > 500 mg/day in these studies, respectively. One of these studies was a meta-analysis that evaluated case-control (n = 7) and cohort (n = 15) studies with coffee, tea, cocoa/chocolate, and cola or soda drinks as the sources of caffeine exposure (Maslova et al., 2010). The fifth study was a randomized double-blind controlled trial (Bech et al., 2007). In this study, pregnant women already consuming at least 3 cups of coffee per day were assigned to drink caffeinated or decaffeinated coffee during the second half of pregnancy, but they were permitted to drink as much caffeine from coffee as well as other sources as desired. No differences in gestation length (95% CI, −2.87 to 0.25; $P = 0.48$) were reported between the two groups (where median daily caffeine intake for the two groups was 117 versus 317 mg/day).

Overall, there is a moderate to high level of confidence (Table 1) that the comparator of 300 mg/day is an acceptable intake level that would not be associated with significant concern for preterm birth and gestational age in healthy pregnant women (Table 2), because the data consistently showed a lack of effects both above and below the comparator. Several studies present findings that suggest the comparator is too low.

3.4.1.6. Fetal growth. Of the 14 studies evaluating fetal growth, nine of the included studies reported no effects of maternal caffeine consumption up to the comparator of 300 mg/day for the following endpoints: birthweight (Bakker et al., 2010; Balat et al., 2003; Bech et al., 2007; Bracken et al., 2003; Clausson et al., 2002), SGA (Bakker et al., 2010; Bech et al., 2015; Hoyt et al., 2014; Infante-Rivard, 2007), IUGR (Bracken et al., 2003; Grosso et al., 2001), and placenta weight, placenta diameter, newborn length, or head circumference (Balat et al., 2003) (Fig. 9C).

Two of the studies, including one of the Danish National Birth Cohort, revealed an increased risk for SGA at consumption levels above the comparator. Hoyt et al. (2014) reported effects at consumption levels ≥300 mg/day, and Bech et al. (2015) reported effects at 25–300 mg/day for decreased birth weight, although associations with increased risk of SGA were only observed in the group that consumed >325 mg/day. One study reported that newborn weight was significantly lower in mothers who consumed >300 mg caffeine/day compared to those consuming <300 mg/day ($P < 0.05$); no details on the method and frequency of intake recall were provided (Balat et al., 2003). In addition, the only analysis performed was a chi-square test, as such no adjustment for confounders was performed. Although a significant effect on mean birth weight (−28 g/100 mg caffeine) was noted in the study by Bracken et al. (2003), the authors concluded that is was not clinically important below intakes of 600 mg/day. Similarly, Bakker et al. (2010) reported a significant association with low birthweight at maternal intake levels ≥540 mg/day. In the randomized double-blind controlled trial described above by Bech et al. (2007), no differences in infant birth weight (95% CI, −40 to 73; $P = 0.57$) were reported between the two groups (117 versus 317 mg/day median intake). One study reported that placenta weight was significantly lower in mothers who consumed >300 mg caffeine/day compared to those consuming <300 mg/day ($P < 0.05$), whereas no significant association was found between caffeine intake and newborn length, head circumference, or placental diameter (Balat et al., 2003).

Four of the 14 included studies reported associations between prenatal caffeine exposure and adverse effects on fetal growth at

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

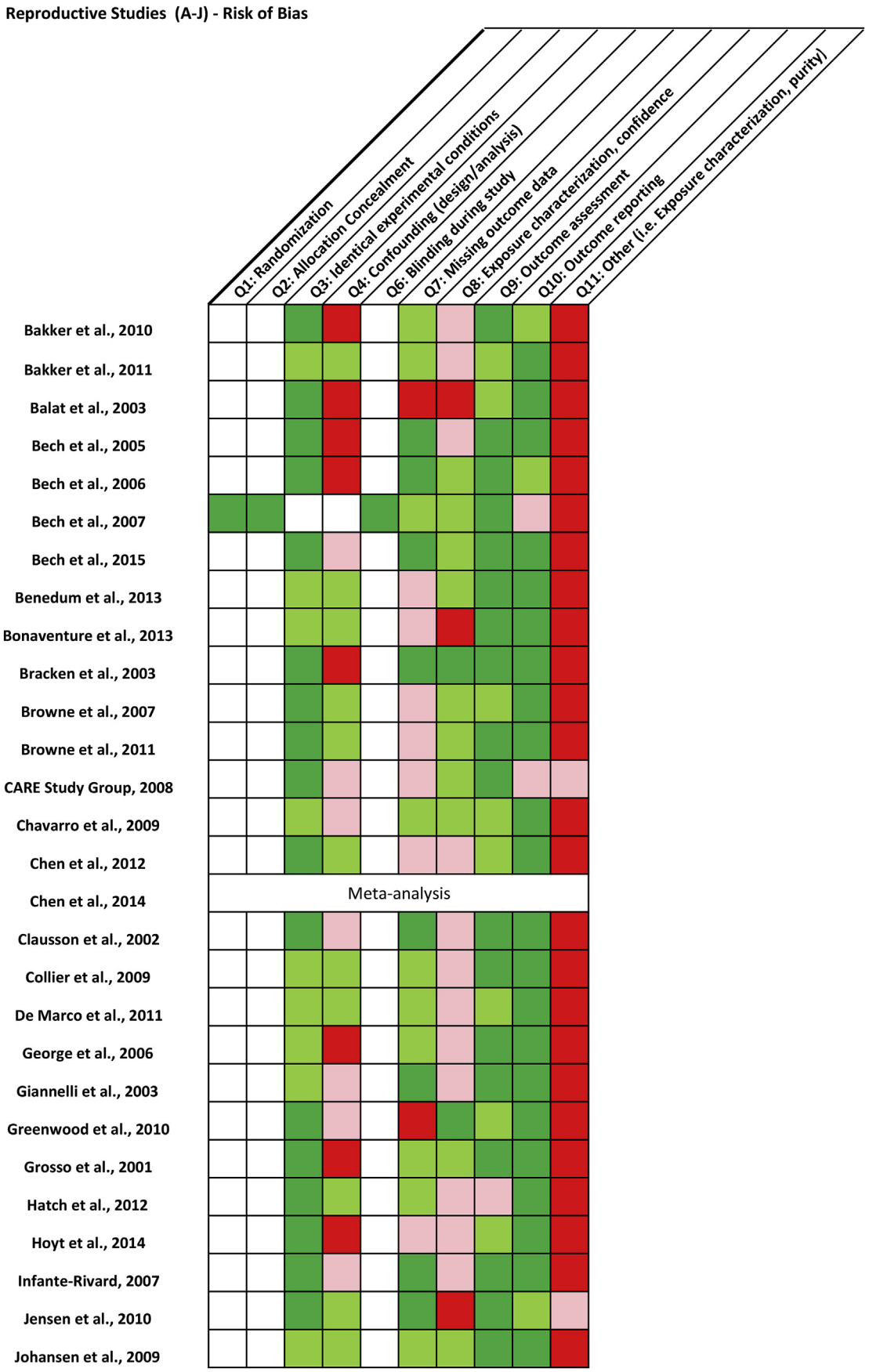


Fig. 10. Risk of bias (RoB) heat map for studies included in the reproductive and developmental outcome. The domain-based validity was evaluated based on study type per the OHAT (2015b) RoB tool. RoB for each domain is indicated by color: "definitely low risk of bias" (dark green, +2), "probably low risk of bias" (light green, +1), "probably high risk of bias" (light red, −1), and "definitely high risk of bias" (dark red, +2). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

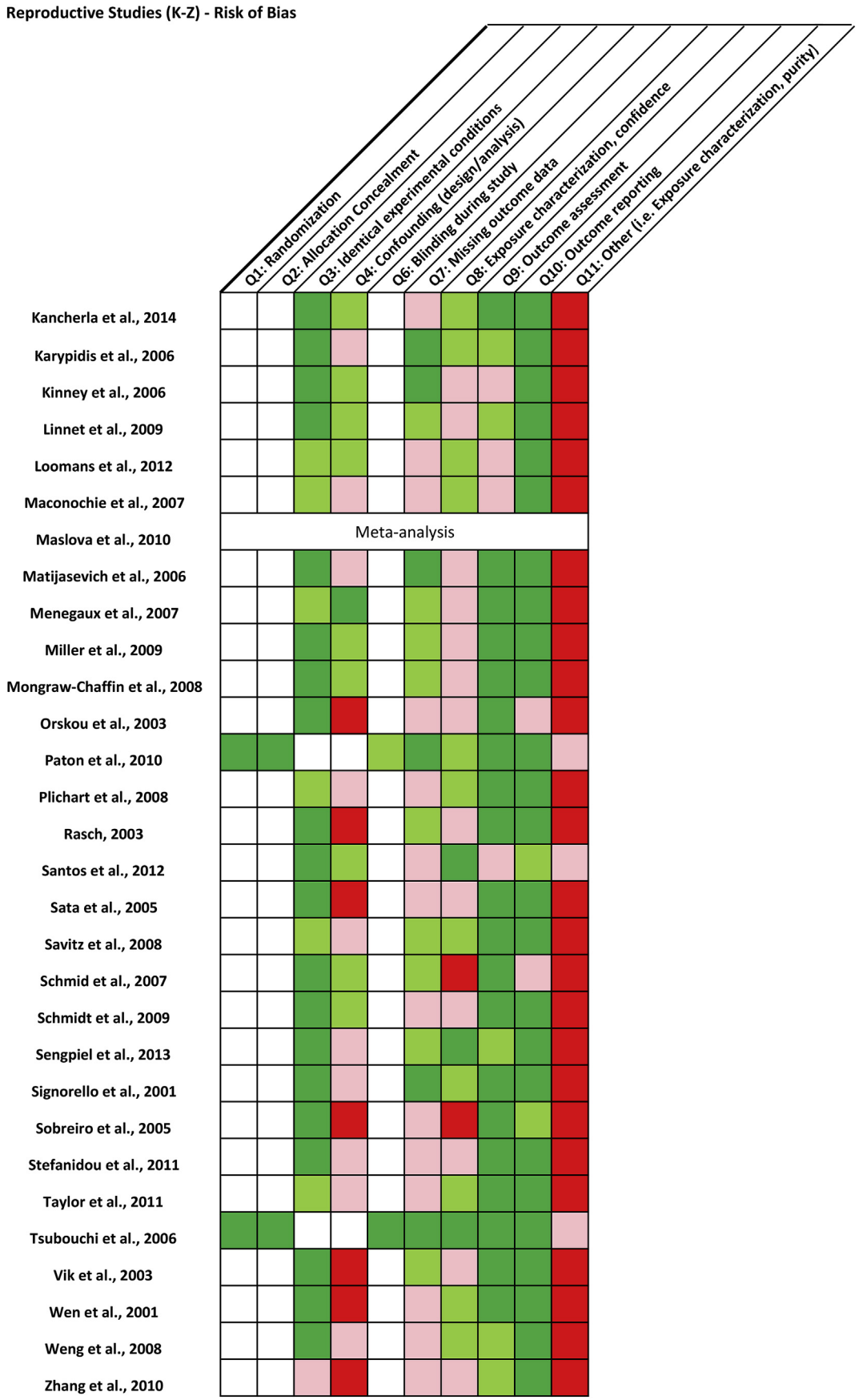


Fig. 10. (*continued*).

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

maternal intake levels lower than the comparator (Fig. 9C). Maternal caffeine consumption was associated with SGA offspring in a large prospective cohort study by Sengpiel et al. (2013); the LOEL in this study was identified as 51–200 mg/day, regardless of the definition used for SGA. In a case-control study, caffeine intake in the second and third trimesters was associated with increased risk of SGA; this was demonstrated in adjusted analyses in which high versus low consumption (<223 mg/day versus >223 mg/day) had an OR of 1.5 (95% CI, 1.0, 2.4) (Vik et al., 2003). Analysis by quartiles of caffeine intake was not adjusted for potential confounders but identified an increased risk starting at 110–204.9 mg/day (OR, 1.9; CI, 1.0, 3.7), with a *P* for trend of 0.001 (Vik et al., 2003). In a large prospective cohort study, maternal caffeine consumption of 200–299 mg/day during pregnancy was associated with an increased risk of IUGR (OR, 1.5; 95% CI, 1.1–2.1); a significant trend was also noted across all trimesters (CARE Study Group, 2008). An association was also observed in a regression analysis for birth weight, with >200 mg/day associated with a reduction in birth weight of about 60–70 g ($P = 0.004$). Finally, as mentioned above, Bech et al. (2015) reported effects at 25–300 mg/day for decreased birth weight, although this effect is not thought to be clinically significant because associations with increased risk of SGA were only observed at levels >325 mg/day. However, a meta-analysis evaluating 13 prospective studies concluded that there was "no clear threshold level of intake below which caffeine does not affect birth weight" (Chen et al., 2014). In this analyses, low birth weight/SGA/IUGR (combined in analysis) was found to be associated with 50–149 mg/day maternal caffeine intake based on RRs of 1.13 (95% CI, 1.06–1.21) for low caffeine intake (50–149 mg/day), 1.38 (95% CI, 1.18–1.62) for moderate caffeine intake (150–349 mg/day), and 1.60 (95% CI, 1.24–2.08) for high caffeine intake (≥350 mg/day), as compared with no or very low caffeine intake.

The final study included for review was one designed to investigate the relationship between maternal caffeine consumption and risk of high birth weight babies. In this large prospective cohort study, higher levels of caffeine intake were demonstrated not to be associated with increased risk of delivering a high birth weight baby (>4000 g; Ørskou et al., 2003). The adjusted OR for pregnant women with caffeine intakes ≥400 mg/day was 0.87 (95% CI, 0.79, 0.95), whereas it was 0.88 (95% CI, 0.81, 0.96) for pregnant women with caffeine intakes of 200–399 mg/day. Although the authors concluded that caffeine intakes ≥400 mg/day were associated with a reduced risk of giving birth to a high birth weight baby, they also noted that there was a significantly increased risk for delivery of a baby weighing >4000 g in the reference group (<200 mg/day).

The body of evidence for fetal growth was difficult to integrate based on the inconsistent findings, and thus difficult to determine a refined conclusion; all of comparison points below the comparator indicated observed effects (suggesting the comparator is too high), whereas the majority of comparison points equal to or greater than the comparator reported a lack of effects (suggesting the comparator was acceptable or too low). The biological significance of the birth weight changes are more robustly evaluated in studies evaluating SGA or IUGR, which as a whole, did not provide support for effects below the comparator. A low magnitude of effect (between 1 and 2 for studies below the comparator) —as well as the observation that, in many cases, effects were limited to single measures and/or subgroups or were not clinically relevant changes—reduced overall confidence in the data suggesting that the comparator may be too high. Across all studies, only five controlled for nausea and/or vomiting. Confidence in the underlying body of evidence supporting the conclusion is increased, however, when the low risk of bias (Fig. 10) was considered.

3.4.1.7. Birth defects. A total of 11 studies that evaluated birth defects and caffeine consumption were included in the SR (Fig. 9C). The following birth defects were evaluated and reported not to be associated with maternal caffeine consumption at or above the comparator of 300 mg/day in any of the studies included in the SR: cardiovascular malformations (CVMs; overall nor in any of the CVM subgroups studied) (Browne et al., 2007), choanal atresia (Kancherla et al., 2014), cleft lip (with or without cleft palate) or cleft palate only (Collier et al., 2009; Johansen et al., 2009), persistent cryptorchidism (Mongraw-Chaffin et al., 2008), and various other individual birth defects, including anotia/microtia, esophageal atresia, diaphragmatic hernia, omphalocele, or gastroschisis (Browne et al., 2011). In the study by Browne et al. (2011), high dietary caffeine consumption (≥300 mg/day) and craniosynostosis were associated (OR, 1.34; CI, 1.01–1.77) but no dose response was observed. These same authors reported statistically significant associations between both small intestinal atresia and total caffeine intake of 10 to <100 mg/day (OR, 1.54; 95% CI, 1.02–2.33) and 200 to <300 mg/day (OR, 1.79; CI, 1.09–2.93) but not at other intake levels. It is important to note that no dose response was seen for any birth defects evaluated in this study; thus, convincing evidence of an etiologic relationship was not supported. An increased risk of persistent cryptorchidism (present at birth and age 2 years) in male children was only found to be associated with maternal first-trimester caffeine consumption of 405 mg caffeine/day (OR, 1.43; 95% CI, 1.06, 1.93) (Mongraw-Chaffin et al., 2008); of note, this intake level was described as an interquartile range, but no information on other consumption levels was provided.

Two of the 11 studies included in the SR reported statistically significant effect ranges below the comparator (Chen et al., 2012; De Marco et al., 2011) and one reported borderline statistical significance for a LOEL (Miller et al., 2009). In a large case-control study, a weak but marginally statistically significant association was observed between maternal caffeine intake and anorectal atresia in offspring; the ORs for 10–99 mg/day, 100–299 mg/day, and ≥300 mg/day were 1.4 (CI, 1.0, 1.9), 1.3 (CI, 1.0, 1.8), and 1.5 (CI, 1.0, 2.2), respectively, when compared to <10 mg/day (Miller et al., 2009). However, even though a LOEL of 10–99 mg/day was identified, no dose response is visible and the CIs at all intake levels start at unity. In a second study on the same case-control population, Chen et al. (2012) concluded there to be a "weak" or "moderate" increased risk of limb defects (LDs) and maternal caffeine intake. No effects were observed for isolated longitudinal or longitudinal preaxial LDs, and all subgroups for multiple congenital anomalies with LD at intakes of ≥300 mg/day. Intake levels associated with no effects on all isolated LDs or transverse LDs were both 10 to <100 mg/day. The LOEL was determined to be the lowest significant effect level based on the authors' conclusions, despite the similar response across all dose groups and lack of dose response.

Finally, neural tube defects (NTDs) were evaluated in three different case-control studies included in the SR, and the results conflicted. In the first study, maternal coffee intake in the periconceptional period was associated with an increased risk of NTDs in offspring; caffeine intake >285 mg/day had an adjusted OR of 10.82 (95% CI, 3.78–31) (De Marco et al., 2011). In the second study, although the authors concluded that maternal caffeine consumption (coffee, tea, soda, chocolate) was associated with all NTDs (adjusted OR, 1.3; 95% CI, 1.0, 1.6) and spina bifida (OR, 1.4; CI, 1.1, 1.9), these data are based on a none (0–9 mg/day) versus all (≥10 mg/day) analysis (Schmidt et al., 2009). When analyzed by quartiles, there was no evidence of a trend and there were no significant findings at higher intake levels of ≥300 mg/day for NTDs (OR, 1.2; CI, 0.8–1.6), spina bifida (OR, 1.4; CI, 0.9–2.1), encephalocele (OR, 0.8; CI, 0.5–1.5), and anencephaly (OR, 1.4; CI, 0.6–3.3). In the third case-control study, maternal caffeine consumption of

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

≥190 mg caffeine/day was not associated with risk of spina bifida in offspring at the highest intake evaluated (OR, 0.6; 95% CI, 0.3, 1.2) (Benedum et al., 2013).

Although the evidence base is broad with respect to the type of birth defects and underlying etiologies, data were relatively consistent in demonstrating a lack of effects following consumption of caffeine at intakes up to 300 mg/day in healthy pregnant women. Based on the underlying study types (observational), low risk of bias, and consistency in findings, there was a moderate level of confidence in this conclusion.

3.4.1.8. Childhood cancers. Three studies that evaluated childhood cancers as related to maternal exposure to caffeine were reviewed and included in the SR: there was one study on CNS tumors and there were two studies on childhood leukemia (Bonaventure et al., 2013; Menegaux et al., 2007; Plichart et al., 2008). Neither maternal coffee consumption nor tea consumption alone was associated with combined childhood malignant CNS tumors in the Epidemiological Study on Childhood Cancer and Leukemia (ESCALE) study, a national population-based case-control study (Plichart et al., 2008). However, total caffeine intake (coffee plus tea) of >332.2 mg/day was significantly associated with all CNS tumors (OR, 4.4; 95% CI, 1.5, 13) (Plichart et al., 2008). When assessed by tumor type, only ependymomas were observed to be associated with caffeine: >47.2 mg caffeine/day yielded an OR of 2.5 (CI, 1.1, 5.9). Coffee plus tea (>332.2 mg/day) was also significantly associated with ependymomas but the sample size was small (n = 3; OR, 23.1; CI, 4.4–120); no association was seen with coffee alone. In a French population-based case-control study, maternal caffeine intake during pregnancy (coffee only) of >285 mg/day was not significantly associated with an increased risk of childhood acute leukemia, acute lymphoblastic leukemia, or acute myeloblastic leukemia (Menegaux et al., 2007). The authors note as an additional key finding that maternal coffee of >3 cups/day during pregnancy was related to acute leukemia (AL) in children whose mothers were nonsmokers; however, the data are incompletely presented. Conversely to the primary findings from Menegaux et al. (2007), Bonaventure et al. (2013) report that maternal coffee intake during pregnancy (timing of retrospective collection not provided) was associated with increased risk of childhood acute leukemia in a case-control study. Acute leukemia was found to be associated with consumption of 95–190 mg caffeine/day (OR, 1.3; 95% CI, 1.0–1.7), whereas acute myeloblastic leukemia and acute lymphoblastic leukemia effects were reported at consumption levels >190 mg/day.

The issue of recall bias (i.e., a type of error associated with differences in the accuracy of recollections by cases and controls) is critical to interpreting findings of these studies. In each of these case-control studies, the controls were healthy children, thus the potential for differential accuracy in recalled caffeine intake reduces confidence in these findings. This topic was acknowledged by both the authors as well as by working group experts at the International Agency for Cancer Research (IARC) in the recent review of the potential carcinogenesis of coffee in which IARC concluded that overall coffee drinking was unclassifiable to its carcinogenicity to humans (Loomis et al., 2016). The limited number of studies, combined with the significant impact of potential recall bias, precluded the development of a conclusion for this SR but highlights the need for additional research that accommodates this significant bias in the future.

