APPLICATIONS AND ADVANCES OF GRADE IN NUTRITION AND CHILD HEALTH

Applications and advances of the Grading of Recommendations Assessment, Development and Evaluation (GRADE) methodology in nutrition and child health

By Behnam Sadeghirad, Pharm.D, M.P.H.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements for

the Degree Doctor of Philosophy

McMaster University DOCTOR OF PHILOSOPHY (2019)

Hamilton, Ontario (Health Research Methodology)

TITLE: Applications and advances of The Grading of Recommendations Assessment, Development and Evaluation (GRADE) methodology in nutrition and child health

AUTHOR: Behnam Sadeghirad, Pharm.D, M.P.H.

SUPERVISOR: Dr. Gordon H Guyatt, Distinguished Professor

NUMBER OF PAGES: xv, 252

ABSTRACT

The relationship between human health and nutrition is complex and limited widely accepted guidance on proper methods of evidence synthesis is available for nutritional issues. While concepts and methods of evidence synthesis in pharmacological treatments can be mostly applied to nutritional interventions, characteristics unique to the nutrition- and dietetics-related topics can lead to distinct challenges that may not be encountered in evidence synthesis of traditional medical interventions. In addition to traditional methods for pooling the results, state-of-the-art methodologies such as GRADE or network meta-analysis, while being widely used in many medical fields, their use in the field of nutrition and food science is surprisingly rare.

This thesis begins with the assessment of methodological quality of available public health guidelines on sugar intake to determine the extent to which nutritional guidelines follow currently available guidance in evidence synthesis and making practice recommendations. Subsequently, we present two examples of proper implementation of evidence synthesis methods in standard pairwise meta-analysis and indirect treatment comparison and handling of relevant challenges including applications of GRADE approach. Further, this thesis presents a network meta-analysis in the field of nutrition and child health in which the challenges of conducting multiple treatment comparison are tackled and a new approach for presenting and making conclusion from network meta-analysis results is proposed.

ACKNOWLEDGMENTS

Mentors dole out words of wisdom. Coaches roll up their sleeves and get their hands dirty. They don't just believe in our potential; they get in the arena to help us realize our potential. They help us see our blind spots and work through our sore spots. They take responsibility for making us better without taking credit for our accomplishments. There is no feeling greater than the gratitude that I have for my coaches throughout my academic career.

To my supervisor, Gordon Guyatt, and my committee members, Bradley Johnston and Joseph Beyene, I thank you for your outstanding mentorship, guidance, honesty, passion, patience, and understanding, without your support and guidance and the opportunities you have afforded me, I could not have accomplished all that I have thus far in my academic career. For me, this has been invaluable.

To all of the wonderful co-authors that I have had the privilege of collaborating with on all of my projects, I am grateful for your unwavering commitment, arduous efforts and insightful suggestions.

To my peers and colleagues in the Health Research Methodology program, in the Department of Health Research Methods, Evidence, and Impact, I thank you for your friendship, support, sincerity, and laughter. When I started my journey here in Canada, I was far from home, you were always there for me and you always showed up when I needed help, personally or professionally. Without you I'd be lost. Thank you!

Last but not least, thank you to my wife, Shahrzad, who has been extremely supportive of me throughout this entire process and has made countless sacrifices to help me get to this point. I Thank you for your patience and support and understanding me and my many flaws. Without you behind me, I doubt that I would be in this place today.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 3