3.4.1.9. Prenatal exposure related to childhood behavior. Three studies were included in the SR that investigated the association between prenatal caffeine exposure and childhood behavior outcomes. No associations were observed in these studies at levels at or higher than the comparator of 300 mg/day (Linnet et al., 2009; Loomans et al., 2012; Santos et al., 2012). In one study, maternal caffeine (>425 mg/day) was not associated with risk of problem behavior in children at age 5 years (Loomans et al., 2012) (OR, 1.04; 95% CI, 0.49–2.22); ORs for emotional problems, conduct problems, hyperactivity/inattention problems, peer relationship problems, and prosocial behavior were similar. In another study, maternal caffeine consumption was not associated with childhood (median age 7 years) hyperkinetic disorder or attention deficit hyperactivity disorder (ADHD) at intakes ≥1000 mg/day; the adjusted RR for 10 cups/day coffee was 2.3 (95% CI, 0.9–5.9) (Linnet et al., 2009). Because data were limited to three studies, no conclusion was developed; however, the lack of effects observed in all studies suggests that this is not an endpoint of concern.

3.4.1.10. Other reproductive endpoints. Additional reproductive endpoints were identified as meeting the SR criteria, but only one study was identified for each; thus, they are grouped and discussed briefly here as an "other" category, and no conclusions were developed. Of these, two studies reported that the outcomes of interest were not associated with caffeine intake levels at or above the respective comparator: pregnancy-induced hypertensions and/or preeclampsia (NOEL of ≥540 mg/day; Bakker et al., 2011) and median age at menopause (NOEL of ≥400 mg/day; Kinney et al., 2006). In the third study, healthy pregnant women in their third trimester were evaluated for markers of maternal stress after consumption of 100 mg caffeine (in coffee) in a controlled trial (Tsubouchi et al., 2006). No effects on maternal or fetal blood flow were observed ($P < 0.05$) using Doppler blood flow analysis and caffeine was shown to have a protective effect against other markers of stress measured as salivary cortisol level chromogranin A.

3.4.2. Body of evidence assessment

With regard to the overall outcome of reproductive and developmental effects, individual studies were generally associated with low risk of bias, with only five studies at the lower end of the spectrum and two at the higher end (Figs. 9 and 10). The study ratings were most impacted by the confidence in exposure. As is inherent to these study types (as discussed elsewhere), very few studies involved direct evaluation of caffeine; rather, they involved self-reported estimates of consumption of caffeine-containing beverages. This topic was highlighted by Peck et al. (2010), in their conduct of a review aimed at characterizing potential reproductive health effects associated with caffeine consumption. The very small number of studies in which the authors made attempts to better quantify caffeine by analyzing beverage samples or biomonitoring in participants tend to better inform the body of evidence (Bracken et al., 2003; Santos et al., 2012; Tsubouchi et al., 2006).

The current body of evidence characterizing this endpoint is generally consistent with what was reported by Nawrot et al. (2003); the majority of studies included in the SR do not report reproductive or developmental effects at levels below the relevant comparator. Although effects below 300 mg/day (or 400 mg/day, in the case of males and nonpregnant females) cannot be ruled out with the currently available data, the effects seen at these levels were primarily limited to isolated congenital malformations (Chen et al., 2012; Miller et al., 2009) or childhood cancers (Bonaventure et al., 2013; Plichart et al., 2008) and findings were of relatively low magnitude (ORs, <2). Effects on birth weight were also reported at intakes below the comparator; however, the biological significance of such an effect is more robustly evaluated in studies evaluating SGA or IUGR, which as a whole did not show effects below the comparator.

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

Findings for some of these endpoints (e.g., birth defects and disease in offspring) can be significantly biased by poor study design, which was taken into consideration when developing conclusions. As noted in many of the studies reviewed, as well as review articles conducted prior to the SR, observational studies evaluating reproductive and developmental endpoints suffer from common limitations in study design. Although exposure assessment in most studies relies on self-reporting techniques such as telephone interviews and food frequency questionnaires, the timing and frequency of intake data collection varied widely between studies, with some analyzing intake data from before pregnancy and/or during one or more trimesters. In addition, the majority of studies utilized data collected at one time (often early in pregnancy) and calculated an average meant to represent the entire pregnancy. In addition to contributing to inconsistency between studies, the timing of intake relevant to the event or outcome should also be considered.

In summary, the SR of 58 studies (55 observational; 3 RCTs) provided evidence to evaluate potential impacts of the consumption of 300 mg caffeine/day in pregnant women, or 400 mg/day in healthy adults, on reproductive and developmental effects. These studies included assessment of fecundability, fertility, male reproductive endpoints, spontaneous abortion, recurrent miscarriage, stillbirth, preterm birth and gestational age, fetal growth, birth defects, childhood cancers, and childhood behavior. When the weight of evidence was considered, an intake of 300 mg caffeine/day in pregnant women and 400 mg/day in healthy adults are generally without significant concern regarding overt adverse effects on fecundability, fertility, male reproductive endpoints, spontaneous abortion, recurrent miscarriage, stillbirth, preterm birth and gestational age, birth defects, and childhood behavior. For some endpoints, including fetal growth, childhood cancers, isolated congenital malformations, evidence indicated that intake at the level of the comparator may be high; however, refined conclusions were difficult to determine given uncertainties in the evidence base, including low attrition rates, lack of ability to account for recall bias particularly for childhood cancers, lack of ability to fully accommodate the impact of the pregnancy signal, and lack of biological relevance of findings (referring primarily to fetal growth changes; the biological significance of such an effect is more robustly evaluated in studies evaluating SGA or IUGR, which, as a whole, did not support effects below the comparator). The relevance of smoking to the analyses in this data set was generally well recognized and, in most studies, the investigators attempted to control for such. There is a moderate level of confidence in the evidence base which supports this conclusion.

The findings of this SR for the reproductive and developmental outcome are consistent with other reviews conducted prior to the SR on reproductive and developmental endpoints in general (Peck et al., 2010) as well as those considering individual endpoints (Brent et al., 2011; Christian and Brent, 2001; Peck et al., 2010). As such, the current guidance for pregnant women is supported by the findings of this SR.

3.5. Acute toxicity

Forty-six full text papers were reviewed for potential evaluation of acute toxicity; 26 were considered to meet the criteria for inclusion in the SR, as they permitted comparison to Nawrot et al. (2003) conclusions. Of the studies that were excluded, most were because they did not meet the inclusion criteria (e.g., the most common reasons were a nonhealthy population or coadministration of caffeine with a drug), and the remaining few reports contained information regarding potential associations. All 26 of the included papers were case reports or case series, most of which were associated with emergency department (ED) visits and/or suicide-related events. As outlined in the protocol and methods, the inclusion/exclusion criteria differed slightly for this outcome relative to other outcomes in that these study types (i.e., case reports and case series) were allowed for inclusion in the acute outcome evaluation. Although these study types are generally regarded as having lower quality than other study types in humans (AHRQ, 2013; Cochrane Collaboration, 2011), inclusion of such studies is warranted when (a) there are no other studies to consider, and/or (b) the objective is to obtain information on rare and adverse effects (Eden et al., 2011; Fitzpatrick-Lewis et al., 2009). The endpoints of interest in this outcome (mainly death and severe intoxication) are rare and adverse; without inclusion of case reports and series, no data are currently available for evaluation.

Regarding the populations of interest, the majority of the reports were described in adults, although four reports characterized acute effects in adolescents and two evaluated effects in pregnant women. No acute toxicity reports in children were found that met the criteria. The included cases involved adverse events associated with very high doses delivered over a relatively short time frame (up to 50 g); a few studies reported on repeated exposures. The majority of papers reviewed contained only brief discussions by the authors regarding the reported amount of caffeine ingested. The form of caffeine consumed varied; approximately half of the reports involved caffeine in a powder or tablet form (e.g., sleep aid), and the remaining reports involved energy drinks and a few involved cola as the source of caffeine. Some reports briefly mentioned ingested coffee or green tea but these beverages were not considered the major source of caffeine in the case reports reviewed. As discussed throughout this section, confidence in the characterizations of exposure is low, since exposures were nearly always self-reports of caffeine ingestion (sometimes with corresponding corroboration by friends or relatives or confirmation via blood tests). Furthermore, it is commonly recognized by experts in emergency medicine and clinical toxicology that quantifying exposure from intoxications is reported with a general recognition of lack of confidence in the assessment of exposure (Brett, 1988; Heyerdahl et al., 2008; Wright, 1980). These shortcomings in study design were accounted for in the evaluation of internal validity, as well as in the consideration of the body of evidence as a whole.

Acute effects associated with caffeine consumption have been described to result in a wide spectrum of symptoms, with headache, nausea, vomiting, fever, tremors, hyperventilation, dizziness, anxiety, tinnitus, and agitation at the milder end of the spectrum (Rudolph and Knudsen, 2010). More severe effects resulting from intoxication can include the following: abdominal pain, altered consciousness, rigidity, seizures, hypokalemia (low potassium), rhabdomyolysis (muscle breakdown), increased blood lactate (acidosis), as well as supraventricular and ventricular arrhythmias and myocardial ischemia (Holmgren et al., 2004). Many of these changes would be expected at high doses, considering caffeine's ability to stimulate the CNS, decrease smooth muscle tone, increase peripheral vascular resistance, and increase cerebrovascular resistance (Hering-Hanit and Gadoth, 2003). Kapur and Smith (2009) state that caffeine overdoses produce multiple symptoms, most of which are commonly associated with a marked increase in adrenergic tone, often including hypertension, tachycardia, dysrhythmias, and CNS and skeletal muscle stimulation. Acute toxicity for the purposes of this SR focused on consumption of caffeine in association with abuse, overdose, and potential death.

Nawrot et al. (2003) did not comprehensively review these acute effects; rather, discussion included a brief characterization of mortality and caffeinism (which the authors characterize as a syndrome associated with a range of effects such as restlessness,

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

anxiety, irritability, sensory disturbances, cardiovascular symptoms, and others), as well as detrusor instability in women. Based on these, two different comparator values associated with acute effects were considered as defined in Nawrot et al. (2003). For nonfatal events and the spectrum of adverse events associated with acute toxicity, <400 mg/day was utilized. For fatal events, the comparator of 10 g from Nawrot et al. (2003) was considered. Comparators for populations other than adults were undefined.

3.5.1. Summary of individual studies by endpoint

Of the 26 reports evaluated, 14 described death following exposure to caffeine, and the rest described other acute adverse effects (Fig. 11). Notably, many of these reports were intentional overdoses of caffeine in suicide attempts. Because most reports included evaluation of both lethality and nonlethality, the subsequent subsections are organized by population and each subsection addresses both lethal and nonlethal endpoints.

3.5.1.1. Acute lethality and nonlethality in adolescents. Six case reports for adolescents (aged 15–18 years) were reviewed; these cases involved exposures ranging from 495 mg to ~51.6 g caffeine. Of the six total reports, one death was reported (Yamamoto et al., 2015) and the remaining five cases described various adverse outcomes (Babu et al., 2011; Kapur and Smith, 2009; Schmidt and Karlson-Stiber, 2008; Usman and Jawaid, 2012; Wilson et al., 2012). Three of the reports were related to attempted suicide events following exposures to caffeine via caffeine tablets at doses that were at or above the Nawrot et al. (2003) comparator (Kapur and Smith, 2009; Schmidt and Karlson-Stiber, 2008; Yamamoto et al., 2015). The only death involved ingestion of 51.6 g caffeine in the form of antisleep tablets (Yamamoto et al., 2015) and was attributed to respiratory failure from brain hemorrhage; however, the authors note uncertainty in the classification. The two other reports were suicide gestures that involved exposures to 12 g caffeine (Schmidt and Karlson-Stiber, 2008), resulting in hyperventilation and increased lactate in the blood, and 10 g caffeine (Kapur and Smith, 2009), associated with cardiovascular collapse, respectively. Both of these patients presented with nausea, vomiting, tachycardia/cardiac instability, and hypokalemia and both fully recovered.

Three of the six adolescent cases were misuse rather than suicide gestures. The self-reported doses were above the 400 mg comparator for nonfatal effects. Each case involved relatively high consumption of energy drinks or shots in short time frames, resulting in estimated caffeine exposures of 480 mg/day (Usman and Jawaid, 2012), 560–800 mg/day (Wilson et al., 2012), and 495 mg/day (Babu et al., 2011). Two of the three cases required treatment; as described by the authors, one case returned to normal, one fully recovered, and one had no further seizure activity. The manifestation of adverse effects varied but similar effects observed included hypertension, heart palpitations, tachycardia, and hypokalemia. The case report by Babu et al. (2011) was unique in that seizures were documented in a patient not known to have a history of them.

Following review of data from these six case reports involving assessment of acute toxicity in adolescents (aged 15–18 years), data

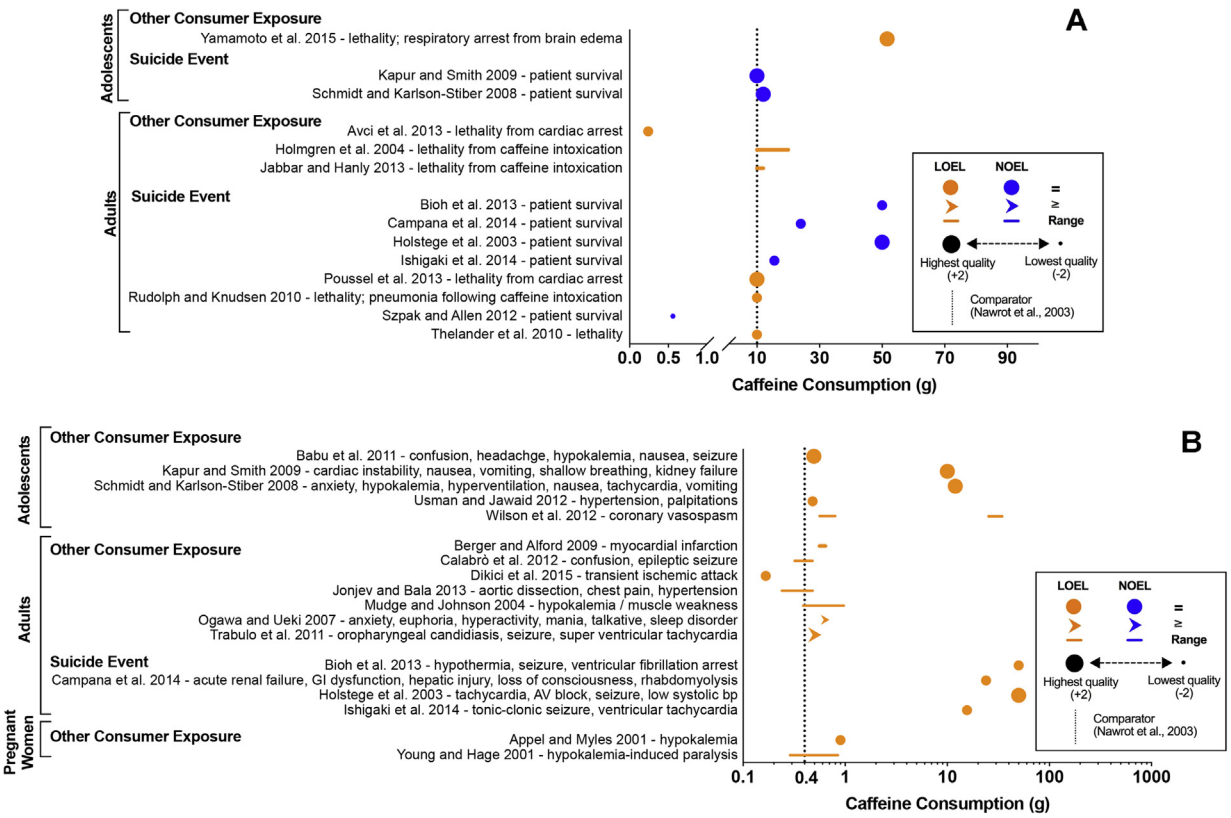


Fig. 11. (A and B) Summary diagram of exposure-response data relative to the comparator for the acute outcome: (A) lethality, and (B) other acute effects. Symbols represent caffeine intake (mg/day) as reported by original study authors. The color of the symbol indicates the type of effect; no effect (NOEL; blue symbols) or the lowest effect level (LOEL; orange symbols). The shape of the symbol represents the type of metric (circles represent a discrete value, arrowheads represent greater than or equal to a value, and a horizontal line represents a range of values; metrics based on that reported by original study authors). The size of the symbol indicates the overall risk of bias (larger symbols indicate a lower risk of bias, or higher methodological quality). The dashed vertical line marks the comparator value. Italicized study names indicate a meta-analysis. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

support that the comparators of 10 g and 2.5 mg/kg/day for lethality and nonlethal endpoints, respectively, are acceptable. There is a very low to low level of confidence in the evidence base (Fig. 11; Table 1). Confidence is limited by high risk of bias and likely publication bias; inherent to the case studies included, reports are limited to those reporting effects (and not reporting a lack of effects), as well as very low confidence in exposure estimates.

3.5.1.2. Lethality and nonlethality in adults. There were 20 publications reviewed that reported on the potential for adverse acute toxicity in adults; the majority of these cases were reports from EDs. All lethality events were associated with suicide. Nonlethal effects included anxiety, hypokalemia, seizures, and myocardial infarction. Generally, fatal cases involved exposures to caffeine powder or tablets, whereas cases associated with nonfatal adverse events involved exposures to energy drinks or colas.

Eight publications involving suicide attempts by caffeine ingestion were identified; all of these cases involved intake of anhydrous caffeine in powder or tablet form, and all cases except one had at consumption doses that were greater than or equal to the Nawrot et al. (2003) comparator of 10 g. Death occurred in three cases (Poussel et al., 2013; Rudolph and Knudsen, 2010; Thelander et al., 2010), all of which were associated with a reported 10-g ingestion (Fig. 11A). Only one of these three exposures was validated by blood assay for caffeine (Poussel et al., 2013). A myocardial event was reported cause of death in one case (Poussel et al., 2013). The other two patients died of brain damage (Thelander et al., 2010) and subsequent pneumonia (Rudolph and Knudsen, 2010). Sweats, trembling, rapid breathing, hypokalemia, and rhabdomyolysis were documented in all three of these cases. The remaining reports, all suicide gestures, described nonfatal adverse effects. Estimated exposure in these cases ranged from 1.1 g (Szpak and Allen, 2012) to 50 g (Bioh et al., 2013; Holstege et al., 2003). Bioh et al. (2013) and Holstege et al. (2003) both reported cases of exposure to 50 g caffeine in the form of caffeine powder and a dietary supplement, respectively. Both individuals suffered seizures, tachycardia, and forms of severe myocardial dysfunction (atrioventricular block or ventricular fibrillation arrest). Another case report of seizure occurred in a patient with schizophrenia who was admitted to the ED after ingesting 15.6 g caffeine from tablets (Ishigaki et al., 2014). Similar symptoms to those noted above were documented: sudden onset of a tonic-clonic seizure and ventricular tachycardia. Campana et al. (2014) reported an estimated exposure was 24 g caffeine (120 No-Doz tablets); adverse effects included uncontrollable vomiting, diarrhea, and intermittent loss of consciousness, followed by sustained severe rhabdomyolysis, acute renal failure, and hepatic injury. Notably, no dysrhythmias were observed, which is a hallmark finding of caffeine overdose. The lowest dose reported in regard to attempted suicide was described by Szpak and Allen (2012). They presented a case report in which the subject ingested 14 cans of energy drink (1.1 g caffeine) over the course 2 nights and was extremely sleep deprived for 72 h, which led to a state of mania-attempted suicide; little detail was provided on other adverse effects.

Four additional fatal cases were discussed by Holmgren et al. (2004). Each death was the result of intentional consumption of a large caffeine dose (two of which were believed to be suicide related and the other two were uncertain). All estimated doses were at or above the comparator of 10 g caffeine. The reported doses were 10 g in three cases and 25 g in the fourth case (all doses were in tablet form). All four cases were noted as having previous drug or alcohol use or prior suicide attempts; very little additional information was provided, although blood levels were reported, confirming high exposure to caffeine, and ventricular fibrillation was noted as the cause of death.

Death associated with accidental overdose (or unknown intentionality of consumption) was reported in two studies (Avci et al., 2013; Jabbar and Hanly, 2013), one of which was below the Nawrot comparator and one was above, respectively. In the first case, the patient was a self-reported chronic consumer of energy drinks (one per day for 7 months) who reported intake of 240 mg caffeine 5 h before playing in a basketball match. Adverse events included heart palpitation, nausea, loss of consciousness, tachycardia, and death from cardiac arrest. Jabbar and Hanly (2013) reported a case of an individual found dead by a friend and who was believed to have ingested 10–12 g caffeine powder. Because the patient was dead upon arrival, the authors state that estimates were based on residual contents of a bag of pure caffeine powder found on the decedent and the subject's stomach contents and extremely high caffeine blood values. No other details were provided.

There were a number of publications in which attempted suicide was not the intent of consumption. These studies described high doses of caffeine ingested in short periods of time with nonfatal adverse events manifest (Berger and Alford, 2009; Dikici et al., 2015; Jonjev and Bala, 2013) (Fig. 11B). All three of these publications involved exposure to caffeinated energy drinks, in which estimated exposures were both above and below the Nawrot et al. (2003) comparator of 400 mg and some form of myocardial adverse event was noted. Berger and Alford (2009) reported an ED visit following reported ingestion of 560–640 mg caffeine over 7 h during motocross racing; the patient experienced a myocardial infarction and hypokalemia as well as transmural ischemia. Jonjev and Bala (2013) discussed three cases in which they suspected caffeine ingestions were associated with aortic dissections (tear in inner layer of the aorta). Caffeine exposures were estimated at 240 mg/day, 320–400 mg/day, and 400–480 mg/day, respectively. Dikici et al. (2015) reported a transient ischemic attack occurring after self-reported consumption of two energy drinks (~167 mg) and sudden-onset loss of vision in the right eye, which spontaneously resolved after 4 h.

Two caffeine-associated seizure-related events were also reviewed (Calabrò et al., 2012; Trabulo et al., 2011). Both report exposure to caffeine from energy drinks that were above and below the comparator of 400 mg/day. Calabrò et al. (2012) discussed a witnessed epileptic seizure in a patient with no known history of seizure or associated conditions, in which the authors estimated consumption of 340–480 mg caffeine/day. Trabulo et al. (2011) reported a case with a reported exposure of 500 mg caffeine over 4 h; postictal state, tonic-clonic seizures, and supraventricular tachycardia were reported. Mudge and Johnson (2004) described a patient with possible ingestion of 400 mg to 1 g/day from cola; symptoms included hypokalemia, muscle weakness, and respiratory depression (requiring intubation).

Ogawa and Ueki (2007) discussed two patient cases not associated with ED visits, but rather with long-term use of caffeine at high doses, both of which exceed the Nawrot et al. (2003) comparator of 400 mg (case 1, >640 mg/day from energy drinks; case 2, 1365–1450 mg/day from a combination of instant coffee, green tea, and caffeine tablets). The authors diagnosed these patients with substance abuse disorder. Adverse effects included euphoria, mania, excessive talking, hyperactivity, anxiety, and sleep disorder for case 1 and tachycardia, flushing, cold sweats, anxiety, agitation, and sleep disturbance for case 2. Medical treatment was administered and both patients recovered.

Following review of data from these 20 reports involving assessment of acute toxicity in adults, it was determined that the comparators of 10 g and 400 mg/day for lethality and nonlethal endpoints, respectively, are acceptable. There was a very low to low level of confidence in the evidence base. Confidence is limited by

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

high risk of bias (likely publication bias) (Fig. 12; Table 1); inherent to the case studies included, reports are limited to those reporting effects (and not reporting a lack of effects), as well as very low confidence in exposure estimates.

3.5.1.3. Acute effects in pregnant women. Our review included two separate reports of pregnant women that experienced acute effects associated with self-reported high amounts due to ingestion of large volumes of cola (Appel and Myles, 2001; Young et al., 2001) (Fig. 11). Both women presented with muscular paralysis and hypokalemia; consumption estimates associated with these nonfatal endpoints were 900 mg/day and 288–850 mg/day for the respective cases. Both cases recovered following routine supportive care. There is a very low to low level of confidence that these data are indirectly supportive of the comparator of 300 mg/day; however, these data were not considered sufficient to develop a conclusion.

3.5.2. Body of evidence assessment

Overall, the current body of evidence characterizing this endpoint is generally consistent with what was reported by Nawrot et al. (2003), which suggests the potential for death following acute exposures of approximately 10 g caffeine. Seven of the fatal case reports documented death following ingestion of ~10 g caffeine (Holmgren et al., 2004; Jabbar and Hanly, 2013; Poussel et al., 2013; Rudolph and Knudsen, 2010; Thelander et al., 2010) or higher, such as 24 g (Holmgren et al., 2004) or 51.6 g (Yamamoto et al., 2015). There was only one report of death at a lower dose than the 10 g. Avci et al. (2013) suggested that an exposure to 240 mg was associated with cardiac arrest; this case appears to be an outlier and/or associated with other factors (e.g., preexisting conditions), given that the patient had been consuming similar amounts for months prior without effects and because it is not consistent with the other case studies evaluated. Five reports described survival following consumption at a dose at or above the comparator: 10–12 g (Kapur and Smith, 2009; Schmidt and Karlson-Stiber, 2008) or significantly higher, such as 15–24 g (Campana et al., 2014; Ishigaki et al., 2014) or even 50 g (Bioh et al., 2013; Holstege et al., 2003). As such, the data support that for healthy individuals, lethality may, but does not always, occur following acute consumption of 10g caffeine.

The review of the data also generally supports a lack of nonlethal acute effects at or below exposures of 400 mg/day. As with the reports evaluating death, the confidence in exposure is very low and, in many of these studies, existing conditions were often thought to have potentiated the effects observed with caffeine. For example, Jonjev and Bala (2013) indicated that a dose of 240 mg from energy drink consumption was possibly associated with greater serious cardiovascular event (aortic tears) in an individual with known underlying heart disease, possibly provoking a potential fatal cardiovascular event. Yet the authors reached the conclusion in regard to uncontrolled consumption of "high-energy" drinks and not in direct regard to the dose of caffeine discussed. Thus, for the majority of endpoints discussed as associated with acute intoxication, the body of evidence suggests a general lack of acute toxicity concern following exposure to 400 mg caffeine in healthy individuals.

With respect to acute toxicity, it was notable that each case appeared to have a rather unique spectrum of adverse events, although commonly reported events included vasospasm, seizure, mania, hypokalemia, and muscle weakness. Nearly all of the case reports describing fatalities involved caffeine powder and tablets, whereas the case reports associated with other acute effects generally involved rapid consumption of caffeinated beverages over a short time. Based on the limited data available, no particular population appeared to be more sensitive; however, there are insufficient data to discern whether adolescents or pregnant women were more at risk for acute toxicity or death. No case reports that fit the criteria were found for children; thus, doses required to produce adverse effects in that population remain an area of uncertainty. However, the fact that no cases were identified in this SR may support that current consumption practices do not warrant concern.

Critical to the assessment of acute toxicity is the recognition that case reports and case series are considered to be of lesser quality than other study types with respect to overall confidence and weight in describing a body of evidence. The majority of studies included were generally associated with a high risk of bias rating, although studies ranged the spectrum from probably low to definitely high (Fig. 12). The study ratings were most impacted by the confidence in exposure, as is inherent to these case studies in which the individual self-reported intake or intake was characterized by health care providers' estimation. There was a large number of reports that evaluated toxicity associated with caffeine in tablets; however, because these reports were not controlled (e.g., estimates based on number of tablets left in a package) and blood caffeine levels often were not measured, uncertainty remains as to the complete dose characterization. Uncertainty is also inherent to the estimations of caffeine from self-reported exposure; efforts were made to standardize these estimates, although there could be large variability in products and descriptions of amounts consumed. As would be expected, the studies in which authors made attempts to specifically quantify or verify the self-reported caffeine level tend to better inform the body of evidence. Furthermore, when interpreting these data, it is notable to consider both that (1) the caffeine doses discussed herein do not represent typical consumer exposures and (2) because of the nature of the study type (i.e., case report), the evidence base for this outcome is inherently biased toward demonstration of effect (i.e., case reports do not often demonstrate a lack of effect).

In summary, the SR of 26 studies provided evidence to evaluate potential acute toxicity. Following a weight of evidence review, the comparators of 10 g for lethality and 400 mg/day or 2.5 mg/kg/day for other acute effects, were determined to be acceptable for healthy adults and adolescents, respectively. There is very low to low confidence in this evidence base, due primarily to uncertainty in the estimates of exposure and to the high risk of bias. A higher level of confidence is also precluded by the inherent bias introduced by utilizing case reports (i.e., limitations to reporting of effects versus no effects). Insufficient and/or lack of data precluded conclusions related to acute toxicity for children and pregnant women.

3.6. Caffeine pharmacokinetics and pharmacodynamics

PK parameters were evaluated contextually with the aim to provide information in target populations that are not as well characterized relative to adults (e.g., children, adolescents, pregnant women), as well as to identify literature that aids in providing context to some of the potential adverse effects observed (PD). The PK of caffeine in healthy adults is well established, a summary of such is provided in Supplementary File S3. Our search criteria aimed to capture all relevant papers for caffeine that may have contained PK and PD data published since Nawrot et al. (2003), with specific focus on individual variation in metabolism and other pharmacogenomic variability. Studies that were deemed to contain important information to advance the understanding of caffeine PK/PD were systematically identified and carefully reviewed, and the findings are summarized in this section. However, these papers were not evaluated for risk of bias or evaluated as an overall body of evidence, given the wide range of endpoints associated with the topic. Fifty-seven studies were reviewed for relevant PK data, and

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

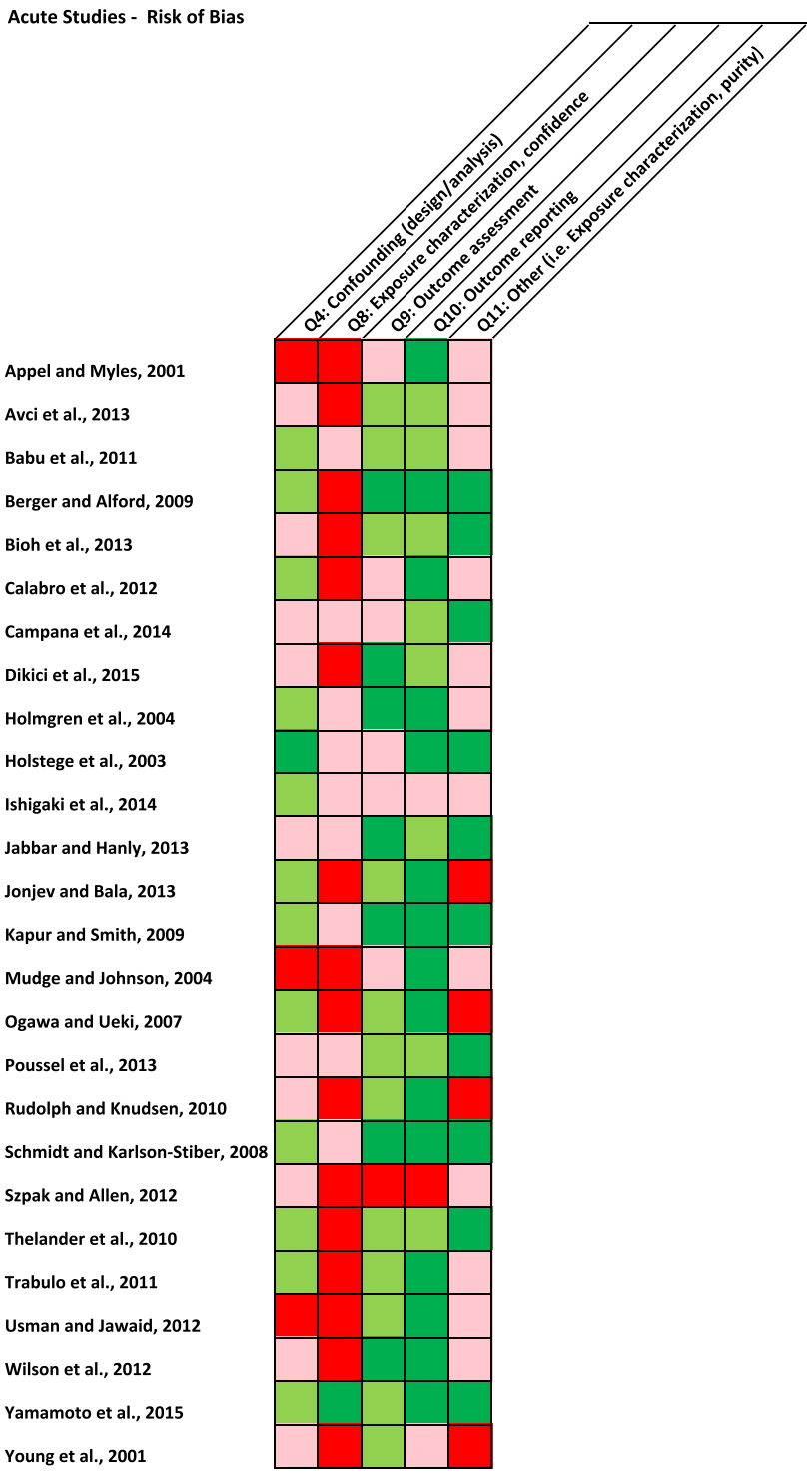


Fig. 12. Risk of bias (RoB) heat map for studies included in the acute outcome. The domain-based validity was evaluated based on study type per the OHAT (2015b) RoB tool. RoB for each domain is indicated by color: "definitely low risk of bias" (dark green, +2), "probably low risk of bias" (light green, +1), "probably high risk of bias" (light red, −1), and "definitely high risk of bias" (dark red, +2). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

46 studies contained relevant PK information in context of a target outcome (PD), resulting in review of ~100 studies (combined from the PK category only and PK information from other outcomes; Fig. 2). For brevity, information by outcome is provided in Supplementary File S3 and a summary of key findings is presented below.

Since caffeine PK has been relatively well understood for decades, more recent research in the area of caffeine metabolism has focused on how one's own genetic makeup leads to interindividual differences in how caffeine is handled by the body. The most common PK/PD topic reviewed was in relation to how small nucleotide polymorphisms (SNPs) have been characterized, further helping to elucidate individual differences in caffeine metabolism and even consumption practices. Simply put, this work evaluated

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

changes at the allele level in genes and subsequent association of changes in how one's body handles exposure to caffeine. As an example, the ADORA2A gene encodes the adenosine A_{2A} receptor; several studies evaluated how polymorphisms in this gene can affect individual sensitivity to caffeine, including differences in anxiogenic responses (e.g., Thorn et al., 2012; Yamada et al., 2014; Yang et al., 2010). For example, following acute caffeine consumption, light caffeine users with ADORA2A SNPs rs5751876 (T/T) and rs35320474 (T/T) reported increased feelings of anxiety (Alsene et al., 2003; Childs and de Wit, 2008). However, genotypic variation in ADORA2A (i.e., 1083 TT) was also linked to decreased caffeine intake (Cornelis et al., 2007).

A number of studies further evaluated pharmacogenomic variations in caffeine associated with *CYP1A2*1F* (variant rs762551, genotype AA) and the *CYP1A2*1K* alleles, which are associated with increased and decreased caffeine metabolism, respectively (Arnaud, 2011; Cornelis et al., 2011; Josse et al., 2012; McLean and Graham, 2002; Sulem et al., 2011; Thorn et al., 2012; Vink et al., 2009). Two studies reported that gender has not been shown to consistently affect most caffeine pharmacokinetic parameters but men may have higher CYP1A2 activity than women (Arnaud, 2011; McLean and Graham, 2002). Interestingly, it was hypothesized more recently that caffeine consumption itself may be a heritable trait (estimated 43–58% from twin studies) that may be influenced by CYP1A2 genotype (Cornelis et al., 2011). For example, CYP1A2 variants (i.e., rs2470893 and rs2472297, allele T) have been linked to increased coffee consumption (Cornelis et al., 2011; Josse et al., 2012; Sulem et al., 2011; Thorn et al., 2012). Coffee consumption and the preference of coffee over tea (perhaps due to differences in perceived bitterness) was also demonstrated to be somewhat heritable in a Dutch twin study (Vink et al., 2009).

Collectively, data are beginning to provide insight into potential epigenetic trends or effects, including further characterizations of SNPs believed to be associated with consumption practices (e.g., self-regulation), as well as specific effects, including several behavioral endpoints (i.e., mood, tolerance, withdrawal). Understanding the metabolism, pharmacology, and mechanism of action of caffeine is helpful to interpreting its effects, adverse or otherwise, throughout the body. This review of newer PK/PD data further highlights the importance of taking PK into consideration, in particular the individual variation presented in caffeine's targets or metabolic enzymes, when interpreting overall findings, as well as future endeavors to characterize sensitive effects or sensitive populations.

4. Discussion

In line with our goal of updating the work of Health Canada (Nawrot et al., 2003), the possibility that caffeine ingestion adversely affects human health at moderate levels of consumption (400 mg/day in health adults, 300 mg/day in healthy pregnant women, and 2.5 mg/kg/day in children and adolescents) was investigated via the IOM SR approach (Eden et al., 2011). The review focused on five outcomes: calcium and bone status, cardiovascular effects, behavioral effects, reproductive and developmental toxicity, and acute toxicity. Data on these outcomes were obtained from the assessment of 380 studies; for each study, information as reported by the authors was extracted, and an evaluation of study quality and applicability (i.e., internal validity and external validity) was conducted. The body of evidence for each health outcome was then reviewed and conclusions were developed using a weight of evidence approach. Considerations of quality of the evidence base, consistency in findings, magnitude of response, and level of adverseness were critical to determining whether the comparator (the intake value established by Nawrot et al., in 2003) remained acceptable—that is, the value was unlikely to be associated with adverse effects. Only studies that were published from January 2001–June 2015 were included in this SR, a necessary limitation give the extensive literature on caffeine that continuously appears in the scientific literature.

In taking the SR approach, data were included only if they allowed for direct comparison—that is, they provided data on an intake amount (mg/day or similar) relative to an outcome (findings that only reported on potential associations between caffeine consumption and an outcome, without commenting on a specific dose relative to an endpoint, were not included). Notably, selected data from each study were then utilized to represent findings in the figures presented throughout this review; however, these representations do not incorporate the contextual data and considerations of individual study quality and level of adverseness (discussed below) that are required for development of conclusions. Such an approach was required, given the voluminous nature of the evidence base; detailed reporting of intricacies or interesting findings from each individual study were beyond the scope of this review and discussion. This approach did, however, allow the opportunity to highlight information most relevant to determinations, to assess new areas (endpoints) evaluated since the publication of the study by Nawrot et al. (2003), and to suggest topics for future research.

Similar to the observations of Nawrot et al. (2003), the evidence base was complicated, as each individual study presented design issues (e.g., inadequate measurement of caffeine intake) or variability across outcomes (e.g., inconsistent techniques used to evaluate endpoints or inconsistent control for confounding variables). Despite these issues, useful data were available to evaluate the PECO; however, expert judgement (which included contextual considerations of each dataset) was required on behalf of the analyst and outcome experts were required in developing weight of the evidence conclusions. All data considered are provided via a combination of materials, including this manuscript, the Supplementary Material, and additional supplemental materials available in the AHRQ repository (See Section 2.2 for AHRQ links to individual outcomes).

Overall, the evidence generally supports the findings of Nawrot et al. (2003); however, there were exceptions for some endpoints within most outcomes evaluated. That is, when the total body of evidence was evaluated and study quality, consistency, level of adversity, and magnitude of response were considered, the evidence generally supports that consumption of up to 400 mg caffeine/day in healthy adults is generally not associated with adverse cardiovascular effects, behavioral effects, reproductive effects, acute effects, or bone status. Evidence also supports that a daily consumption of up to 300 mg caffeine/day in healthy pregnant women is generally not associated with adverse reproductive and developmental effects. Very limited data were identified for child and adolescent populations; the available evidence suggests that 2.5 mg caffeine/day remains an acceptable recommendation.

Consistent with Nawrot et al. (2003), as well as other recent reviews (EFSA, 2015; McGuire, 2014; Milanez, 2011), some of the endpoints (blood pressure, anxiety, calcium homeostasis, and, to a lesser extent, sleep) had effects more readily observed at exposures below the comparator than others, indicating the potential that these may be sensitive effects. However, when interpreting these data, considerations were given to the following: the dose and conditions under which such effects were observed, the type of effect (clinical/physiological), and the potential link to downstream effects. These parameters, collectively, aid in characterizing the level of adversity with respect to the individual study, as well as the body of evidence for each endpoint and each outcome. Extensive considerations were given to making *a priori* determinations of

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

adversity during problem formulation and scoping. It was recognized early during the process that in conducting a broad review, there would be a wide array of endpoints to potentially evaluate. Based on expert review (J.G., M.T., C.O., H.L., J.P.) during screening, endpoints considered to be not clearly adverse were excluded. For many endpoints, such as blood pressure, it was difficult to make blanket designations, as the potential for an endpoint to be adverse was conditional; thus, such endpoints were included. However, thresholds or specific parameters under which endpoints were considered to be adverse were not determined *a priori*. Thus, recognizing that the pharmacological effects of caffeine are anticipated to cause physiological changes, characterization of the level of adverseness added complexity to this assessment, because not all physiological changes are adverse.

Given the basic pharmacology of caffeine behavioral changes are to be expected, particularly its primary mechanism of action, blocking adenosine receptors. A good example of such is increased alertness. Under some conditions, such as when caffeine is consumed before bedtime these actions can also be associated with adverse effects such as difficulty sleeping. Thus, differentiation of desirable and undesirable effects (Guyatt et al., 2011) is an important aspect to the interpretation of adversity. Quantitative differentiation of desirable and undesirable effects (including conditions under which these could be characterized) for each of these endpoints would have been a large undertaking in itself, one that was beyond the scope of the assessment. In these cases, considerations of the magnitude of change, as well as the consistency of such relative to downstream effects, aided in informing the level of adversity. For blood pressure as well as other physiological endpoints within the cardiovascular outcome, the magnitude of changes was relatively small (although statistically significant in the studies) and transient in nature, may only affect specific subsets of the population (specific genotypes), and may be subject to tolerance in adult habitual caffeine consumers. In addition, because much of the data for these endpoints, and in particular the physiological endpoints within cardiovascular outcome, were obtained from short-term (often single exposure) controlled trials, the datasets for these endpoints would not provide sufficient evidence to characterize potential long-term effects of caffeine on cardiovascular health. Similar conditions apply to the interpretation of changes in calcium homeostasis; such changes were observed in short-term (single exposure) studies and were not determined to be of consequence to the overall calcium economy.

In the behavior outcome, sleep disruption is an example where a majority of studies indicated the potential for effects at exposures below the comparator depending of course on the time of consumption and dose consumed. The evidence characterizing effects on sleep also highlights the difficulty of characterizing adversity relative to desirable and anticipated effects, considering that caffeine is intentionally ingested to avoid sleepiness and is pharmacologically proven to enhance alertness. With respect to the data in this SR, many of the studies showing effects below the comparator were studies in which there was a deliberate intention of studying the effect of caffeine on sleep. These data support the potential for an effect; however, as discussed by Nawrot et al. (2003), consumers are likely to self-limit caffeine intake to avoid negative effects on sleep. Wesensten (2014) recently discussed this phenomenon, pointing out that caffeine indeed can have a negative impact on sleep and explaining that caffeine is so effective at doing what is intended (blocking sleep) that partakers should expect it reasonable that caffeine could disrupt subsequent sleep. Turnbull et al. (2016) also discuss this concept in a review of neurobehavioral hazard, indicating that the potential effect of caffeine on sleep (particularly when consumed close to normal sleep time) is well known, thus leading many consumers to refrain from consuming caffeinated beverages late in the day.