Chapter 4

Chapter 6

LIST OF TABLES

Chapter 2

Chapter 3

Chapter 4

Chapter 6

LIST OF APPENDICES

Chapter 2

Chapter 3

Chapter 4

Chapter 6

LIST OF ABBREVIATIONS

AGREE: Appraisal of Guidelines for REsearch & Evaluation

APGAR: Appearance, Pulse, Grimace, Activity, and Respiration

BMI: Body Mass Index

BW: Birth Weight

CI: Confidence Interval

CoE: Certainty of Evidence

CoI: Conflict of Interest

DRI: Dietary Reference Intake

D-L: DerSimonian-Laird

Fos: Fructo-oligosaccharides

FM fed: exclusively fed with formula milk

GA: Gestational Age

Gos: galacto-oligosaccharides

GRADE: Grading of Recommendations Assessment, Development, and Evaluation

IF: Incoherence Factor

ILSI: International Life Sciences Institute

IQR: Interquartile Range

IUGR: Intra-Uterine Growth Restriction

I-V: Inverse Variance

LF: Lactoferrin

MD: Mean Difference

M-H: Mantel-Haenszel

MM fed: exclusively fed with mother’s milk

MultiPrb: Multiple-strain probiotics

NEC: Necrotizing Enterocolitis

NICE: The National Institute for Health and Care Excellence

NMA: Network Meta-Analysis

OR: Odds Ratio

PHG: Public Health Guideline

PLC: Placebo

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

RCT: Randomized Controlled Trial

RoB: Risk of Bias

RR: Relative Risk

SACN: Scientific Advisory Committee on Nutrition

SD: Standard Deviation

SGA: Small for Gestational Age

SIGN: Scottish Intercollegiate Guidelines Network

SinglePrb: Single-strain probiotics

SMD: Standardized Mean Difference

SSB: Sugar-Sweetened Beverage

SUCRA: Surface Under the Cumulative Ranking curve

WHO: World Health Organization

DECLARATION OF ACADEMIC ACHIEVEMENT

This thesis is a “sandwich thesis”, which combines five individual projects published or submitted for publication in peer-reviewed journals. This work is original research that I conducted. I am the principle contributor and first author of all the manuscripts contained in this dissertation. The following are detail contributions of B. Sadeghirad in all of the papers included in the dissertation:

Chapter 1: This chapter is unpublished. BS is the sole author.

Chapter 2: This chapter in published in the *Annals of Internal Medicine*. B. Sadeghirad, L. Lytvyn, J. Slavin and B.C. Johnston conceived and designed the study. Analysis and interpretation of the data was done by J. Erickson, B. Sadeghirad, L. Lytvyn, J. Slavin and B.C. Johnston. The manuscript was drafted by J. Erickson, B. Sadeghirad and B.C. Johnston. Critical revision of the article for important intellectual content was done by J. Erickson, B. Sadeghirad, L. Lytvyn, J. Slavin and B.C. Johnston. B. Sadeghirad, L. Lytvyn, J. Erickson and B.C. Johnston collected the data and B Sadeghirad performed all statistical analysis. All authors reviewed or critically revised the manuscript.

Chapter 3: This chapter in published in the journal *Obesity Reviews*. N.R.C. Campbell, B.C. Johnston, and B. Sadeghirad conceived and designed the study. Data collection was done by B. Sadeghirad, T. Duhaney, and S. Motaghipisheh. Data analysis and interpretation was done by B. Sadeghirad and B.C. Johnston. B. Sadeghirad and T. Duhaney drafted the manuscript. All authors reviewed or critically revised the manuscript.

Chapter 4: This chapter is published in the journal *Pediatrics*. B. Sadeghirad and I.D. Florez conceived the study idea. B. Sadeghirad, R.L. Morgan, and I.D. Florez coordinated the systematic review. B. Sadeghirad and I.D. Florez wrote the first draft of the manuscript. B. Sadeghirad and R. Couban designed the search strategy. B. Sadeghirad, R.L. Morgan, and I.D. Florez screened abstracts and full texts. B. Sadeghirad, R.L. Morgan, D. Zeraatkar, and A.M. Zea acquired the data and judged risk of bias in the studies. B. Sadeghirad, R.L. Morgan, B.C. Johnston assess the certainty of evidence. B. Sadeghirad performed the data analysis. B. Sadeghirad, I.D. Florez, B.C. Johnston critically revised the manuscript.