Anxiety was another endpoint where a majority of studies indicated the potential for effects at exposures below the comparator, although there were a number of inconsistencies between studies that could be a result of the individuals studied (e.g., genetic polymorphisms; Alsene et al., 2003; Childs and de Wit, 2008; Domschke et al., 2012; Rogers et al., 2010). The data from our SR also support the Nawrot et al. (2003) finding that consumer self-regulation and awareness of potential sensitivities (e.g., anxiety disorders) is important for avoiding caffeine-induced anxiety. Nawrot et al. (2003) also pointed out that consideration must be given to the possibility that some of these subjective effects categorized as “anxiety” may also be related to caffeine's ability to increase alertness and arousal. A review by Cappelletti et al. (2015) supports this assertion, emphasizing that these anxiogenic responses tend to occur at higher doses, above the lower concentrations which produce stimulating effects. One's own perception of sensitivity may be very different than another; to this end, the concept of self-limitation remains a research gap. However, in light of inconsistent results in the literature and individual differences in sensitivity to caffeine, some people (e.g., those with anxiety disorders) should be aware of the possible adverse effects of caffeine and should limit their intake accordingly. A SR by Ruxton (2014) focused on children and adolescents and also confirmed that moderate caffeine intake is unlikely to cause harmful effects, also supporting the 2.5 mg/kg/day as an acceptable protective limit. The authors go on to comment that more research should be done at higher doses, as they note that increased anxiety and withdrawal symptoms were manifest at exposures on the order of >5 mg/kg/day. Our review also suggests that this remains a potential research gap (i.e., to investigate caffeine effects at levels >2.5 mg/kg/day on anxiety in children and at >400 mg in adults with preexisting conditions).

Thus, for all of the above reasons, we encourage considerations of magnitude and level of adversity when interpreting data from the figures presented herein, as opposed to solely acknowledging that an “effect” was reported. One can easily see from the collection of figures that effects are manifest below the comparator, many of which are directly associated with the pharmacodynamic properties of caffeine. The goal of the exercise, however, was to go beyond the simple comparisons and to integrate the data using a weight of evidence approach that accommodates the concepts described above (e.g., level of adversity [severity] and consistency; OHAT, 2015a). It is also important to note that while the plots presented throughout are meant to provide a summary graphic of the studies reviewed, they do not necessarily reflect all data presented in each study, nor do they provide an “equal weight” per study. For example, per the internal process implemented, data were selected for graphical representation based on the most refined set of analyses presented by the authors; thus, not all subsets of analyses conducted, or findings at all exposures, may be reflected (note: such data were extracted and can be found via AHRQ; See Section 2.2 for AHRQ links to individual outcomes). Or in several cases, a row on the plot represents a single subgroup or subset of the data (e.g., BMD at a single site versus overall, represented by different rows; see El Maghraoui et al., 2010); as a result, the plot entries should not simply be counted, as they do not equally represent all data points. This underscores one of the challenges in reviewing such a large body of data—that is, it was not possible to discuss or present all aspects of each study, including displaying all findings reported or critically evaluation strengths or limitations within the body of the manuscript.

In evaluating individual studies, several themes in study design elements were observed. One example is the consideration of the pregnancy signal in the context of, primarily, spontaneous abortion,

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

as well as stillbirth and fetal growth (Brent et al., 2011; Lawson et al., 2004; Peck et al., 2010; Stein and Susser, 1991). As discussed in the section on reproduction and development (Section 3.4), this phenomenon relates to the fact that caffeine consumption has been shown to change over the course of pregnancy (Nawrot et al., 2003; Peck et al., 2010). This is important to note because pregnancy symptoms including nausea, aversion to smells or tastes, and vomiting (typically described as morning sickness) are more common in healthy pregnancies that result in live births (Hinkle et al., 2016; Nippita and Dodge, 2016). This "aversion" is associated with what is called the pregnancy signal (Lawson et al., 2004). Research has found that women who have higher concentrations of pregnancy hormones that are known to be associated with healthier pregnancies are more likely to exhibit the aversion habits and practices. By nature of aversion to strong smells (e.g., coffee), such aversions can lead to a decrease in caffeine since it is found in strong-smelling coffee. As such, accounting for the pregnancy signal, although it is complex, has been found to be critical for assessing endpoints such as spontaneous abortion and stillbirth, and potentially for fetal growth, because the symptoms may reflect the status of the pregnancy and resultant influence on caffeine intake levels (Brent et al., 2011; Lawson et al., 2004; Peck et al., 2010; Stein and Susser, 1991). Although most studies on spontaneous abortion (seven of the eight in this review) considered this phenomenon using some measure of nausea as an indicator, the approach for doing so varied and existing measures of the presence or frequency of nausea may not provide a valid assessment of the relevant aspects of aversion (and hence the weighting of this factor in the RoB assessment), which may explain some of variability in evaluation of these endpoints (Hinkle et al., 2016; Nawrot et al., 2003; Nippita and Dodge, 2016; Peck et al., 2010).

Also of importance, but not unique, to the reproductive and developmental body of evidence is the issue of recall bias (i.e., a type of error associated with differences in the accuracy of recollections by cases and controls), as well as the impact of timing on recall bias (Werler et al., 1989, 2011). The observed differences were considered as a weakness across the body of evidence within the SR and were reflected in the quality of evidence assessment. Relevant to future research areas, the work of Werler et al. suggests that an approach for potentially minimizing recall bias in case-control studies of rare endpoints such as childhood cancer or malformations would be to choose a control group from the same source population who also had offspring with other cancers or malformations unrelated to the exposure of interest. Since the effect of caffeine on childhood cancer rate seems to be an emerging area of interest for the outcome of reproductive and developmental effects, well-controlled studies using this guidance may be useful to address concerns of differential accuracy for retrospective reporting of prior caffeine exposures.

One of the largest areas of uncertainties in the underlying body of evidence assessed herein was the characterization of exposure, the findings of which are demonstrated by the (primarily red) color coding on risk of bias questions 8 and 11 (see Figs. 4, 6, 8, 10, and 12). Although not a new complication (Peck et al., 2010), the simple concept that caffeine is consumed via a variety of sources complicates clear evaluation of the exposure-response relationship. Several of the sources included in this SR are complex mixtures with other potentially active compounds and the amount of caffeine within each substance can be highly variable. This is a particular issue for coffee (Mitchell et al., 2014), which was the primary substance evaluated in >20% of studies assessed in this SR. Despite attempts to standardize this metric using our decision tree (a unique strength of this assessment), very few studies validated the amount of caffeine in the respective sources (though validation is generally not feasible in many observational population studies, some studies used validated exposure surveys) or the purity of caffeine (topics that directly impact the risk of bias scoring). It should be noted, however, that the evidence also contains a large number of controlled trials in which exposure was well characterized, although these studies were primarily associated with physiological endpoints. Because the complexities in exposure were well characterized, it was recognized *a priori* that the uncertainties in exposure assessment associated with this research question had to be accepted and attempts were made to address and/or standardize uncertainties. We did not, however, choose to exclude studies based on such given this recognition (the OHAT handbook describes a tiered process in which studies having a high risk of bias for key domains, including exposure, can be excluded) combined with the objective of providing a comprehensive assessment of available literature.

The condition under which these sources are consumed further complicates evaluation of the exposure-response relationship. This SR evaluated consumption of caffeine amounts within a day; however, consistent with the kinetic behavior of caffeine, effects may vary based on how the caffeine is consumed within a day. The most dramatic example in the SR dataset being that lethality events were associated with rapid and excessive consumption of capsules or powders (the comparator for lethality [10 g] is equivalent to ~100 cups of coffee). These findings are supportive of regulatory measures to restrict the number of capsules per entire pack (Thelander et al., 2010), as well as public warnings regarding the dangers of caffeine powder (FDA, 2015). More commonly in the dataset, however, are considerations for consumption of caffeine prior to bedtime, or studies that evaluated repeated exposure in a day. Although these aspects are important for developing conclusions, they were only considered contextually as the overall exposure was standardized to daily intake for comparison purposes. More in-depth assessments of specific endpoints or sensitivities, particularly if data are to be used quantitatively as candidate datasets for health protective values, should consider such conditions.

When determining outcome and overall conclusions, considerations of the level of adversity, or the relative importance (severity) of each endpoint within an outcome, was critical (Eden et al., 2011; Guyatt et al., 2011). These considerations are standard in the practice of SR. In the clinical field, such designations are often evaluated using the GRADE system, applied herein, in which determinations of outcome importance are conducted by categorizing outcomes on a spectrum from limited importance for decision making to critical importance for decision making (Guyatt et al., 2011). When the caffeine data are assessed using such considerations, the evidence regarding the safety of the comparators evaluated herein is stronger for the endpoints categorized as having a higher level of adversity (i.e., clinical effects determined to be of "high" importance to decision making) relative to those considered to have a lower level of adversity. For example, when data from cardiovascular morbidity and mortality (Fig. 5A), reproductive and developmental outcomes (Fig. 9), and nonfatal acute effects (Fig. 11B) are considered, the data more strongly support the appropriateness of the comparator values. The SR datasets on these endpoints also have more data on and above the comparator than other datasets, which increases confidence. In endpoints of lesser adversity, which are primarily physiological and/or of lesser clinical significance, there is a higher frequency of effects at consumption levels below the comparator. It is also notable that while the presence of a dose response relationship was recorded, it was ultimately determined that dose-response data as collected were not a good fit for determining confidence in the body of evidence relative to the research question (see Methods). While demonstration of a dose provides strength in demonstrating a relationship, the objective of this assessment was not to evaluate the

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

presence of a relationship per se, but rather to evaluate the potential for a relationship at or below the comparator intakes. Specifically, it was difficult to integrate the data collected regarding the potential presence of a relationship (at any intake, as recorded in this SR), relative to the specific question under investigation herein (a specific intake). Moreover, the frameworks utilized herein to evaluate the quality of evidence, including the assessment of dose-response, seemed to be better suited for evaluating evidence as it relates to positive findings rather than absence of effects – and in this case, a framework is need that integrates both. These complexities highlights the need for continued development of – as well as flexibility in - guidance and frameworks for the application of toxicology in SR, as many assessment will have a similar focus – that is – characterization of safe levels (vs characterization of potential hazard).

Another complex problem associated with developing conclusions was consideration of habitual or chronic consumption relative to the effect being evaluated; specifically control for issues such in study design or consideration of confounding issues (e.g., withdrawal). Cessation of caffeine use in habitual consumers has been associated with withdrawal. Since the publication by Nawrot et al. (2003), the fifth edition of the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) has been published and caffeine withdrawal became an officially recognized diagnosis in DSM-5. An official diagnosis related to caffeine may be confusing to the majority of people who use it safely every day (Addicott, 2014). Addicott explains that DSM-IV actually recognized caffeine intoxication, which included use of caffeine >250 mg and five or more symptoms (e.g., restlessness, nervousness, insomnia, gastrointestinal distress, and tachycardia) but did not include withdrawal as a disorder because data were lacking. In contrast, DSM-5 now includes withdrawal as a disorder and includes headache as one of the symptoms. Although our review supported that headache was not associated with caffeine at doses <400 mg/day, we did note that headache in the context of withdrawal was sometimes mentioned by the authors in their study design. Headache is often discussed in context with caffeine withdrawal but seems to remain controversial (Dews et al., 2002; Torelli et al., 2009). Nawrot et al. (2003) discussed the cessation of caffeine being associated with headache and also discussed withdrawal as a hypothesis for their onset at both low and high doses. Furthermore, controversy exists and it should be expected that individuals will have variability in severity of symptoms; withdrawal in general is short lived and relatively mild in those affected. Knowledge of potential consequences of withdrawal or conversely the consequences of testing caffeine-naïve individuals should be taken into account by researchers when possible.

In this SR's underlying dataset, control for habitual consumption and/or withdrawal varied greatly, ranging from no consideration of, to cessation prior to participation in controlled trials, to being a main focus of a study evaluating effects (e.g., Brown et al., 2016; Wu et al., 2009). Habituation for caffeine typically results in larger doses needed to produce an effect (Dews et al., 2002). Of interest, habituation has been associated with the observation of U-shaped dose-response curves. For example, in one study, Happonen et al. (2004) offered one interpretation for this observation such that naïve and heavy users of caffeine (>800 mg/day) were at a relatively greater risk of cardiovascular risk than those who habitually consumed moderate levels of caffeine. A review (Zulli et al., 2016) of this topic concluded that recent trends in studies of habitual caffeine consumption and adverse cardiovascular effects (CVD or coronary heart disease) suggest a neutral or even possibly a protective effect of caffeine. A similar conclusion was reached by Cappelletti et al., 2015 who found extensive evidence that there was little to no effect of caffeine on coronary artery disease, myocardial infarction, and stroke. Thus, the role, or consideration, of habituation and withdrawal is an important factor to consider in future research, as well as in characterization of adversity of the data presented herein.

The data in this study also support findings from previous investigations, which demonstrate that there are a variety of kinetic parameters that influence the potential for adverse effects, highlighting interindividual differences in responses. Some of the data reviewed in this SR had emphasis on genotype/phenotype implications and, in particular, differences related to CYP1A2 and COMT genes and subsequent impacts on caffeine metabolism, bone and calcium, or cardiovascular effects. Several studies also investigated polymorphisms in the ADORA2A gene and resulting impacts on anxiety and sleep caused by changes in the adenosine A2A receptor. The health effects of these genotype/phenotype differences for many of these endpoints has also been recently discussed in a review by Pourshahidi et al. (2016). More research is needed, however, to meaningfully translate genotypic variation into public health recommendations. The complexities of these interindividual differences are further compounded by potential differences in PK introduced by the rate of consumption as well as the temperature and manner consumed—areas of interest to future investigations (White et al., 2016).

In the last few years, three large governmental assessments of caffeine have been issued (DGAC, 2015; EFSA, 2015; Milanez, 2011), as well as an updated assessment of carcinogenicity associated with coffee consumption (Loomis et al., 2016). Most recently, EFSA (2015) commissioned an internal SR of literature (Bull et al., 2015) and, based on such, concluded that habitual caffeine consumption up to 400 mg/day does not give rise to safety concerns for nonpregnant women. The US Dietary Guidelines Advisory Committee (DGAC) reached a similar conclusion, stating that "strong and consistent evidence shows that consumption of coffee within the moderate range 3–5 cups/day or up to 400 mg/day caffeine is not associated with increased risk of major chronic diseases, such as CVD and cancer and premature death in healthy adults" (DGAC, 2015). Both of these assessments were issued after our SR was initiated, and the consistency in conclusions provides overall confidence in the original values developed by Health Canada (Nawrot et al., 2003) and assessed herein. EFSA also made additional conclusions, several of which address uncertainties described above. For example, with respect to the conditions under which caffeine is consumed within a day, EFSA concluded that single doses of caffeine of up to 200 mg do not give rise to safety concerns (200 mg was also the amount associated with lack of concern for pregnant and lactating women). The underlying data (Bull et al., 2015) addressed a larger scope than that evaluated herein (e.g., multiple exposure conditions, including coadministration with alcohol or *p*-synephrine or in combination with ingredients found in energy drinks, and different outcomes), although for the research question most similar to that evaluated herein, fewer studies (n = 112) relative to the number of studies evaluated for this SR. It is our speculation that the complexity of trying to account for all of these factors and desire to provide a safe value for all populations likely help to explain the conservatism in the single dose designation.

A key strength of this assessment is its approach: by utilizing a SR approach, we have been both rigorous and transparent in the identification and evaluation of data. In addition, the review was conducted by a multidisciplinary team of subject matter experts, including experts in epidemiology, clinical medicine, and SR. The evidence base was voluminous; conclusions were developed following integration of results from 381 studies. In addition, this SR has a high level of transparency via the figures and tables herein, extensive materials in the AHRQ system, and PROSPERO registration of five protocols prior to the implementation of the review.

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

Several challenges, however, were met in applying the SR approach to a broad array of outcomes and endpoints, which is a topic in itself being addressed in the field, as the best practices for use of SR in the field of toxicology (versus medicine) are still being developed (Stephens et al., 2016; Wikoff and Britt, 2016). Unlike other evidence-based toxicological assessments, this SR focused on a single evidence stream; given the volume of human data available, as well as strength of human data in developing public health guidance, it was determined that other evidence streams involving experimental laboratory studies (e.g., animal studies, mechanistic studies) would not be included. However, evaluation of such data may be useful in future endeavors that aim to very specifically characterize specific effects and, particularly, the underlying mode of action or thresholds in the dose-response relationship. In addition, this SR aids in addressing topics identified in the IOM workshop on caffeine (McGuire, 2014), including exploration of safe caffeine exposure levels and associated boundaries, as well as systematic data collection and analysis.

With respect to further utilization of the data from this SR, it is important to emphasize that we did not set out to identify a new value for caffeine but instead to ascertain whether or not the heavily cited values used in Nawrot et al. (2003) remain acceptable 13 years after publication. If our objective had been to develop an independent value, a different approach would have been implemented. The process and data presented in this SR, however, provide the core elements to such an endeavor. These data comprehensively characterize the body of literature available, allowing for informed selection of candidate datasets—including the identification of critical endpoints as well as characterization of the quality and strength of evidence for each endpoint. These data also provide a foundation for future meta-analyses that could better characterize the dose-response relationship for selected endpoint(s) (Eden et al., 2011; Moher et al., 2009). This approach would likely allow for better evaluation of data above the comparator; in this assessment, data were limited to exposure doses or ranges provided by the authors.

Establishing regulatory boundaries is challenging for an ingredient like caffeine, and this process involves navigation of the spectrum from allowing the consumer to obtain desired effects (typically cognitive or performance related) to avoiding concerning and unwanted adverse events. Furthering the complexity, it is quite widely recognized that perceptions of caffeine's effects and experiences with caffeine's pharmacological impact can differ greatly between individuals. Although it is certainly important to understand sensitive subpopulations and unique cases since 85% of the public apparently seeks caffeine (Mitchell et al., 2014), helping to reassure what level of exposure is safe for the majority of healthy consumers can alleviate undue alarm and allow scientists to shift attention to more meaningful research areas. Since Health Canada's work is so heavily and commonly referenced in nearly every discussion of caffeine safety, we believe there is value to the interested scientist, regulatory body, or even layperson in evaluating whether or not the benchmarks put forward by Health Canada remain acceptable in assuring a reasonable certainty of no harm nearly 13 years since they were developed. Numerous reviews have been written on caffeine and while they often focus on one major endpoint of interest; an approach is understandable and valuable, we believe that part of the reason the Health Canada value continues to be cited is that it more comprehensively takes into account caffeine's potential effects on multiple outcomes. It has served as a reference value aimed at reassuring the typical healthy individual consuming about 400 mg/day that he or she should not expect to experience alarming or unwanted effects.

In conclusion, the results of this SR support the guidance values characterized over a decade ago by Health Canada and reinforce integrative assessments from other authoritative groups (EFSA, 2015). Furthermore, the data evaluated herein represent the most advanced science (i.e., more up-to-date methodology than that in the body of evidence evaluated by Health Canada). Although there are exceptions related to specific endpoints, the evidence generally supports that consumption of up to 400 mg caffeine/day in healthy adults is not associated with overt, adverse cardiovascular, behavioral, reproductive, acute, or bone status effects. Evidence also supports that a daily consumption of up to 300 mg caffeine/day in healthy pregnant women is associated with a general lack of adverse reproductive and developmental effects. Recognizing that individuals may differ in their own level of sensitivity to caffeine, the Health Canada values were originally intended to provide guidance on safe levels of consumption to healthy consumers.

The findings of the SR would further support the safety of standard consumption practices in the United States, as both mean and upper end estimated intakes (mean of 165 mg/day and 90th percentile of 395 mg/day, all ages) are below the comparator value evaluated herein. Findings of this assessment, however, highlight that as established previously—there is no “bright line,” as potential effects are dependent on many conditional factors; further, there is some limited evidence that self-regulation reduces consumption (Griffiths et al., 1986). With regard to child and adolescent populations, limited data were identified; however, based on the available studies reviewed, there is no evidence to suggest a need for a change from the recommendation of 2.5 mg/kg caffeine/day. It should be noted that additional research would be valuable in this area, as well as in other areas identified as having insufficient information in this SR – a finding similar to that of other investigators (e.g., Ruxton, 2014). To that end, the results of this SR support a shift in caffeine research to focus on characterizing effects in sensitive populations and establishing better quantitative characterization of interindividual variability as well as subpopulations (e.g., unhealthy populations, those with preexisting conditions), conditions (e.g., coexposures), and outcomes (e.g., exacerbation of risk-taking behavior) that could render individuals to be at greater risk relative to healthy adults and pregnant women.

Acknowledgments

The authors thank librarian Laurel Graham for informing and reviewing the SR search strategy. The SR was sponsored by the North American Branch of the International Life Sciences Institute (ILSI) Caffeine Working Group. ILSI North America also received unrestricted grants from the American Beverage Association (ABA) and the National Coffee Association (NCA). The two unrestricted grantees received periodic progress reports but did not participate in any aspect of the systematic review. Neither ILSI North America nor the two grantees (ABA and NCA) had any influence over the grading, evaluation, or interpretation of the data collected in this review. ILSI North America is a public, nonprofit foundation that provides a forum for government, academia and industry, to advance understanding of scientific issues related to the nutritional quality and safety of the food supply by sponsoring research programs, educational seminars and workshops, and publications. ILSI North America receives support primarily from its industry membership. ABA is the national trade association that represents the US nonalcoholic beverage industry. NCA is the national trade association that represents the US coffee industry. ToxStrategies and some of the members of the SAB received funds or an honorarium from ILSI North America for conducting the work on the SR of caffeine. ToxStrategies is a private consulting firm providing services on toxicology and risk assessment issues to private and public organizations.

The researchers' scientific conclusion and professional

judgements were not subject to the funders' control; the contents of this manuscript reflect solely the view of the authors.

The opinions or assertions contained herein are the private views of the author(s) and are not to be construed as official or reflecting the views of the Army or the Department of Defense. Any citations of commercial organizations and trade names in this report do not constitute an official Department of the Army endorsement of approval of the products or services of these organizations. The investigators adhered to the policies for protection of human subjects as prescribed by DOD Instruction 3216.02, and the research was conducted in adherence with the provisions of 32 CFR Part 219.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.fct.2017.04.002.

Transparency document

Transparency document related to this article can be found online at http://dx.doi.org/10.1016/j.fct.2017.04.002.

References

Addicott, M.A., 2014. Caffeine use disorder: a review of the evidence and future implications. Curr. Addict. Rep. 1, 186–192. http://dx.doi.org/10.1007/s40429-014-0024-9.

AHRQ, 2013. Grading the Strength of a Body of Evidence when Assessing Health Care Interventions for the Effective Health Care Program of the Agency for Healthcare Research and Quality: an Update. Agency for Healthcare Research and Quality, Rockville, MD.

Ajayi, O.B., Ukwade, M.T., 2001. Caffeine and intraocular pressure in a Nigerian population. J. Glaucoma 10, 25–31. http://dx.doi.org/10.1097/00061198-200102000-00006.

Albrand, G., Munoz, F., Sornay-Rendu, E., DuBoeuf, F., Delmas, P.D., 2003. Independent predictors of all osteoporosis-related fractures in healthy postmenopausal women: the OFELY study. Bone 32, 78–85. http://dx.doi.org/10.1016/S8756-3282(02)00919-5.

Alsene, K., Deckert, J., Sand, P., de Wit, H., 2003. Association between A2a receptor gene polymorphisms and caffeine-induced anxiety. Neuropsychopharmacology 28, 1694–1702. http://dx.doi.org/10.1038/sj.npp.1300232.

Appel, C.C., Myles, T.D., 2001. Caffeine-induced hypokalemic paralysis in pregnancy. Obstet. Gynecol. 97, 805–807. http://dx.doi.org/10.1016/S0029-7844(00)01210-2.

Arciero, P.J., Ormsbee, M.J., 2009. Relationship of blood pressure, behavioral mood state, and physical activity following caffeine ingestion in younger and older women. Appl. Physiol. Nutr. Metab. 34, 754–762. http://dx.doi.org/10.1139/H09-068.

Arnaud, M.J., 2011. Pharmacokinetics and metabolism of natural methylxanthines in animal and man. Handb. Exp. Pharmacol. 33–91. http://dx.doi.org/10.1007/978-3-642-13443-2_3.

Attwood, A.S., Higgs, S., Terry, P., 2007. Differential responsiveness to caffeine and perceived effects of caffeine in moderate and high regular caffeine consumers. Psychopharmacol. Berl. 190, 469–477. http://dx.doi.org/10.1007/s00213-006-0643-5.

Avci, S., Sarikaya, R., Büyükcam, F., 2013. Death of a young man after overuse of energy drink. Am. J. Emerg. Med. 31, 2012–2013. http://dx.doi.org/10.1016/j.ajem.2013.06.031.

Babu, K.M., Zuckerman, M.D., Cherkes, J.K., Hack, J.B., 2011. First-onset seizure after use of an energy drink. Pediatr. Emerg. Care 27, 539–540. http://dx.doi.org/10.1097/PEC.0b013e31821dc72b.

Bakker, R., Steegers, E.A.P., Obradov, A., Raat, H., Hofman, A., Jaddoe, V.W.V., 2010. Maternal caffeine intake from coffee and tea, fetal growth, and the risks of adverse birth outcomes: the Generation R Study. Am. J. Clin. Nutr. 91, 1691–1698. http://dx.doi.org/10.3945/ajcn.2009.28792.

Bakker, R., Steegers, E.A.P., Raat, H., Hofman, A., Jaddoe, V.W.V., 2011. Maternal caffeine intake, blood pressure, and the risk of hypertensive complications during pregnancy. The Generation R Study. Am. J. Hypertens. 24, 421–428. http://dx.doi.org/10.1038/ajh.2010.242.

Balat, O., Balat, A., Ugur, M., Pence, S., 2003. The effect of smoking and caffeine on the fetus and placenta in pregnancy. Clin. Exp. Obs. Gynecol. 30, 57–59.

Barbour, K.E., Zmuda, J.M., Strotmeyer, E.S., Horwitz, M.J., Boudreau, R., Evans, R.W., Ensrud, K.E., Petit, M.A., Gordon, C.L., Cauley, J.A., 2010. Correlates of trabecular and cortical volumetric bone mineral density of the radius and tibia in older men: the osteoporotic fractures in men study. J. Bone Min. Res. 25, 1017–1028. http://dx.doi.org/10.1002/jbmr.6.

Barone, J.J., Roberts, H.R., 1996. Caffeine consumption. Food Chem. Toxicol. 34, 119–129.

Baylin, A., Hernandez-Diaz, S., Kabagambe, E.K., Siles, X., Campos, H., 2006. Transient exposure to coffee as a trigger of a first nonfatal myocardial infarction. Epidemiology 17, 506–511. http://dx.doi.org/10.1097/01.ede.0000229444.55718.96.

Bech, B.H., Autrup, H., Nohr, E.A., Henriksen, T.B., Olsen, J., 2006. Stillbirth and slow metabolizers of caffeine: comparison by genotypes. Int. J. Epidemiol. 35, 948–953. http://dx.doi.org/10.1093/ije/dyl116.

Bech, B.H., Frydenberg, M., Henriksen, T.B., Obel, C., Olsen, J., 2015. Coffee consumption during pregnancy and birth weight: does smoking modify the association? J. Caffeine Res. 5, 65–72. http://dx.doi.org/10.1089/jcr.2015.0001.

Bech, B.H., Nohr, E.A., Vaeth, M., Henriksen, T.B., Olsen, J., 2005. Coffee and fetal death: a cohort study with prospective data. Am. J. Epidemiol. 162, 983–990. http://dx.doi.org/10.1093/aje/kwi317.

Bech, B.H., Obel, C., Henriksen, T.B., Olsen, J., 2007. Effect of reducing caffeine intake on birth weight and length of gestation: randomised controlled trial. BMJ 334, 409. http://dx.doi.org/10.1136/bmj.39062.520648.BE.

Benedum, C.M., Yazdy, M.M., Mitchell, A.A., Werler, M.M., 2013. Risk of spina bifida and maternal cigarette, alcohol, and coffee use during the first month of pregnancy. Int. J. Environ. Res. Public Health 10, 3263–3281. http://dx.doi.org/10.3390/ijerph10083263.

Berger, A.J., Alford, K., 2009. Cardiac arrest in a young man following excess consumption of caffeinated "energy drinks". Med. J. Aust. 190, 41–43 doi:ber10142_fm [pii].

Bertoia, M., Triche, E., 2013. Long-term alcohol and caffeine intake and risk of sudden cardiac death in women. Am. J. Clin. Nutr. 97, 1356–1363. http://dx.doi.org/10.3945/ajcn.112.044248.

Bioh, G., Gallagher, M.M., Prasad, U., 2013. Survival of a highly toxic dose of caffeine. BMJ Case Rep. 2013 http://dx.doi.org/10.1136/bcr-2012-007454.

Bloomer, R.J., Farney, T.M., Harvey, I.C., Alleman, R.J., 2013. Safety profile of caffeine and 1,3-dimethylamylamine supplementation in healthy men. Hum. Exp. Toxicol. 32, 1126–1136. http://dx.doi.org/10.1177/0960327113475680.

Bonaventure, A., Rudant, J., Goujon-Bellec, S., Orsi, L., Leverger, G., Baruchel, A., Bertrand, Y., Nelken, B., Pasquet, M., Michel, G., Sirvent, N., Bordigoni, P., Ducassou, S., Rialland, X., Zelenika, D., Hémon, D., Clavel, J., 2013. Childhood acute leukemia, maternal beverage intake during pregnancy, and metabolic polymorphisms. Cancer Causes Control 24, 783–793. http://dx.doi.org/10.1007/s10552-013-0161-9.

Bonnet, M., Tancer, M., Uhde, T., Yeragani, V.K., 2005. Effects of caffeine on heart rate and QT variability during sleep. Depress. Anxiety 22, 150–155. http://dx.doi.org/10.1002/da.20127.

Bonnet, M.H., Arand, D.L., 2003a. Insomnia, metabolic rate and sleep restoration. J. Intern. Med. 254, 23–31. http://dx.doi.org/10.1046/j.1365-2796.2003.01176.x.

Bonnet, M.H., Arand, D.L., 2003b. Situational insomnia: consistency, predictors, and outcomes. Sleep 26, 1029–1036.

Botella, P., Parra, A., 2003. Coffee increases state anxiety in males but not in females. Hum. Psychopharmacol. 18, 141–143. http://dx.doi.org/10.1002/hup.444.

Bracken, M.B., Triche, E.W., Belanger, K., Hellenbrand, K., Leaderer, B.P., 2003. Association of maternal caffeine consumption with decrements in fetal growth. Am. J. Epidemiol. 157, 456–466. http://dx.doi.org/10.1093/aje/kwf220.

Brathwaite, J.M., Da Costa, L.A., El-Sohemy, A., 2011. Catechol-O-methyltransferase genotype is associated with self-reported increased heart rate following caffeine consumption. J. Caffeine Res. 1, 123–130. http://dx.doi.org/10.1089/caf.2011.0015.

Brent, R.L., Christian, M.S., Diener, R.M., 2011. Evaluation of the reproductive and developmental risks of caffeine. Birth Defects Res. B Dev. Reprod. Toxicol. 92, 152–187. http://dx.doi.org/10.1002/bdrb.20288.

Brett, A.S., 1988. Implications of discordance between clinical impression and toxicology analysis in drug overdose. Arch. Intern. Med. 148, 437–441.

Brice, C.F., Smith, A.P., 2002. Effects of caffeine on mood and performance: a study of realistic consumption. Psychopharmacol. Berl. 164, 188–192. http://dx.doi.org/10.1007/s00213-002-1175-2.

Brown, O.I., Allgar, V., Wong, K.Y.-K., 2016. Coffee reduces the risk of death after acute myocardial infarction: a meta-analysis. Coron. Artery Dis. 27, 566–572. http://dx.doi.org/10.1097/MCA.0000000000000397.

Browne, M.L., Bell, E.M., Druschel, C.M., Gensburg, L.J., Mitchell, A.A., Lin, A.E., Romitti, P.A., Correa, A., National birth defects prevention study, 2007. Maternal caffeine consumption and risk of cardiovascular malformations. Birth Defects Res. A. Clin. Mol. Teratol. 79, 533–543. http://dx.doi.org/10.1002/bdra.20365.

Browne, M.L., Hoyt, A.T., Feldkamp, M.L., Rasmussen, S.A., Marshall, E.G., Druschel, C.M., Romitti, P.A., 2011. Maternal caffeine intake and risk of selected birth defects in the national birth defects prevention study. Birth Defects Res. A Clin. Mol. Teratol. 91, 93–101. http://dx.doi.org/10.1002/bdra.20752.

Bull, S., Brown, T., Burnett, K., Ashdown, L., Rushton, L., 2015. Extensive literature search as preparatory work for the safety assessment for caffeine. EFSA Support. Publ. 12 http://dx.doi.org/10.2903/sp.efsa.2015.EN-561, 561E–n/a.

Calabrò, R.S., Italiano, D., Gervasi, G., Bramanti, P., 2012. Single tonic-clonic seizure after energy drink abuse. Epilepsy Behav. 23, 384–385. http://dx.doi.org/10.1016/j.yebeh.2011.12.010.

Calamaro, C.J., Mason, T.B.A., Ratcliffe, S.J., 2009. Adolescents living the 24/7 lifestyle: effects of caffeine and technology on sleep duration and daytime functioning. Pediatrics 123, e1005–e1010. http://dx.doi.org/10.1542/peds.2008-3641.

Caldeira, D., Martins, C., Alves, L.B., Pereira, H., Ferreira, J.J., Costa, J., 2013. Caffeine

does not increase the risk of atrial fibrillation: a systematic review and meta-analysis of observational studies. Heart 99, 1383–1389. http://dx.doi.org/10.1136/heartjnl-2013-303950.

Campana, C., Griffin, P.L., Simon, E.L., 2014. Caffeine overdose resulting in severe rhabdomyolysis and acute renal failure. Am. J. Emerg. Med. 32, 3–4. http://dx.doi.org/10.1016/j.ajem.2013.08.042.

Cappelletti, S., Daria, P., Sani, G., Aromatario, M., 2015. Caffeine: cognitive and physical performance enhancer or psychoactive drug? Curr. Neuropharmacol. 13, 71–88. http://dx.doi.org/10.2174/1570159X13666141210215655.

CARE Study Group, 2008. Maternal caffeine intake during pregnancy and risk of fetal growth restriction: a large prospective observational study. BMJ 337, a2332. http://dx.doi.org/10.1136/bmj.a2332.

Carrier, J., Fernandez-Bolanos, M., Robillard, R., Dumont, M., Paquet, J., Selmaoui, B., Fillpini, D., 2007. Effects of caffeine are more marked on daytime recovery sleep than on nocturnal sleep. Neuropsychopharmacology 32, 964–972. http://dx.doi.org/10.1038/sj.npp.1301198.

CDC, 2012. Energy drink consumption and its association with sleep problems among U.S. service members on a combat deployment - Afghanistan, 2010. MMWR. Morb. Mortal. Wkly. Rep. 61, 895–898.

Chavarro, J.E., Rich-Edwards, J.W., Rosner, B.A., Willett, W.C., 2009. Caffeinated and alcoholic beverage intake in relation to ovulatory disorder infertility. Epidemiology 20, 374–381. http://dx.doi.org/10.1097/EDE.0b013e31819d68cc.

Chen, L.-W., Wu, Y., Neelakantan, N., Chong, M.F.-F., Pan, A., van Dam, R.M., 2014. Maternal caffeine intake during pregnancy is associated with risk of low birth weight: a systematic review and dose-response meta-analysis. BMC Med. 12, 174. http://dx.doi.org/10.1186/s12916-014-0174-6.

Chen, L., Bell, E.M., Browne, M.L., Druschel, C.M., Romitti, P.A., Schmidt, R.J., Burns, T.L., Moslehi, R., Olney, R.S., 2012. Maternal caffeine consumption and risk of congenital limb deficiencies. Birth Defects Res. A Clin. Mol. Teratol. 94, 1033–1043. http://dx.doi.org/10.1002/bdra.23050.

Cheng, M., Hu, Z., Lu, X., Huang, J., Gu, D., 2014. Caffeine intake and atrial fibrillation incidence: dose response meta-analysis of prospective cohort studies. Can. J. Cardiol. 30, 448–454. http://dx.doi.org/10.1016/j.cjca.2013.12.026.

Childs, E., de Wit, H., 2008. Enhanced mood and psychomotor performance by a caffeine-containing energy capsule in fatigued individuals. Exp. Clin. Psychopharmacol. 16, 13–21. http://dx.doi.org/10.1037/1064-1297.16.1.13.

Christensen, B., Mosdol, A., Retterstol, L., Landaas, S., Thelle, D.S., 2001. Abstention from filtered coffee reduces the concentrations of plasma homocysteine and serum cholesterol—a randomized controlled trial. Am. J. Clin. Nutr. 74, 302–307.

Christian, M.S., Brent, R.L., 2001. Teratogen update: evaluation of the reproductive and developmental risks of caffeine. Teratology 64, 51–78. http://dx.doi.org/10.1002/tera.1047.

Clausson, B., Granath, F., Ekbom, A., Lundgren, S., Nordmark, A., Signorello, L.B., Cnattingius, S., 2002. Effect of caffeine exposure during pregnancy on birth weight and gestational age. Am. J. Epidemiol. 155, 429–436.

Cochrane Collaboration, 2011. Cochrane Handbook for Systematic Reviews of Interventions [WWW Document]. http://handbook.cochrane.org.

Collier, S.A., Browne, M.L., Rasmussen, S.A., Honein, M.A., 2009. Maternal caffeine intake during pregnancy and orofacial clefts. Birth Defects Res. A Clin. Mol. Teratol. 85, 842–849. http://dx.doi.org/10.1002/bdra.20600.

Conen, D., Chiuve, S.E., Everett, B.M., Zhang, S.M., Buring, J.E., Albert, C.M., 2010. Caffeine consumption and incident atrial fibrillation in women123, 92, 509–514. http://dx.doi.org/10.3945/ajcn.2010.29627.

Cornelis, M.C., El-Sohemy, A., Campos, H., 2007. Genetic polymorphism of the adenosine A2A receptor is associated with habitual caffeine consumption. Am. J. Clin. Nutr. 86, 240–244.

Cornelis, M.C., El-Sohemy, A., Kabagambe, E.K., Campos, H., 2006. Coffee, CYP1A2 genotype, and risk of myocardial infarction. JAMA 295, 1135–1141. http://dx.doi.org/10.1001/JAMA.295.10.1135.

Cornelis, M.C., Monda, K.L., Yu, K., Paynter, N., Azzato, E.M., Bennett, S.N., Berndt, S.I., Boerwinkle, E., Chanock, S., Chatterjee, N., Couper, D., Curhan, G., Heiss, G., Hu, F.B., Hunter, D.J., Jacobs, K., Jensen, M.K., Kraft, P., Landi, M.T., Nettleton, J.A., Purdue, M.P., Rajaraman, P., Rimm, E.B., Rose, L.M., Rothman, N., Silverman, D., Stolzenberg-Solomon, R., Subar, A., Yeager, M., Chasman, D.I., van Dam, R.M., Caporaso, N.E., 2011. Genome-wide meta-analysis identifies regions on 7p21 (AHR) and 15q24 (CYP1A2) as determinants of habitual caffeine consumption. PLoS Genet. 7, e1002033. http://dx.doi.org/10.1371/journal.pgen.1002033.

CRD, 2009. CRD's Guidance for Undertaking Reviews in Health Care [WWW Document]. https://www.york.ac.uk/media/crd/Systematic_Reviews.pdf.

Davies, M.J., Judd, J.T., Baer, D.J., Clevidence, B.A., Paul, D.R., Edwards, A.J., Wiseman, S.A., Muesing, R.A., Chen, S.C., 2003. Black tea consumption reduces total and LDL cholesterol in mildly hypercholesterolemic adults. J. Nutr. 133, 3298S–3302S.

De Marco, P., Merello, E., Calevo, M.G., Mascelli, S., Pastorino, D., Crocetti, L., De Biasio, P., Piatelli, G., Cama, A., Capra, V., 2011. Maternal periconceptional factors affect the risk of spina bifida-affected pregnancies: an Italian case-control study. Childs. Nerv. Syst. 27, 1073–1081. http://dx.doi.org/10.1007/s00381-010-1372-y.

Del Brutto, O.H., Mera, R.M., Zambrano, M., 2014. Cardiovascular health and caffeine consumption. A population-based study in rural Ecuador. Int. J. Cardiol. 172, 284–285. http://dx.doi.org/10.1016/j.ijcard.2013.12.237.

Denson, T.F., Jacobson, M., von Hippel, W., Kemp, R.I., Mak, T., 2012. Caffeine expectancies but not caffeine reduce depletion-induced aggression. Psychol. Addict. Behav. 26, 140–144. http://dx.doi.org/10.1037/a0024725.

Dews, P.B., O'Brien, C.P., Bergman, J., 2002. Caffeine: behavioral effects of withdrawal and related issues. Food Chem. Toxicol. 40, 1257–1261.

DGAC, 2015. Scientific Report of the 2015 Dietary Guidelines Advisory Committee, Part B, Chapter 2 [WWW Document]. https://health.gov/dietaryguidelines/2015-scientific-report/04-integration.asp (Accessed 5 August 2015).

Dikici, S., Saritas, A., Kilinc, S., Guneysu, S., Gunes, H., 2015. Does an energy drink cause a transient ischemic attack? Am. J. Emerg. Med. 33, 129.e5–129.e6. http://dx.doi.org/10.1016/j.ajem.2014.06.037.

Domschke, K., Gajewska, A., Winter, B., Herrmann, M.J., Warrings, B., Mühlberger, A., Wosnitza, K., Glotzbach, E., Conzelmann, A., Dlugos, A., Fobker, M., Jacob, C., Arolt, V., Reif, A., Pauli, P., Zwanzger, P., Deckert, J., 2012. ADORA2A gene variation, caffeine, and emotional processing: a multi-level interaction on startle reflex. Neuropsychopharmacology 37, 759–769. http://dx.doi.org/10.1038/npp.2011.253.

Drake, C., Roehrs, T., Shambroom, J., Roth, T., 2013. Caffeine effects on sleep taken 0, 3, or 6 hours before going to bed. J. Clin. Sleep. Med. 9, 1195–1200. http://dx.doi.org/10.5664/jcsm.3170.

Drake, C.L., Jefferson, C., Roehrs, T., Roth, T., 2006. Stress-related sleep disturbance and polysonographic response to caffeine. Sleep. Med. 7, 567–572. http://dx.doi.org/10.1016/j.sleep.2006.03.019.Stress-related.

Drapeau, C., Hamel-Hebert, I., Robillard, R., Selmaoui, B., Filipini, D., Carrier, J., 2006. Challenging sleep in aging: the effects of 200 mg of caffeine during the evening in young and middle-aged moderate caffeine consumers. J. Sleep. Res. 15, 133–141. http://dx.doi.org/10.1111/j.1365-2869.2006.00518.x.

Drewnowski, A., Rehm, C.D., 2016. Sources of caffeine in diets of US children and adults: trends by beverage type and purchase location. Nutrients 8, 154. http://dx.doi.org/10.3390/nu8030154.

Eden, J., Levit, L., Berg, A., Morton, S., 2011. Finding what Works in Health Care: Standards for Systematic Reviews. National Academies Press, Washington, DC.

EFSA, 2015. Scientific Opinion on the safety of caffeine. EFSA J. 13, 4102–4120. http://dx.doi.org/10.2903/j.efsa.2015.4102.

El Maghraoui, A., Ghazi, M., Gassim, S., Ghozlani, I., Mounach, A., Rezqi, A., Dehhaoui, M., 2010. Risk factors of osteoporosis in healthy Moroccan men. BMC Musculoskelet. Disord. 11, 148. http://dx.doi.org/10.1186/1471-2474-11-148.

FDA, 2015. FDA consumer advice on pure powdered caffeine [WWW document]. http://www.fda.gov/Food/RecallsOutbreaksEmergencies/SafetyAlertsAdvisories/ucm405787.htm.

Fitzpatrick-Lewis, D., Ciliska, D., Thomas, H., 2009. The methods for the synthesis of studies without control groups [WWW document]. http://www.nccmt.ca/resources/publications/10.

Floegel, A., Pischon, T., Bergmann, M., Teucher, B., Kaaks, R., Boeing, H., 2012. Coffee consumption and risk of chronic disease in the European prospective investigation into cancer and nutrition (EPIC)—Germany study. Am. J. Clin. Nutr. 95, 901–908. http://dx.doi.org/10.3945/ajcn.111.023648.Am.

Fredholm, B.B., Battig, K., Holmen, J., Nehlig, A., Zvartau, E.E., 1999. Actions of caffeine in the brain with special reference to factors that contribute to its widespread use. Pharmacol. Rev. 51, 83–133.