Chapter 5: This chapter in published in International Journal Preventive Medicine. B.C. Johnston, B. Sadeghirad, and G.H. Guyatt conceived and designed the study. B. Sadeghirad and B.C. Johnston drafted the manuscript. All authors reviewed or critically revised the manuscript.

Chapter 6: This chapter is under review in The Lancet. B.C. Johnston, B. Sadeghirad, and G.H. Guyatt conceptualized and designed the study; B.C. Johnston and B. Sadeghirad coordinated and supervised the systematic review. BS, IDF, RLM, FF, YC, DZ, MMB, SS, and TL selected the articles, extracted the data, assessed the risk of bias. B. Sadeghirad performed the data analysis; B. Sadeghirad, I.D. Florez and R.L. Morgan assessed the quality of the evidence and interpret the results. B. Sadeghirad and B.C.

Johnston drafted the initial version of the manuscript. All authors reviewed or critically revised the manuscript.

Chapter 7: This chapter is unpublished. BS is the sole author.

Chapter 1: Introduction of the thesis

Over the past four decades, systematic reviews and meta-analyses have become increasingly popular and the process of performing them has become well-established (1, 2). Well-designed rigorous systematic reviews can summarize available evidence and minimize biases through a comprehensive and reproducible search with *a priori* defined selection criteria, methods for assessing biases, and synthesis methods. Transparency of this process through the publication of an open-access protocol, in a peer-reviewed journal or registration of the protocol in the PROSPERO registry and following guidelines on reporting is also crucial to reproducible science. Such evidence synthesis reports are recognized internationally as credible sources for evidence-informed decision making (1, 3). The increased utility of evidence synthesis methods has contributed to an increase in their publication rates (4-6) and the development of new methodologies such as the GRADE approach for assessment of certainty in evidence, trial sequential analysis, or network meta-analysis (NMA).

There are unique challenges in evidence synthesis of nutrition-related topics (7). The relationship between human health and nutrition is complex and given limited widely accepted guidance on proper methods of addressing nutritional issues, the adaptation of evidence-based methodologies from the traditional biomedical model (e.g. pharmaceutical sciences) to nutrition and dietetics has been slow (8-10). Although concepts and methods of evidence synthesis in pharmacological treatments can be mostly applied to nutritional interventions, characteristics unique to the nutrition- and dietetics-related topics (e.g. dietary exposures that change over the life course depending stage of life, changing dietary preferences and geographic locations, attempting to study single

nutrients that work in unison with other nutrients) can lead to distinct challenges that may not be encountered in evidence synthesis of traditional medical interventions (9, 10).

Variability in evidence synthesis methods and selection of target population comparison, outcomes and consideration of contextual variables related to nutritional interventions and outcomes has resulted in an abundance of reviews on the same subject area with contradictory or counterintuitive results, and poor-quality reporting (11-13). This was first suggested by Moher and Tricco in 2008 (4) and by Chung et al in 2009 (8) and seems to still be the case in 2015 when Salam et al (9) conducted a systematic review of the conduct and methodological quality of selected nutrition interventions. They all pointed to the lack of standardization of outcomes, issues in risk of bias assessment, variation in eligibility criteria and handling heterogeneity which can influence the results of any evidence synthesis study and lead to biased conclusions. Many nutritional and dietetics guidelines have been making their recommendations based on such systematic reviews (4, 8, 9).

One limitation of systematic reviews and meta-analysis is that their quality is largely influenced by the quality and accuracy of underlying evidence. The field of nutrition, like many other areas of health science, relies largely on non-randomized observational evidence and poorly done small randomized trials (11, 14, 15). In addition, the majority of trials only compare treatments or active interventions to standard of care, placebo, or waitlisted controls, which leads to scarcity or more often lack of evidence for comparative effectiveness of interventions.