Frost, L., Vestergaard, P., 2005. Caffeine and risk of atrial fibrillation or flutter: the Danish diet, cancer, and health study. Am. J. Clin. Nutr. 81, 578–582 doi:81/3/578 [pii].

Fulgoni, V.L., Keast, D.R., Lieberman, H.R., 2015. Trends in intake and sources of caffeine in the diets of US adults: 2001-2010. Am. J. Clin. Nutr. 101, 1081–1087. http://dx.doi.org/10.3945/ajcn.113.080077.

Fung, T.T., Arasaratnam, M.H., Grodstein, F., Katz, J.N., Rosner, B., Willett, W.C., Feskanich, D., 2014. Soda consumption and risk of hip fractures in postmenopausal women in the Nurses' Health Study. Am. J. Clin. Nutr. 100, 953–958. http://dx.doi.org/10.3945/ajcn.114.083352.

Gardener, H., Rundek, T., Wright, C.B., Elkind, M.S.V., Sacco, R.L., 2013. Coffee and tea consumption are inversely associated with mortality in a multiethnic urban population. J. Nutr. 143, 1299–1308. http://dx.doi.org/10.3945/jn.112.173807.

George, L., Granath, F., Johansson, A.L.V., Olander, B., Cnattingius, S., 2006. Risks of repeated miscarriage. Paediatr. Perinat. Epidemiol. 20, 119–126. http://dx.doi.org/10.1111/j.1365-3016.2006.00703.x.

Gershon, P., Shinar, D., Ronen, A., 2009. Evaluation of experience-based fatigue countermeasures. Accid. Anal. Prev. 41, 969–975. http://dx.doi.org/10.1016/j.aap.2009.05.012.

Giannelli, M., Doyle, P., Roman, E., Pelerin, M., Hermon, C., 2003. The effect of caffeine consumption and nausea on the risk of miscarriage. Paediatr. Perinat. Epidemiol. 17, 316–323.

Greenberg, J.A., Chow, G., Ziegelstein, R.C., 2008. Caffeinated coffee consumption, cardiovascular disease, and heart valve disease in the elderly (from the Framingham Study). Am. J. Cardiol. 102, 1502–1508. http://dx.doi.org/10.1016/j.amjcard.2008.07.046.

Greenwood, D.C., Alwan, N., Boylan, S., Cade, J.E., Charvill, J., Chipps, K.C., Cooke, M.S., Dolby, V.A., Hay, A.W.M., Kassam, S., Kirk, S.F.L., Konje, J.C., Potdar, N., Shires, S., Simpson, N., Taub, N., Thomas, J.D., Walker, J., White, K.L.M., Wild, C.P., 2010. Caffeine intake during pregnancy, late miscarriage and stillbirth. Eur. J. Epidemiol. http://dx.doi.org/10.1007/s10654-010-9443-7.

Griffiths, R.R., Bigelow, G.E., Liebson, I.A., O'Keeffe, M., O'Leary, D., Russ, N., 1986. Human coffee drinking: manipulation of concentration and caffeine dose. J. Exp. Anal. Behav. 45, 133–148. http://dx.doi.org/10.1901/jeab.1986.45-133.

Grosso, L.M., Rosenberg, K.D., Belanger, K., Saftlas, A.F., Leaderer, B., Bracken, M.B., 2001. Maternal caffeine intake and intrauterine growth retardation. Epidemiology 12, 447–455.

Guyatt, G., Oxman, A.D., Akl, E.A., Kunz, R., Vist, G., Brozek, J., Norris, S., Falck-Ytter, Y., Glasziou, P., DeBeer, H., Jaeschke, R., Rind, D., Meerpohl, J., Dahm, P.,

Schunemann, H.J., 2011. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J. Clin. Epidemiol. 64, 383–394. http://dx.doi.org/10.1016/j.jclinepi.2010.04.026.

Hagen, K., Thoresen, K., Stovner, L.J., Zwart, J.A., 2009. High dietary caffeine consumption is associated with a modest increase in headache prevalence: results from the Head-HUNT Study. J. Headache Pain 10, 153–159. http://dx.doi.org/10.1007/s10194-009-0114-6.

Hallstrom, H., Byberg, L., Glynn, A., Lemming, E.W., Wolk, A., Michaelsson, K., 2013. Long-term coffee consumption in relation to fracture risk and bone mineral density in women. Am. J. Epidemiol. 178, 898–909. http://dx.doi.org/10.1093/aje/kwt062.

Hallstrom, H., Melhus, H., Glynn, A., Lind, L., Syvänen, A.-C., Michaëlsson, K., 2010. Coffee consumption and CYP1A2 genotype in relation to bone mineral density of the proximal femur in elderly men and women: a cohort study. Nutr. Metab. (Lond) 7, 12. http://dx.doi.org/10.1186/1743-7075-7-12.

Hallstrom, H., Wolk, A., Glynn, A., Michaëlsson, K., 2006. Coffee, tea and caffeine consumption in relation to osteoporotic fracture risk in a cohort of Swedish women. Osteoporos. Int. 17, 1055–1064. http://dx.doi.org/10.1007/s00198-006-0109-y.

Happonen, P., Läärä, E., Hiltunen, L., Luukinen, H., 2008. Coffee consumption and mortality in a 14-year follow-up of an elderly northern Finnish population. Br. J. Nutr. 99, 1354–1361. http://dx.doi.org/10.1017/S0007114507871650.

Happonen, P., Voutilainen, S., Salonen, J.T., 2004. Coffee drinking is dose-dependently related to the risk of acute coronary events in middle-aged men. J. Nutr. 134, 2381–2386.

Happonen, P., Voutilainen, S., Tuomainen, T.P., Salonen, J.T., 2006. Catechol-O-methyltransferase gene polymorphism modifies the effect of coffee intake on incidence of acute coronary events. PLoS One 1, e117. http://dx.doi.org/10.1371/journal.pone.0000117.

Harrell, P.T., Juliano, L.M., 2009. Caffeine expectancies influence the subjective and behavioral effects of caffeine. Psychopharmacol. Berl. 207, 335–342. http://dx.doi.org/10.1007/s00213-009-1658-5.

Harter, D.L., Busnello, F.M., Dibi, R.P., Stein, A.T., Kato, S.K., Vanin, C.M.D.M., 2013. Association between low bone mass and calcium and caffeine intake among perimenopausal women in Southern Brazil: cross-sectional study. Sao Paulo Med. J. 131, 315–322. http://dx.doi.org/10.1590/1516-3180.2013.1315428.

Haskell, C.F., Kennedy, D.O., Milne, A.L., Wesnes, K.A., Scholey, A.B., 2008. The effects of l-theanine, caffeine and their combination on cognition and mood. Biol. Psychol. 77, 113–122. http://dx.doi.org/10.1016/j.biopsycho.2007.09.008.

Hatch, E.E., Wise, L.A., Mikkelsen, E.M., Christensen, T., Riis, A.H., Sørensen, H.T., Rothman, K.J., 2012. Caffeinated beverage and soda consumption and time to pregnancy. Epidemiology 23, 393–401. http://dx.doi.org/10.1097/EDE.0b013e31824cbaac.

Heaney, R.P., Rafferty, K., 2001. Carbonated beverages and urinary calcium excretion. Am. J. Clin. Nutr. 74, 343–347.

Heatherley, S.V., Hancock, K.M.F., Rogers, P.J., 2006. Psychostimulant and other effects of caffeine in 9- to 11-year-old children. J. Child. Psychol. Psychiatry Allied Discip. 47, 135–142. http://dx.doi.org/10.1111/j.1469-7610.2005.01457.x.

Hering-Hanit, R., Gadoth, N., 2003. Caffeine-induced headache in children and adolescents. Cephalalgia 23, 332–335. http://dx.doi.org/10.1046/j.1468-2982.2003.00576.x.

Hewlett, P., Smith, A., 2007. Effects of repeated doses of caffeine on performance and alertness: new data and secondary analyses. Hum. Psychopharmacol. Clin. Exp. 22, 339–350. http://dx.doi.org/10.1002/hup.854.

Heyerdahl, F., Hovda, K.E., Bjornaas, M.A., Brors, O., Ekeberg, O., Jacobsen, D., 2008. Clinical assessment compared to laboratory screening in acutely poisoned patients. Hum. Exp. Toxicol. 27, 73–79. http://dx.doi.org/10.1177/0960327107087800.

Hinkle, S.N., Mumford, S.L., Grantz, K.L., Silver, R.M., Mitchell, E.M., Sjaarda, L.A., Radin, R.G., Perkins, N.J., Galai, N., Schisterman, E.F., 2016. Association of nausea and vomiting during pregnancy with pregnancy loss: a secondary analysis of a randomized clinical trial. JAMA Intern. Med. 176, 1621–1627. http://dx.doi.org/10.1001/jamainternmed.2016.5641.

Ho, S.C., Chung, J.W.Y., 2013. The effects of caffeine abstinence on sleep: a pilot study. Appl. Nurs. Res. 26, 80–84. http://dx.doi.org/10.1016/j.apnr.2012.08.004.

Holmgren, P., Nordén-Pettersson, L., Ahlner, J., 2004. Caffeine fatalities: four case reports. Forensic Sci. Int. 139, 71–73. http://dx.doi.org/10.1016/j.forsciint.2003.09.019.

Holstege, C.P., Hunter, Y., Baer, A.B., Savory, J., Bruns, D.E., Boyd, J.C., 2003. Massive caffeine overdose requiring vasopressin infusion and hemodialysis. J. Toxicol. Clin. Toxicol. 41, 1003–1007. http://dx.doi.org/10.1081/CLT-120026526.

Hoyt, A.T., Browne, M., Richardson, S., Romitti, P., Druschel, C., 2014. Maternal caffeine consumption and small for gestational age births: results from a population-based case-control study. Matern. Child. Health J. 18, 1540–1551. http://dx.doi.org/10.1007/s10995-013-1397-4.

IARC 1991. IARC Monographs on the evaluation of carcinogenic risks to humans. Volume 51: coffee, tea, matte, methylxanthines and methylgloyoxal.

Infante-Rivard, C., 2007. Caffeine intake and small-for-gestational-age birth: modifying effects of xenobiotic-metabolising genes and smoking. Paediatr. Perinat. Epidemiol. 21, 300–309.

Ishigaki, S., Fukasawa, H., Kinoshita-Katahashi, N., Yasuda, H., Kumagai, H., Furuya, R., 2014. Caffeine intoxication successfully treated by hemoperfusion and hemodialysis. Intern. Med. 53, 2745–2747. http://dx.doi.org/10.2169/internalmedicine.53.2882.

Institute of Medicine (US) Committee on Military Nutrition Research, 2001. Caffeine for the Sustainment of Mental Task Performance: Formulations for Military Operations. National Academies Press (US), Washington (DC). http://dx.doi.org/10.17226/10219.

Jabbar, S.B., Hanly, M.G., 2013. Fatal caffeine overdose. Am. J. Forensic Med. Pathol. 34, 321–324. http://dx.doi.org/10.1097/PAF.0000000000000058.

James, J.E., Gregg, M.E., 2004a. Effects of dietary caffeine on mood when rested and sleep restricted. Hum. Psychopharmacol. 19, 333–341. http://dx.doi.org/10.1002/hup.589.

James, J.E., Gregg, M.E., 2004b. Hemodynamic effects of dietary caffeine, sleep restriction, and laboratory stress. Psychophysiology 41, 914–923. http://dx.doi.org/10.1111/j.1469-8986.2004.00248.x.

Jaussent, I., Dauvilliers, Y., Ancelin, M.-L., Dartigues, J.-F., Tavernier, B., Touchon, J., Ritchie, K., Besset, A., 2011. Insomnia symptoms in older adults: associated factors and gender differences. Am. J. Geriatr. Psychiatry 19, 88–97. http://dx.doi.org/10.1097/JGP.0b013e3181e049b6.

Jensen, T.K., Swan, S.H., Skakkebaek, N.E., Rasmussen, S., Jørgensen, N., 2010. Caffeine intake and semen quality in a population of 2,554 young Danish men. Am. J. Epidemiol. 171, 883–891. http://dx.doi.org/10.1093/aje/kwq007.

Jha, R.M., Mithal, A., Malhotra, N., Brown, E.M., 2010. Pilot case-control investigation of risk factors for hip fractures in the urban Indian population. BMC Musculoskelet. Disord. 11, 49. http://dx.doi.org/10.1186/1471-2474-11-49.

Johansen, A.M.W., Wilcox, A.J., Lie, R.T., Andersen, L.F., Drevon, C.A., 2009. Maternal consumption of coffee and caffeine-containing beverages and oral clefts: a population-based case-control study in Norway. Am. J. Epidemiol. 169, 1216–1222. http://dx.doi.org/10.1093/aje/kwp040.

Jonjev, Z.S., Bala, G., 2013. High-energy drinks may provoke aortic dissection. Coll. Antropol. 37 (Suppl. 2), 227–229.

Jordan, J.B., Korgaokar, A., Farley, R.S., Coons, J.M., Caputo, J.L., 2014. Caffeine supplementation and reactive agility in elite youth soccer players. Pediatr. Exerc. Sci. 26, 168–176. http://dx.doi.org/10.1123/pes.2013-0134.

Josse, A.R., Da Costa, L.A., Campos, H., El-Sohemy, A., 2012. Associations between polymorphisms in the AHR and CYP1A1-CYP1A2 gene regions and habitual caffeine consumption. Am. J. Clin. Nutr. 96, 665–671. http://dx.doi.org/10.3945/ajcn.112.038794.

Judice, P.B., Magalhaes, J.P., Santos, D. a, Matias, C.N., Carita, A.I., Armada-Da-Silva, P. a S., Sardinha, L.B., Silva, A.M., 2013. A moderate dose of caffeine ingestion does not change energy expenditure but decreases sleep time in physically active males: a double-blind randomized controlled trial. Appl. Physiol. Nutr. Metab. 38, 49–56. http://dx.doi.org/10.1139/apnm-2012-0145.

Juliano, L.M., Huntley, E.D., Harrell, P.T., Westerman, A.T., 2012. Development of the Caffeine Withdrawal Symptom Questionnaire: caffeine withdrawal symptoms cluster into 7 factors. Drug Alcohol Depend. 124, 229–234. http://dx.doi.org/10.1016/j.drugalcdep.2012.01.009.

Kabagambe, E.K., Baylin, A., Campos, H., 2007. Nonfatal acute myocardial infarction in Costa Rica: modifiable risk factors, population-attributable risks, and adherence to dietary guidelines. Circulation 115, 1075–1081. http://dx.doi.org/10.1161/CIRCULATIONAHA.106.643544.

Kancherla, V., Romitti, P.A., Sun, L., Carey, J.C., Burns, T.L., Siega-Riz, A.M., Druschel, C.M., Lin, A.E., Olney, R.S., National Birth Defects Prevention Study, 2014. Descriptive and risk factor analysis for choanal atresia: the National Birth Defects Prevention Study, 1997-2007. Eur. J. Med. Genet. 57, 220–229. http://dx.doi.org/10.1016/j.ejmg.2014.02.010.

Kapur, R., Smith, M.D., 2009. Treatment of cardiovascular collapse from caffeine overdose with lidocaine, phenylephrine, and hemodialysis. Am. J. Emerg. Med. 27, 3–6. http://dx.doi.org/10.1016/j.ajem.2008.06.028.

Karapetian, G.K., Engels, H.J., Gretebeck, K.A., Gretebeck, R.J., 2012. Effect of caffeine on LT, VT and HRVT. Int. J. Sports Med. 33, 507–513. http://dx.doi.org/10.1055/s-0032-1301904.

Karypidis, A.H., Soderstrom, T., Nordmark, A., Granath, F., Cnattingius, S., Rane, A., 2006. Association of cytochrome P450 1B1 polymorphism with first-trimester miscarriage. Fertil. Steril. 86, 1498–1503. http://dx.doi.org/10.1016/j.fertnstert.2006.03.059.

Keenan, E.K., Tiplady, B., Priestley, C.M., Rogers, P.J., 2014. Naturalistic effects of five days of bedtime caffeine use on sleep, next-day cognitive performance, and mood. J. caffeine Res. 4, 13–20. http://dx.doi.org/10.1089/jcr.2011.0030.

Kempf, K., Herder, C., Erlund, I., 2010. Effects of coffee consumption on subclinical inflammation and other risk factors for type 2 diabetes: a clinical trial. Am. J. Clin. Nutr. 91, 950–957. http://dx.doi.org/10.3945/ajcn.2009.28548.

Keogh, E., Chaloner, N., 2002. The moderating effect of anxiety sensitivity on caffeine-induced hypoalgesia in healthy women. Psychopharmacol. Berl. 164, 429–431. http://dx.doi.org/10.1007/s00213-002-1280-2.

Killen, L.G., Green, J.M., O'Neal, E.K., McIntosh, J.R., Hornsby, J., Coates, T.E., 2013. Effects of caffeine on session ratings of perceived exertion. Eur. J. Appl. Physiol. 113, 721–727. http://dx.doi.org/10.1007/s00421-012-2480-z.

Kinney, A., Kline, J., Levin, B., 2006. Alcohol, caffeine and smoking in relation to age at menopause. Maturitas 54, 27–38.

Krall, E.A., Garvey, A.J., Garcia, R.I., 2002. Smoking relapse after 2 years of abstinence: findings from the VA normative aging study. Nicotine Tob. Res. 4, 95–100. http://dx.doi.org/10.1080/14622200110098428.

Kristjansson, A.L., Sigfusdottir, I.D., Mann, M.J., James, J.E., 2014. Caffeinated sugar-sweetened beverages and common physical complaints in Icelandic children aged 10-12years. Prev. Med. Balt. 58, 40–44. http://dx.doi.org/10.1016/j.ypmed.2013.10.011.

Larsson, S.C., Virtamo, J., Wolk, A., 2011. Coffee consumption and risk of stroke in women. Stroke 42, 908–912. http://dx.doi.org/10.1161/STROKEAHA.110.603787.

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

Lawson, C.C., LeMasters, G.K., Wilson, K.A., 2004. Changes in caffeine consumption as a signal of pregnancy. Reprod. Toxicol. 18, 625–633. http://dx.doi.org/10.1016/j.reprotox.2004.03.004.

Lee, D.R., Lee, J., Rota, M., Lee, J., Ahn, H.S., Park, S.M., Shin, D., 2014. Coffee consumption and risk of fractures: a systematic review and dose–response meta-analysis. Bone 63, 20–28. http://dx.doi.org/10.1016/j.bone.2014.02.007.

Lee, Y., Moon, S.J., Montell, C., 2009. Multiple gustatory receptors required for the caffeine response in Drosophila. Proc. Natl. Acad. Sci. U. S. A. 106, 4495–4500. http://dx.doi.org/10.1073/pnas.0811744106.

Lieberman, H.R., 2001. The effects of ginseng, ephedrine, and caffeine on cognitive performance, mood and energy. Nutr. Rev. 59, 91–102.

Liguori, A., Robinson, J.H., 2001. Caffeine antagonism of alcohol-induced driving impairment. Drug Alcohol Depend. 63, 123–129. http://dx.doi.org/10.1016/S0376-8716(00)00196-4.

Linnet, K.M., Wisborg, K., Secher, N.J., Thomsen, P.H., Obel, C., Dalsgaard, S., Henriksen, T.B., 2009. Coffee consumption during pregnancy and the risk of hyperkinetic disorder and ADHD: a prospective cohort study. Acta Paediatr. 98, 173–179. http://dx.doi.org/10.1111/j.1651-2227.2008.00980.x.

Loomans, E.M., Hofland, L., van der Stelt, O., van der Wal, M.F., Koot, H.M., Van den Bergh, B.R.H., Vrijkotte, T.G.M., 2012. Caffeine intake during pregnancy and risk of problem behavior in 5- to 6-year-old children. Pediatrics 130. http://dx.doi.org/10.1542/peds.2011-3361 e305–13.

Loomis, D., Guyton, K.Z., Grosse, Y., Lauby-Secretan, B., El Ghissassi, F., Bouvard, V., Benbrahim-Tallaa, L., Guha, N., Mattock, H., Straif, K., 2016. Carcinogenicity of drinking coffee, mate, and very hot beverages. Lancet. Oncol. 17, 877–878. http://dx.doi.org/10.1016/S1470-2045(16)30239-X.

Lopez-Garcia, E., van Dam, R.M., Li, T.Y., Rodriguez-Artalejo, F., Hu, F.B., 2008. The relationship of coffee consumption with mortality. Ann. Intern. Med. 148, 904–914.

Lucas, M., Mirzaei, F., Pan, A., Okereke, O.I., Willett, W.C., O'Reilly, É.J., Koenen, K., Ascherio, A., 2011. Coffee, caffeine, and risk of depression among women. Arch. Intern. Med. 171, 1571–1578. http://dx.doi.org/10.1001/archinternmed.2011.393.

Lucas, M., O'Reilly, E.J., Pan, A., Mirzaei, F., Willett, W.C., Okereke, O.I., Ascherio, A., 2014. Coffee, caffeine, and risk of completed suicide: results from three prospective cohorts of American adults. World J. Biol. Psychiatry 15, 377–386. http://dx.doi.org/10.3109/15622975.2013.795243.

Maconochie, N., Doyle, P., Prior, S., Simmons, R., 2007. Risk factors for first trimester miscarriage–results from a UK-population-based case-control study. BJOG 114, 170–186. http://dx.doi.org/10.1111/j.1471-0528.2006.01193.x.

Mancia, G., 2012 Aug. Short- and long-term blood pressure variability: present and future. Hypertension 60 (2), 512–517. http://dx.doi.org/10.1161/HYPERTENSIONAHA.112.194340.

Marczinski, C.A., Stamates, A.L., Ossege, J., Maloney, S.F., Bardgett, M.E., Brown, C.J., 2014. Subjective state, blood pressure, and behavioral control changes produced by an "energy shot". J. Caffeine Res. 4, 57–63. http://dx.doi.org/10.1089/jcr.2014.0005.

Maslova, E., Bhattacharya, S., Lin, S.-W., Michels, K.B., 2010. Caffeine consumption during pregnancy and risk of preterm birth: a meta-analysis. Am. J. Clin. Nutr. 92, 1120–1132. http://dx.doi.org/10.3945/ajcn.2010.29789.

Matijasevich, A., Barros, F.C., Santos, I.S., Yemini, A., 2006. Maternal caffeine consumption and fetal death: a case-control study in Uruguay. Paediatr. Perinat. Epidemiol. 20, 100–109. http://dx.doi.org/10.1111/j.1365-3016.2006.00706.x.

McGuire, S., 2014. Institute of Medicine. 2014. Caffeine in Food and Dietary Supplements: Examining Safety-Workshop Summary. Washington, DC: The National Academies Press, 2014. Adv. Nutr. 5, 585–586.

McLean, C., Graham, T.E., 2002. Effects of exercise and thermal stress on caffeine pharmacokinetics in men and eumenorrheic women. J. Appl. Physiol. 93, 1471–1478. http://dx.doi.org/10.1152/japplphysiol.00762.2000.

McLellan, T.M., Caldwell, J.A., Lieberman, H.R., 2016 Dec. A review of caffeine's effects on cognitive, physical and occupational performance. Neurosci. Biobehav Rev. 71, 294–312. http://dx.doi.org/10.1016/j.neubiorev.2016.09.001.

McNair, P., Lorr, M., Droppleman, L., 1981. POMS Manual, second ed. Education and Industrial Testing Service, San Diego, CA.

Menegaux, F., Ripert, M., Hémon, D., Clavel, J., 2007. Maternal alcohol and coffee drinking, parental smoking and childhood leukaemia: a French population-based case-control study. Paediatr. Perinat. Epidemiol. 21, 293–299. http://dx.doi.org/10.1111/j.1365-3016.2007.00824.x.

Milanez, S., 2011. Adverse Health Effects of Caffeine: Review and Analysis of Recent Human and Animal Research. Oak Ridge National Laboratory, Oak Ridge, TN.

Millen, B.E., Abrams, S., Adams-Campbell, L., Anderson, C.A., Brenna, J.T., Campbell, W.W., Clinton, S., Hu, F., Nelson, M., Neuhouser, M.L., Perez-Escamilla, R., Siega-Riz, A.M., Story, M., Lichtenstein, A.H., 2016. The 2015 dietary guidelines advisory committee scientific report: development and major conclusions. Adv. Nutr. 7, 438–444. http://dx.doi.org/10.3945/an.116.012120.

Miller, E.A., Manning, S.E., Rasmussen, S.A., Reefhuis, J., Honein, M.A., 2009. Maternal exposure to tobacco smoke, alcohol and caffeine, and risk of anorectal atresia: national Birth Defects Prevention Study 1997-2003. Paediatr. Perinat. Epidemiol. http://dx.doi.org/10.1111/j.1365-3016.2008.00976.x.

Mineharu, Y., Koizumi, A., Wada, Y., Iso, H., Watanabe, Y., Date, C., Yamamoto, A., Kikuchi, S., Inaba, Y., Toyoshima, H., Kondo, T., Tamakoshi, A., JACC Study Group, 2011. Coffee, green tea, black tea and oolong tea consumption and risk of mortality from cardiovascular disease in Japanese men and women. J. Epidemiol. Community Health 65, 230–240. http://dx.doi.org/10.1136/jech.2009.097311.

Mitchell, D.C., Hockenberry, J., Teplansky, R., Hartman, T.J., 2015. Assessing dietary exposure to caffeine from beverages in the U.S. population using brand-specific versus category-specific caffeine values. Food Chem. Toxicol. 80, 247–252. http://dx.doi.org/10.1016/j.fct.2015.03.024.

Mitchell, D.C., Knight, C.A., Hockenberry, J., Teplansky, R., Hartman, T.J., 2014. Beverage caffeine intakes in the US. Food Chem. Toxicol. 63, 136–142. http://dx.doi.org/10.1016/j.fct.2013.10.042.

Miyai, N., Arita, M., Miyashita, K., Morioka, I., Shiraishi, T., Nishio, I., 2002 Mar 1. Blood pressure response to heart rate during exercise test and risk of future hypertension. Hypertension 39 (3), 761–766.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 6, e1000097. http://dx.doi.org/10.1371/journal.pmed.1000097.

Money, C.D., Tomenson, J.A., Penman, M.G., Boogaard, P.J., Jeffrey Lewis, R., 2013. A systematic approach for evaluating and scoring human data. Regul. Toxicol. Pharmacol. 66, 241–247. http://dx.doi.org/10.1016/j.yrtph.2013.03.011.

Mongraw-Chaffin, M.L., Cohn, B.A., Cohen, R.D., Christianson, R.E., 2008. Maternal smoking, alcohol consumption, and caffeine consumption during pregnancy in relation to a son's risk of persistent cryptorchidism: a prospective study in the Child Health and Development Studies cohort, 1959-1967. Am. J. Epidemiol. 167, 257–261. http://dx.doi.org/10.1093/aje/kwm311.

Mostofsky, E., Rice, M.S., Levitan, E.B., Mittleman, M.A., 2012. Habitual coffee consumption and risk of heart failure: a dose response meta-analysis. Circ. Hear. Fail 5, 401–405. http://dx.doi.org/10.1161/CIRCHEARTFAILURE.112.967299.Habitual.

Mostofsky, E., Schlaug, G., Mukamal, K.J., Rosamond, W.D., Mittleman, M.A., 2010. Coffee and acute ischemic stroke onset. Neurology 75, 1583–1588. http://dx.doi.org/10.1212/WNL.0b013e3181fb443d.

Motl, R.W., Dishman, R.K., 2004. Effects of acute exercise on the soleus H-reflex and self-reported anxiety after caffeine ingestion. Physiol. Behav. 80, 577–585. http://dx.doi.org/10.1016/j.physbeh.2003.10.015.

Mougios, V., Ring, S., Petridou, A., Nikolaidis, M.G., 2003. Duration of coffee- and exercise-induced changes in the fatty acid profile of human serum. J. Appl. Physiol. 94, 476–484. http://dx.doi.org/10.1152/japplphysiol.00624.2002.

Mozaffarian, D., Benjamin, E.J., Go, A.S., Arnett, D.K., Blaha, M.J., Cushman, M., Das, S.R., de Ferranti, S., Després, J.-P., Fullerton, H.J., Howard, V.J., Huffman, M.D., Isasi, C.R., Jiménez, M.C., Judd, S.E., Kissela, B.M., Lichtman, J.H., Lisabeth, L.D., Liu, S., Mackey, R.H., Magid, D.J., McGuire, D.K., Mohler, E.R., Moy, C.S., Muntner, P., Mussolino, M.E., Nasir, K., Neumar, R.W., Nichol, G., Palaniappan, L., Pandey, D.K., Reeves, M.J., Rodriguez, C.J., Rosamond, W., Sorlie, P.D., Stein, J., Towfighi, A., Turan, T.N., Virani, S.S., Woo, D., Yeh, R.W., Turner, M.B., 2016. Heart disease and stroke Statistics-2016 update: a report from the American heart association. Circulation 133, e38–e360. http://dx.doi.org/10.1161/CIR.0000000000000350.

Mudge, D.W., Johnson, D.W., 2004. Coca-Cola and kangaroos. Lancet 364, 1190. http://dx.doi.org/10.1016/S0140-6736(04)17111-0.

Namdar, M., Koepfli, P., Grathwohl, R., Siegrist, P.T., Klainguti, M., Schepis, T., Delaloye, R., Wyss, C.A., Fleischmann, S.P., Gaemperli, O., Kaufmann, P.A., 2006. Caffeine decreases exercise-induced myocardial flow reserve. J. Am. Coll. Cardiol. 47, 405–410. http://dx.doi.org/10.1016/j.jacc.2005.08.064.

Nardi, A.E., Lopes, F.L., Freire, R.C., Veras, A.B., Nascimento, I., Valença, A.M., de-Melo-Neto, V.L., Soares-Filho, G.L., King, A.L., Araújo, D.M., Mezzasalma, M.A., Rassi, A., Zin, W.A., 2009. Panic disorder and social anxiety disorder subtypes in a caffeine challenge test. Psychiatry Res. 169, 149–153. http://dx.doi.org/10.1016/j.psychres.2008.06.023.

Nardi, A.E., Lopes, F.L., Valença, A.M., Freire, R.C., Veras, A.B., de-Melo-Neto, V.L., Nascimento, I., King, A.L., Mezzasalma, M.A., Soares-Filho, G.L., Zin, W.A., 2007. Caffeine challenge test in panic disorder and depression with panic attacks. Compr. Psychiatry 48, 257–263. http://dx.doi.org/10.1016/j.comppsych.2006.12.001.

Nawrot, P., Jordan, S., EastNawrot, P., Jordan, S., Eastwood, J., Rotstein, J., Hugenholtz, A., Feeley, M., 2003. Effects of caffeine on human health. Food Addit. Contam. 20, 1–30. http://dx.doi.org/10.1080/0265203021000007840.

Nippita, S., Dodge, L.E., 2016. Toward a deeper understanding of nausea, vomiting, and pregnancy loss. JAMA Intern. Med. 176, 1628–1629. http://dx.doi.org/10.1001/jamainternmed.2016.6101.

Nishijima, Y., Ikeda, T., Takamatsu, M., Kiso, Y., Shibata, H., Fushiki, T., Moritani, T., 2002. Influence of caffeine ingestion on autonomic nervous activity during endurance exercise in humans. Eur. J. Appl. Physiol. 87, 475–480. http://dx.doi.org/10.1007/s00421-002-0678-1.

Noordzij, M., Uiterwaal, C., Arends, L., Kok, F., Grobbee, D., Geleijnse, J., 2005. Blood pressure response to chronic intake of coffee and caffeine: a meta-analysis of randomized controlled trials. J. Hypertens. 23, 921–928.

Norcross, J.C., Guadagnoli, E., Prochaska, J.O., 1984. Factor structure of the Profile of Mood States (POMS): two partial replications. J. Clin. Psychol. 40, 1270–1277.

Ogawa, N., Ueki, H., 2007. Clinical importance of caffeine dependence and abuse. Psychiatry Clin. Neurosci. 61, 263–268. http://dx.doi.org/10.1111/j.1440-1819.2007.01652.x.

OHAT, 2015a. Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration [WWW Document]. https://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookjan2015_508.pdf.

OHAT, 2015b. OHAT risk of bias rating tool for human and animal studies [WWW document]. https://ntp.niehs.nih.gov/ntp/ohat/pubs/riskofbiastool_508.pdf.

Omvik, S., Pallesen, S., Bjorvatn, B., Thayer, J., Hilde Nordhus, I., 2007. Night-time thoughts in high and low worriers: reaction to caffeine-induced sleeplessness.

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

Behav. Res. Ther. 45, 715–727. http://dx.doi.org/10.1016/j.brat.2006.06.006.
Ørskou, J., Henriksen, T.B., Kesmodel, U., Secher, N.J., 2003. Maternal characteristics and lifestyle factors and the risk of delivering high birth weight infants. Obstet. Gynecol. 102, 115–120. http://dx.doi.org/10.1016/S0029-7844(03)00402-2115.
Ozsungur, S., Brenner, D., El-Sohemy, A., 2009. Fourteen well-described caffeine withdrawal symptoms factor into three clusters. Psychopharmacol. Berl. 201, 541–548. http://dx.doi.org/10.1007/s00213-008-1329-y.
Paganini-Hill, A., 2011. Lifestyle practices and cardiovascular disease mortality in the elderly: the Leisure World Cohort Study. Cardiol. Res. Pract. 2011, 983764. http://dx.doi.org/10.4061/2011/983764.
Pallarés, J.G., Fernández-Elías, V.E., Ortega, J.F., Muñoz, G., Muñoz-Guerra, J., Mora-Rodríguez, R., 2013. Neuromuscular responses to incremental caffeine doses: performance and side effects. Med. Sci. Sports Exerc 45, 2184–2192. http://dx.doi.org/10.1249/MSS.0b013e31829a6672.
Paterson, L.M., Nutt, D.J., Ivarsson, M., Hutson, P.H., Wilson, S.J., 2009. Effects on sleep stages and microarchitecture of caffeine and its combination with zolpidem or trazodone in healthy volunteers. J. Psychopharmacol. 23, 487–494. http://dx.doi.org/10.1177/0269881109104852.
Paterson, L.M., Wilson, S.J., Nutt, D.J., Hutson, P.H., Ivarsson, M., 2007. A translational, caffeine-induced model of onset insomnia in rats and healthy volunteers. Psychopharmacol. Berl. 191, 943–950. http://dx.doi.org/10.1007/s00213-006-0672-0.
Paton, C.D., Lowe, T., Irvine, A., 2010. Caffeinated chewing gum increases repeated sprint performance and augments increases in testosterone in competitive cyclists. Eur. J. Appl. Physiol. 110, 1243–1250. http://dx.doi.org/10.1007/s00421-010-1620-6.
Peacock, A., Bruno, R., Martin, F.H., Carr, A., 2013. The impact of alcohol and energy drink consumption on intoxication and risk-taking behavior. Alcohol. Clin. Exp. Res. 37, 1234–1242. http://dx.doi.org/10.1111/acer.12086.
Peck, J.D., Leviton, A., Cowan, L.D., 2010. A review of the epidemiologic evidence concerning the reproductive health effects of caffeine consumption: a 2000-2009 update. Food Chem. Toxicol. 48, 2549–2576. http://dx.doi.org/10.1016/j.fct.2010.06.019.
Peeling, P., Dawson, B., 2007. Influence of caffeine ingestion on perceived mood states, concentration, and arousal levels during a 75-min university lecture. Adv. Physiol. Educ. 31, 332–335. http://dx.doi.org/10.1152/advan.00003.2007.
Pelchovitz, D., Goldberger, J., 2011. Caffeine and cardiac arrhythmias: a review of the evidence. Am. Journ. Med. 124, 284–289.
Penolazzi, B., Natale, V., Leone, L., Russo, P.M., 2012. Individual differences affecting caffeine intake. Analysis of consumption behaviours for different times of day and caffeine sources. Appetite 58, 971–977. http://dx.doi.org/10.1016/j.appet.2012.02.001.
Pham, N.M., Nanri, A., Kurotani, K., Kuwahara, K., Kume, A., Sato, M., Hayabuchi, H., Mizoue, T., 2014. Green tea and coffee consumption is inversely associated with depressive symptoms in a Japanese working population. Public Health Nutr. 17, 625–633. http://dx.doi.org/10.1017/S1368980013000360.
Plichart, M., Menegaux, F., Lacour, B., Hartmann, O., Frappaz, D., Doz, F., Bertozzi, A.-I., Defaschelles, A.-S., Pierre-Kahn, A., Icher, C., Chastagner, P., Plantaz, D., Rialland, X., Hémon, D., Clavel, J., 2008. Parental smoking, maternal alcohol, coffee and tea consumption during pregnancy and childhood malignant central nervous system tumours: the ESCALE study (SFCE). Eur. J. Cancer Prev. 17, 376–383. http://dx.doi.org/10.1097/CEJ.0b013e3282f75e6f.
Pontifex, K.J., Wallman, K.E., Dawson, B.T., Goodman, C., 2010. Effects of caffeine on repeated sprint ability, reactive agility time, sleep and next day performance. J. Sports Med. Phys. Fit. 50, 455–464.
Pourshahidi, L.K., Navarini, L., Petracco, M., Strain, J.J., 2016. A comprehensive overview of the risks and benefits of coffee consumption. Compr. Rev. Food Sci. Food Saf. 15 http://dx.doi.org/10.1111/1541-4337.12206.
Poussel, M., Kimmoun, A., Levy, B., Gambier, N., Dudek, F., Puskarczyk, E., Poussel, J.F., Chenuel, B., 2013. Fatal cardiac arrhythmia following voluntary caffeine overdose in an amateur body-builder athlete. Int. J. Cardiol. 166, e41–e42. http://dx.doi.org/10.1016/j.ijcard.2013.01.238.
Rafferty, K., Heaney, R.P., 2008. Nutrient effects on the calcium economy: emphasizing the potassium controversy. J. Nutr. 138, 166S–171S.
Ragsdale, F.R., Gronli, T.D., Batool, N., Haight, N., Mehaffey, A., McMahon, E.C., Nalli, T.W., Mannello, C.M., Sell, C.J., McCann, P.J., Kastello, G.M., Hooks, T., Wilson, T., 2010. Effect of Red Bull energy drink on cardiovascular and renal function. Amino Acids 38, 1193–1200. http://dx.doi.org/10.1007/s00726-009-0330-z.
Rapuri, P.B., Gallagher, J.C., Kinyamu, H.K., Ryschon, K.L., 2001. Caffeine intake increases the rate of bone loss in elderly women and interacts with vitamin D receptor genotypes. Am. J. Clin. Nutr. 74, 694–700.
Rasche, V., 2003. Cigarette, alcohol, and caffeine consumption: risk factors for spontaneous abortion. Acta Obs. Gynecol. Scand. 82, 182–188.
Rauh, R., Burkert, M., Siepmann, M., Mueck-Weymann, M., 2006. Acute effects of caffeine on heart rate variability in habitual caffeine consumers. Clin. Physiol. Funct. Imaging 26, 163–166. http://dx.doi.org/10.1111/j.1475-097X.2006.00663.x.
Reddy, J.G., Ebbert, J.O., Klesges, L.M., Enders, F.T.B., Klesges, R.C., Lanctot, J.Q., McClanahan, B.S., 2008. The relationship between caffeine and blood pressure in preadolescent African American girls. Ethn. Dis. 18, 283–288.
Ribeiro-Alves, M. a, Trugo, L.C., Donangelo, C.M., 2003. Use of oral contraceptives blunts the calciuric effect of caffeine in young adult women. J. Nutr. 133, 393–398.
Richards, G., Smith, A.P., 2016. A review of energy drinks and mental health, with a focus on stress, anxiety, and depression. J. Caffeine Res. 6, 49–63. http://dx.doi.org/10.1089/jcr.2015.0033.
Richardson, T., Ryder, J., Rozkovec, A., Meckes, C., Thomas, P., Kerr, D., 2004. Influence of caffeine on heart rate variability in patients with long-standing type 1 diabetes. Diabetes Care 27, 1127–1131. http://dx.doi.org/10.2337/diacare.27.5.1127.
Ritchie, K., Ancelin, M.L., Amieva, H., Rouaud, O., Carrière, I., 2014. The association between caffeine and cognitive decline: examining alternative causal hypotheses. Int. Psychogeriatr. 26, 581–590. http://dx.doi.org/10.1017/S1041610213002469.
Rogers, P.J., Heatherley, S.V., Mullings, E.L., Smith, J.E., 2013. Faster but not smarter: effects of caffeine and caffeine withdrawal on alertness and performance. Psychopharmacol. Berl. 226, 229–240. http://dx.doi.org/10.1007/s00213-012-2889-4.
Rogers, P.J., Hohoff, C., Heatherley, S.V., Mullings, E.L., Maxfield, P.J., Evershed, R.P., Deckert, J., Nutt, D.J., 2010. Association of the anxiogenic and alerting effects of caffeine with ADORA2A and ADORA1 polymorphisms and habitual level of caffeine consumption. Neuropsychopharmacology 35, 1973–1983. http://dx.doi.org/10.1038/npp.2010.71.
Rogers, P.J., Smith, J.E., Heatherley, S.V., Pleydell-Pearce, C.W., 2008. Time for tea: mood, blood pressure and cognitive performance effects of caffeine and theanine administered alone and together. Psychopharmacol. Berl. 195, 569–577. http://dx.doi.org/10.1007/s00213-007-0938-1.
Ronen, A., Oron-Gilad, T., Gershon, P., 2014. The combination of short rest and energy drink consumption as fatigue countermeasures during a prolonged drive of professional truck drivers. J. Saf. Res. 49, 39–43. http://dx.doi.org/10.1016/j.jsr.2014.02.006.
Rudolph, T., Knudsen, K., 2010. A case of fatal caffeine poisoning. Acta Anaesthesiol. Scand. 54, 521–523. http://dx.doi.org/10.1111/j.1399-6576.2009.02201.x.
Rueda-Sanchez, M., Diaz-Martinez, L.A., 2008. Prevalence and associated factors for episodic and chronic daily headache in the Colombian population. Cephalalgia 28, 216–225. http://dx.doi.org/10.1111/j.1468-2982.2007.01499.x.
Ruusunen, A., Lehto, S.M., Tolmunen, T., Mursu, J., Kaplan, G. a, Voutilainen, S., 2010. Coffee, tea and caffeine intake and the risk of severe depression in middle-aged Finnish men: the Kuopio Ischaemic Heart Disease Risk Factor Study. Public Health Nutr. 13, 1215–1220. http://dx.doi.org/10.1017/S1368980010000509.
Ruxton, C.H.S., 2014. The suitability of caffeinated drinks for children: a systematic review of randomised controlled trials, observational studies and expert panel guidelines. J. Hum. Nutr. Diet. 27, 342–357. http://dx.doi.org/10.1111/jhn.12172.
Ruxton, C.H.S., 2008. The impact of caffeine on mood, cognitive function, performance and hydration: a review of benefits and risks. Nutr. Bull. 33, 15–25. http://dx.doi.org/10.1111/j.1467-3010.2007.00665.x.
Salinero, J.J., Lara, B., Abian-Vicen, J., Gonzalez-Millán, C., Areces, F., Gallo-Salazar, C., Ruiz-Vicente, D., Del Coso, J., 2014. The use of energy drinks in sport: perceived ergogenicity and side effects in male and female athletes. Br. J. Nutr. 112, 1494–1502. http://dx.doi.org/10.1017/S0007114514002189.
Sanchez-Ortuno, M., Moore, N., Taillard, J., Valtat, C., Leger, D., Bioulac, B., Philip, P., 2005. Sleep duration and caffeine consumption in a French middle-aged working population. Sleep. Med. 6, 247–251. http://dx.doi.org/10.1016/j.sleep.2004.10.005.
Sands, H.R., Downey, L.A., Wilson, R.P., Abbott, L.R., Tysse, B., Parrott, A.C., 2015. Mood and psychomotor tremor changes following acute caffeine consumption in moderate and minimal caffeine consumers. J. Caffeine Res. 5, 42–49. http://dx.doi.org/10.1089/jcr.2014.0020.
Santos, I.S., Matijasevich, A., Domingues, M.R., 2012. Maternal caffeine consumption and infant nighttime waking: prospective cohort study. Pediatrics 129, 860–868. http://dx.doi.org/10.1542/peds.2011-1773.
Sata, F., Yamada, H., Suzuki, K., Saijo, Y., Kato, E.H., Morikawa, M., Minakami, H., Kishi, R., 2005. Caffeine intake, CYP1A2 polymorphism and the risk of recurrent pregnancy loss. Mol. Hum. Reprod. 11, 357–360.
Savitz, D.A., Chan, R.L., Herring, A.H., Howards, P.P., Hartmann, K.E., 2008. Caffeine and miscarriage risk. Epidemiology 19, 55–62. http://dx.doi.org/10.1097/EDE.0b013e31815c09b9.
Savoca, M.R., Evans, C.D., Wilson, M.E., Harshfield, G.A., Ludwig, D.A., 2004. The association of caffeinated beverages with blood pressure in adolescents. Arch. Pediatr. Adolesc. Med. 158, 473–477.
Scher, A.I., Stewart, W.F., Lipton, R.B., 2004. Caffeine as a risk factor for chronic daily headache: a population-based study. Neurology 63, 2022–2027.
Schmid, T.E., Eskenazi, B., Baumgartner, A., Marchetti, F., Young, S., Weldon, R., Anderson, D., Wyrobek, A.J., 2007. The effects of male age on sperm DNA damage in healthy non-smokers. Hum. Reprod. 22, 180–187. http://dx.doi.org/10.1093/humrep/del338.
Schmidt, A., Karlson-Stiber, C., 2008. Caffeine poisoning and lactate rise: an overlooked toxic effect? Acta Anaesthesiol. Scand. 52, 1012–1014. http://dx.doi.org/10.1111/j.1399-6576.2008.01680.x.
Schmidt, R.J., Romitti, P.A., Burns, T.L., Browne, M.L., Druschel, C.M., Olney, R.S., 2009. Maternal caffeine consumption and risk of neural tube defects. Birth Defects Res. Part A Clin. Mol. Teratol. http://dx.doi.org/10.1002/bdra.20624.
Schmidt, R.J., Romitti, P.A., Burns, T.L., Murray, J.C., Browne, M.L., Druschel, C.M., Olney, R.S., 2010. Caffeine, selected metabolic gene variants, and risk for neural tube defects. Birth Defects Res. Part A Clin. Mol. Teratol. 88, 879–889. http://dx.doi.org/10.1002/bdra.20681.
Scholey, A.B., Kennedy, D.O., 2004. Cognitive and physiological effects of an "energy drink": an evaluation of the whole drink and of glucose, caffeine and herbal flavouring fractions. Psychopharmacol. Berl. 176, 320–330. http://dx.doi.org/