NMA is a novel method that synthesizes evidence from multiple interventions. It is an extension of the traditional pairwise meta-analysis that allows estimating relative effects of multiple treatments by combining both direct and indirect sources of evidence (16). Ideally, direct comparisons (those compared head-to-head in trials) would be available for all alternatives versus all others; however, all the relevant and desired direct comparisons from randomized trials will seldom be available.

NMA provides a solution to the results dilemma and allows researchers to better understand the relative effectiveness of a particular intervention versus any competing intervention. In addition to providing relative estimates, NMA can rank treatments in order of their effectiveness, providing attractive information for clinicians and policy makers to inform their decisions and has been considered a revolutionary tool for health technology assessments. The methodology of NMA is discussed in detail in multiple other publications (16-18).

The potential of NMA methods is allowing nutrition research to benefit substantially from this methodology. While this methodology is being widely used in many medical fields, its use in the field of nutrition and food science is rare (19, 20). NMA allows investigators to make inferences on every possible pairwise comparison, even when interventions have not been compared in head-to-head trials, which is common in most fields including nutrition. Although a valid indirect inference of effectiveness depends on the extent to which the key assumptions of transitivity and incoherence (also known as inconsistency) are likely to be plausible (16, 19). In addition, NMA allows a better understanding of the

evidence gap and can help in the design of future RCTs, as it has been proposed as a tool to plan the optimal design and the required sample size of new trials (21).

Regardless of statistical methods used to assess the effectiveness and harms of an intervention, it is of utmost importance for decision-makers, to know how much they can rely on the results of an evidence synthesis. Although results of any evidence synthesis provides essential information about the effectiveness and harms associated with an intervention(s), this is not sufficient for making well-informed decisions. Implicit or explicit judgement of decision-makers about the quality of (certainty in) the evidence synthesis conclusion(s) can lead to different implementation decisions.

To avoid errors and make this process transparent, the Grading of Recommendations Assessment, Development and Evaluation (GRADE) working group suggested a systematic and explicit approach for rating the quality (certainty) of evidence (22, 23). This approach is widely accepted worldwide by systematic reviewers, authors of health technology assessments, and guideline developers in healthcare, including over 110 international organizations and societies (e.g. World Health Organization, NICE (National Institute for Health and Care Excellence), and Canadian Agency for Drugs and Technologies in Health and the Cochrane Collaboration).

Although the GRADE approach is being applied in nutrition research, the uptake of the methodology is slower than expected (24). Some have suggested this is due to the inherent methodologic constraints of nutrition and dietetics interventions/exposures and limitations of GRADE methodology that gives a higher certainty to the evidence from

systematic review of randomized trials compared to those of observational studies (24, 25). Even when systematic reviewers and guideline developers use GRADE approach their adherence is poor, or they often apply arguable modifications to the suggested methods with no plausible rationale or supporting evidence (26-28).

Outline of the thesis

This is a sandwich thesis of five papers presented in chapters 2 to 6 covering a range of topics on the application of GRADE approach in nutrition and child health. We intended to better understand the current status of applications of GRADE methodology in the nutrition field, investigate the strengths and limitations of implementation of GRADE approach in traditional meta-analysis methods, as well as indirect treatment comparisons, and network meta-analysis methods. With this work, we provide a platform for improving the use and reporting of GRADE for complex evidence synthesis projects on nutrition topics using traditional meta-analysis and NMA methods.

In chapter 2, we conducted a systematic survey of available public health guidelines on sugar intake recommendations. We aimed to assess the methodological quality of the guidelines and the certainty (quality) of the evidence provided for their recommendations. We used the Appraisal of Guidelines for Research and Evaluation, 2nd edition (AGREE II), instrument to evaluate the methodological quality of published guidelines and assessed evidence quality of articles supporting recommendations based on GRADE methodology.

In chapter 3, we performed a systematic review and meta-analysis of randomized trials on the effects of unhealthy food and beverage marketing on children's dietary intake and dietary preferences to investigate the challenges in applying evidence synthesis methods and GRADE approach in the field of nutrition and food science.