10.1007/s00213-004-1935-2.
Scott, W.H., Coyne, K.M., Johnson, M.M., Lausted, C.G., Sahota, M., Johnson, A.T., 2002. Effects of caffeine on performance of low intensity tasks. Percept. Mot. Ski. 94, 521–532. http://dx.doi.org/10.2466/pms.2002.94.2.521.
Sengpiel, V., Elind, E., Bacelis, J., Nilsson, S., Grove, J., Myhre, R., Haugen, M., Meltzer, H.M., Alexander, J., Jacobsson, B., Brantsaeter, A.-L., 2013. Maternal caffeine intake during pregnancy is associated with birth weight but not with gestational length: results from a large prospective observational cohort study. BMC Med. 11, 42. http://dx.doi.org/10.1186/1741-7015-11-42.
Shimbo, M., Nakamura, K., Shi, H.J., Kizuki, M., Seino, K., Inose, T., Takano, T., 2005. Green tea consumption in everyday life and mental health. Public Health Nutr. 8, 1300–1306. http://dx.doi.org/10.1079/phn2005752.
Sigmon, S.C., Griffiths, R.R., 2011. Caffeine choice prospectively predicts positive subjective effects of caffeine and d-amphetamine. Drug Alcohol Depend. 118, 341–348. http://dx.doi.org/10.1016/j.drugalcdep.2011.04.018.
Sigmon, S.C., Herning, R.I., Better, W., Cadet, J.L., Griffiths, R.R., 2009. Caffeine withdrawal, acute effects, tolerance, and absence of net beneficial effects of chronic administration: cerebral blood flow velocity, quantitative EEG, and subjective effects. Psychopharmacol. Berl. 204, 573–585. http://dx.doi.org/10.1007/s00213-009-1489-4.
Signorello, L.B., Nordmark, A., Granath, F., Blot, W.J., McLaughlin, J.K., Annerén, G., Lundgren, S., Ekbom, A., Rane, A., Cnattingius, S., 2001. Caffeine metabolism and the risk of spontaneous abortion of normal karyotype fetuses. Obstet. Gynecol. 98, 1059–1066.
Skewes, M.C., Decou, C.R., Gonzalez, V.M., 2013. Energy drink use, problem drinking and drinking motives in a diverse sample of Alaskan college students. Int. J. Circumpolar Health 1, 72. http://dx.doi.org/10.3402/ijch.v72i0.21204.
Smith, A.P., 2009. Caffeine, cognitive failures and health in a non-working community sample. Hum. Psychopharmacol. Clin. Exp. 24, 29–34. http://dx.doi.org/10.1002/hup.991.
Smith, J.E., Lawrence, A.D., Diukova, A., Wise, R.G., Rogers, P.J., 2012. Storm in a coffee cup: caffeine modifies brain activation to social signals of threat. Soc. Cogn. Affect. Neurosci. 7, 831–840. http://dx.doi.org/10.1093/scan/nsr058.
Sobreiro, B.P., Lucon, A.M., Pasqualotto, F.F., Hallak, J., Athayde, K.S., Arap, S., 2005. Semen analysis in fertile patients undergoing vasectomy: reference values and variations according to age, length of sexual abstinence, seasonality, smoking habits and caffeine intake. São Paulo Med. J. 123, 161–166 doi:/S1516-31802005000400002.
Sofi, F., Conti, A.A., Gori, A.M., Eliana Luisi, M.L., Casini, A., Abbate, R., Gensini, G.F., 2007. Coffee consumption and risk of coronary heart disease: a meta-analysis. Nutr. Metab. Cardiovasc. Dis. 17, 209–223. http://dx.doi.org/10.1016/j.numecd.2006.07.013.
Sondermeijer, H.P., van Marle, A.G.J., Kamen, P., Krum, H., 2002. Acute effects of caffeine on heart rate variability. Am. J. Cardiol. 90, 906–907. http://dx.doi.org/10.1016/S0002-9149(02)02725-X.
Stafford, L.D., Yeomans, M.R., 2005. Caffeine deprivation state modulates coffee consumption but not attentional bias for caffeine-related stimuli. Behav. Pharmacol. 16, 559–571. http://dx.doi.org/10.1097/01.fbp.0000181595.08868.25.
Stein, Z., Susser, M., 1991 May. Miscarriage, caffeine, and the epiphenomena of pregnancy: the causal model. Epidemiology 2 (3), 163–167.
Stefanidou, E.M., Caramellino, L., Patriarca, A., Menato, G., 2011 Oct. Maternal caffeine consumption and sine causa recurrent miscarriage. Eur. J. Obstet. Gynecol. Reprod. Biol. 158 (2), 220–224. http://dx.doi.org/10.1016/j.ejogrb.2011.04.024.
Stephens, M.L., Betts, K., Beck, N.B., Cogliano, V., Dickersin, K., Fitzpatrick, S., Freeman, J., Gray, G., Hartung, T., McPartland, J., Rooney, A.A., Scherer, R.W., Verloo, D., Hoffmann, S., 2016. The emergence of systematic review in toxicology. Toxicol. Sci. 152, 10–16. http://dx.doi.org/10.1093/toxsci/kfw059.
Sulem, P., Gudbjartsson, D.F., Geller, F., Prokopenko, I., Feenstra, B., Aben, K.K.H., Franke, B., den Heijer, M., Kovacs, P., Stumvoll, M., Magi, R., Yanek, L.R., Becker, L.C., Boyd, H.A., Stacey, S.N., Walters, G.B., Jonasdottir, A., Thorleifsson, G., Holm, H., Gudjonsson, S.A., Rafnar, T., Bjornsdottir, G., Becker, D.M., Melbye, M., Kong, A., Tonjes, A., Thorgeirsson, T., Thorsteinsdottir, U., Kiemeney, L.A., Stefansson, K., 2011. Sequence variants at CYP1A1-CYP1A2 and AHR associate with coffee consumption. Hum. Mol. Genet. 20, 2071–2077. http://dx.doi.org/10.1093/hmg/ddr086.
Sünram-Lea, S.I., Owen-Lynch, J., Robinson, S.J., Jones, E., Hu, H., 2012. The effect of energy drinks on cortisol levels, cognition and mood during a fire-fighting exercise. Psychopharmacol. Berl. 219, 83–97. http://dx.doi.org/10.1007/s00213-011-2379-0.
Syce, D.V., 2015. A cup of black tea does not modify heart rate variability. J. Caffeine Res. 5, 55–59. http://dx.doi.org/10.1089/jcr.2014.0016.
Szpak, A., Allen, D., 2012. A case of acute suicidality following excessive caffeine intake. J. Psychopharmacol. 26, 1502–1510. http://dx.doi.org/10.1177/0269881112442788.
Taylor, K.C., Small, C.M., Dominguez, C.E., Murray, L.E., Tang, W., Wilson, M.M., Bouzyk, M., Marcus, M., 2011. Alcohol, smoking, and caffeine in relation to fecundability, with effect modification by NAT2. Ann. Epidemiol. http://dx.doi.org/10.1016/j.annepidem.2011.04.011.
Temple, J.L., Bulkley, A.M., Briatico, L., Dewey, A.M., 2009. Sex differences in reinforcing value of caffeinated beverages in adolescents. Behav. Pharmacol. 20, 731–741. http://dx.doi.org/10.1097/FBP.0b013e328333b27c.
Temple, J.L., Ziegler, A.M., 2011. Gender differences in subjective and physiological responses to caffeine and the role of steroid hormones. J. Caffeine Res. 1, 41–48. http://dx.doi.org/10.1089/jcr.2011.0005.
Temple, J.L., Ziegler, A.M., Graczyk, A., Bendlin, A., Sion, T., Vattana, K., 2014. Cardiovascular responses to caffeine by gender and pubertal stage. Pediatrics 134, e112–e119. http://dx.doi.org/10.1542/peds.2013-3962.
Thelander, G., Jönsson, A.K., Personne, M., Forsberg, G.S., Lundqvist, K.M., Ahlner, J., 2010. Caffeine fatalities–do sales restrictions prevent intentional intoxications? Clin. Toxicol. (Phila) 48, 354–358. http://dx.doi.org/10.3109/15563650903586752.
Thorn, C.F., Aklillu, E., McDonagh, E.M., Klein, T.E., Altman, R.B., 2012. PharmGKB summary: caffeine pathway. Pharmacogenet. Genomics 22, 389–395. http://dx.doi.org/10.1097/FPC.0b013e3283505d5e.
Tinley, E., Yeomans, M., Durlach, P., 2003. Caffeine reinforces flavour preference in caffeine-dependent, but not long-term withdrawn, caffeine consumers. Psychopharmacol. Berl. 166, 416–423. http://dx.doi.org/10.1007/s00213-002-1354-1.
Torelli, P., Evangelista, A., Bini, A., Castellini, P., Lambru, G., Manzoni, G.C., 2009. Fasting headache: a review of the literature and new hypotheses. Headache 49, 744–752. http://dx.doi.org/10.1111/j.1526-4610.2009.01390.x.
Trabulo, D., Marques, S., Pedroso, E., 2011. Caffeinated energy drink intoxication. Emerg. Med. J. 28, 712–714. http://dx.doi.org/10.1136/emj.09.2010.3322rep.
Trapp, G.S.A., Allen, K., O'Sullivan, T.A., Robinson, M., Jacoby, P., Oddy, W.H., 2014. Energy drink consumption is associated with anxiety in Australian young adult males. Depress. Anxiety 31, 420–428. http://dx.doi.org/10.1002/da.22175.
Trovato, G.M., Pirri, C., Martines, G.F., Trovato, F., Catalano, D., 2010. Coffee, nutritional status, and renal artery resistive index. Ren. Fail. 32, 1137–1147. http://dx.doi.org/10.3109/0886022X.2010.516853.
Tsubouchi, H., Shimoya, K., Hayashi, S., Toda, M., Morimoto, K., Murata, Y., 2006. Effect of coffee intake on blood flow and maternal stress during the third trimester of pregnancy. Int. J. Gynecol. Obstet. 92, 19–22.
Turley, K.R., Bland, J.R., Evans, W.J., 2008. Effects of different doses of caffeine on exercise responses in young children. Med. Sci. Sports Exerc 40, 871–878. http://dx.doi.org/10.1249/MSS.0b013e318165984c.
Turley, K.R., Desisso, T., Gerst, J.W., 2007. Effects of caffeine on physiological responses to exercise: boys versus men. Pediatr. Exerc. Sci. 19, 481–492.
Turley, K.R., Gerst, J.W., 2006. Effects of caffeine on physiological responses to exercise in young boys and girls. Med. Sci. Sports Exerc 38, 520–526. http://dx.doi.org/10.1249/01.mss.0000191189.40436.73.
Turnbull, D., Rodricks, J.V., Marioano, G.F., 2016. Neurobehavioral hazard identifiation and characterization for caffeine. Reg. Tox. Pharm. 74, 81–92.
Usman, A., Jawaid, A., 2012. Hypertension in a young boy: an energy drink effect. BMC Res. Notes 5, 591. http://dx.doi.org/10.1186/1756-0500-5-591.
Vik, T., Bakketeig, L.S., Trygg, K.U., Lund-Larsen, K., Jacobsen, G., 2003. High caffeine consumption in the third trimester of pregnancy: gender-specific effects on fetal growth. Paediatr. Perinat. Epidemiol. 17, 324–331. http://dx.doi.org/10.1046/j.1365-3016.2003.00507.x.
Vinader-Caerols, C., Monleón, S., Carrasco, C., Parra, A., 2012. Effects of alcohol, coffee, and tobacco, alone or in combination, on physiological parameters and anxiety in a young population. J. Caffeine Res. 2, 70–76. http://dx.doi.org/10.1089/jcr.2012.0018.
Vink, J.M., Staphorsius, A.S., Boomsma, D.I., 2009. A genetic analysis of coffee consumption in a sample of Dutch twins. Twin Res. Hum. Genet. 12, 127–131. http://dx.doi.org/10.1375/twin.12.2.127.
Vlachopoulos, C., Panagiotakos, D., Ioakeimidis, N., Dima, I., Stefanadis, C., 2005. Chronic coffee consumption has a detrimental effect on aortic stiffness and wave reflections. Am. J. Clin. Nutr. 81, 1307–1312.
Waring, W.S., Goudsmit, J., Marwick, J., Webb, D.J., Maxwell, S.R.J., 2003. Acute caffeine intake influences central more than peripheral blood pressure in young adults. Am. J. Hypertens. 16, 919–924. http://dx.doi.org/10.1016/S0895-7061(03)01014-8.
Watson, J., Deary, I., Kerr, D., 2002. Central and peripheral effects of sustained caffeine use: tolerance is incomplete. Br. J. Clin. Pharmacol. 54, 400–406. http://dx.doi.org/10.1046/j.1365-2125.2002.01681.x.
Wen, W., Shu, X.O., Jacobs, D.R., Brown, J.E., 2001. The associations of maternal caffeine consumption and nausea with spontaneous abortion. Epidemiology 12, 38–42. http://dx.doi.org/10.1097/00001648-200101000-00008.
Weng, X., Odouli, R., Li, D.-K., 2008. Maternal caffeine consumption during pregnancy and the risk of miscarriage: a prospective cohort study. Am. J. Obstet. Gynecol. 198, 279.e1–279.e8. http://dx.doi.org/10.1016/j.ajog.2007.10.803.
Werler, M.M., Louik, C., Mitchell, A.A., 2011. Case-control studies for identifying novel teratogens. Am. J. Med. Genet. C. Semin. Med. Genet. 157C, 201–208. http://dx.doi.org/10.1002/ajmg.c.30307.
Werler, M.M., Pober, B.R., Nelson, K., Holmes, L.B., 1989. Reporting accuracy among mothers of malformed and nonmalformed infants. Am. J. Epidemiol. 129, 415–421.
Wesensten, N.J., 2014. Legitimacy of concerns about caffeine and energy drink consumption. Nutr. Rev. 72, 78–86. http://dx.doi.org/10.1111/nure.12146.
Wetmore, C.M., Ichikawa, L., LaCroix, A.Z., Ott, S.M., Scholes, D., 2008. Association between caffeine intake and bone mass among young women: potential effect modification by depot medroxyprogesterone acetate use. Osteoporos. Int. 19, 519–527. http://dx.doi.org/10.1007/s00198-007-0473-2.
White, J.R.J., Padowski, J.M., Zhong, Y., Chen, G., Luo, S., Lazarus, P., Layton, M.E., McPherson, S., 2016. Pharmacokinetic analysis and comparison of caffeine administered rapidly or slowly in coffee chilled or hot versus chilled energy drink in healthy young adults. Clin. Toxicol. (Phila) 54, 308–312. http://dx.doi.org/10.3109/15563650.2016.1146740.
Wikoff, D.S., Britt, J.K., 2016. The role of systematic review in the practice of

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002

toxicology and risk assessment: an appreciation for the primary tool in evidence-based practice. Toxicol. Open Access 2, 1–4.
Wilson, R.E., Kado, H.S., Samson, R., Miller, A.B., 2012. A case of caffeine-induced coronary artery vasospasm of a 17-year-old male. Cardiovasc. Toxicol. 12, 175–179. http://dx.doi.org/10.1007/s12012-011-9152-9.
Wright, G.A., Baker, D.D., Palmer, M.J., Stabler, D., Mustard, J.A., Power, E.F., Borland, A.M., Stevenson, P.C., 2013. Caffeine in floral nectar enhances a pollinator's memory of reward. Science 339, 1202–1204. http://dx.doi.org/10.1126/science.1228806.
Wright, N., 1980. An assessment of the unreliability of the history given by self-poisoned patients. Clin. Toxicol. 16, 381–384. http://dx.doi.org/10.3109/15563658008989963.
Wu, H.B.A., 2006. Tietz Guide to Clinical Laboratory Tests, fourth ed. Saunders, Elsevier, St. Louis, MO.
Wu, J., Ho, S.C., Zhou, C., Ling, W., Chen, W., Wang, C., Chen, Y., 2009. Coffee consumption and risk of coronary heart diseases: a meta-analysis of 21 prospective cohort studies. Int. J. Cardiol. 137, 216–225. http://dx.doi.org/10.1016/j.ijcard.2008.06.051.
Yamada, K., Kobayashi, M., Kanda, T., 2014. Involvement of adenosine A2A receptors in depression and anxiety. Int. Rev. Neurobiol. 119, 373–393. http://dx.doi.org/10.1016/B978-0-12-801022-8.00015-5.
Yamamoto, T., Yoshizawa, K., Kubo, S.-I., Emoto, Y., Hara, K., Waters, B., Umehara, T., Murase, T., Ikematsu, K., 2015. Autopsy report for a caffeine intoxication case and review of the current literature. J. Toxicol. Pathol. 28, 33–36. http://dx.doi.org/10.1293/tox.28.2014-0044.
Yang, A., Palmer, A.A., De Wit, H., 2010. Genetics of caffeine consumption and responses to caffeine. Psychopharmacol. Berl. 211, 245–257. http://dx.doi.org/10.1007/s00213-010-1900-1.
Yeragani, V.K., Krishnan, S., Engels, H.J., Gretebeck, R., 2005. Effects of caffeine on linear and nonlinear measures of heart rate variability before and after exercise. Depress. Anxiety 21, 130–134. http://dx.doi.org/10.1002/da.20061.
Yoto, A., Motoki, M., Murao, S., Yokogoshi, H., 2012. Effects of L-theanine or caffeine intake on changes in blood pressure under physical and psychological stresses. J. Physiol. Anthropol. 31, 28. http://dx.doi.org/10.1186/1880-6805-31-28.
Young, H.A., Benton, D., 2013. Caffeine can decrease subjective energy depending on the vehicle with which it is consumed and when it is measured. Psychopharmacol. Berl. 228, 243–254. http://dx.doi.org/10.1007/s00213-013-3025-9.
Young, S.L., Hage, M.L., Li, J., 2001. Another case of excessive caffeine and hypokalemia in pregnancy. Obs. Gynecol. 98, 874.
Yukawa, G.S., Mune, M., Otani, H., Tone, Y., Liang, X.M., Iwahashi, H., Sakamoto, W., 2004. Effects of coffee consumption on oxidative susceptibility of low-density lipoproteins and serum lipid levels in humans. Biochem 69, 70–74. http://dx.doi.org/10.1023/B: BIRY.0000016354.05438.0f.
Zhang, B.-Y., Wei, Y.-S., Niu, J.-M., Li, Y., Miao, Z.-L., Wang, Z.-N., 2010. Risk factors for unexplained recurrent spontaneous abortion in a population from southern China. Int. J. Gynaecol. Obstet. 108, 135–138. http://dx.doi.org/10.1016/j.ijgo.2009.09.019.
Zhang, Y., Balilionis, G., Casaru, C., Geary, C., Schumacker, R.E., Neggers, Y.H., Curtner-Smith, M.D., Richardson, M.T., Bishop, P.A., Green, J.M., 2014. Effects of caffeine and menthol on cognition and mood during simulated firefighting in the heat. Appl. Ergon. 45, 510–514. http://dx.doi.org/10.1016/j.apergo.2013.07.005.
Zulli, A., Smith, R.M., Kubatka, P., Novak, J., Uehara, Y., Loftus, H., Qaradakhi, T., Pohanka, M., Kobyliak, N., Zagatina, A., Klimas, J., Hayes, A., La Rocca, G., Soucek, M., Kruzliak, P., 2016. Caffeine and cardiovascular diseases: critical review of current research. Eur. J. Nutr. 55, 1331–1343. http://dx.doi.org/10.1007/s00394-016-1179-z.

Please cite this article in press as: Wikoff, D., et al., Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.002