In chapter 4, we conducted a systematic review and indirect comparison to assess the effects of human and bovine colostrum on prevention of morbidity and mortality in preterm infants, where limited evidence of effectiveness was available in small-to-medium size randomized trials with no head-to-head comparison of active interventions. This project provided a unique opportunity to explore the use of indirect comparison and the application of GRADE in nutrition and child health.

Chapter 5 provides the published protocol for a large evidence synthesis project on the relative effectiveness of preventive therapies for necrotizing enterocolitis. This project was an opportunity to perform NMA and explore many evidence synthesis concepts (basics to advanced) and use these concepts when considering the application of GRADE (e.g. inferences from direct vs indirect vs network estimates) and make advances in methods for summarizing findings from NMAs.

In chapter 6, we performed a network meta-analysis of randomized trials on the effects of probiotics, prebiotics, and synbiotics for prevention of mortality and morbidity in preterm infants. We applied the most recent recommendations from GRADE working group for assessment of certainty of evidence in NMA and suggested a new method that provides guidance on how to draw conclusions regarding which treatments are more likely to be

superior in terms of effectiveness, while considering the estimates of effects, certainty of evidence, and rankings. The result of this work is the ascertainment of a more intuitive understanding of NMA results, to be pilot-tested in future work.

Lastly, Chapter 7 summarizes the key findings, limitations and the implications of the thesis with direction for opportunities in future.

References

1. Lavis J, Davies H, Oxman A, Denis JL, Golden-Biddle K, Ferlie E. Towards systematic reviews that inform health care management and policy-making. J Health Serv Res Policy. 2005;10 Suppl 1:35-48.
2. Cook DJ, Greengold NL, Ellrodt AG, Weingarten SR. The relation between systematic reviews and practice guidelines. Ann Intern Med. 1997;127(3):210-6.
3. Oxman AD, Sackett DL, Guyatt GH. Users' guides to the medical literature. I. How to get started. The Evidence-Based Medicine Working Group. Jama. 1993;270(17):2093-5.
4. Moher D, Tricco AC. Issues related to the conduct of systematic reviews: a focus on the nutrition field. Am J Clin Nutr. 2008;88(5):1191-9.
5. Page MJ, Shamseer L, Altman DG, Tetzlaff J, Sampson M, Tricco AC, et al. Epidemiology and Reporting Characteristics of Systematic Reviews of Biomedical Research: A Cross-Sectional Study. PLoS Med. 2016;13(5):e1002028.
6. Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. PLoS Med. 2007;4(3):e78.
7. Balk EM, Horsley TA, Newberry SJ, Lichtenstein AH, Yetley EA, Schachter HM, et al. A collaborative effort to apply the evidence-based review process to the field of nutrition: challenges, benefits, and lessons learned. Am J Clin Nutr. 2007;85(6):1448-56.
8. Chung M, Balk EM, Ip S, Raman G, Yu WW, Trikalinos TA, et al. Reporting of systematic reviews of micronutrients and health: a critical appraisal. Am J Clin Nutr. 2009;89(4):1099-113.
9. Salam RA, Welch V, Bhutta ZA. Systematic reviews on selected nutrition interventions: descriptive assessment of conduct and methodological challenges. BMC Nutrition. 2015;1(1):9.
10. Lichtenstein AH, Yetley EA, Lau J. Application of systematic review methodology to the field of nutrition. J Nutr. 2008;138(12):2297-306.
11. Ioannidis JPA. The Challenge of Reforming Nutritional Epidemiologic Research. Jama. 2018;320(10):969-70.
12. Chartres N, Fabbri A, Bero LA. Association of Industry Sponsorship With Outcomes of Nutrition Studies: A Systematic Review and Meta-analysis. JAMA Intern Med. 2016;176(12):1769-77.
13. Dwyer JT, Rubin KH, Fritsche KL, Psota TL, Liska DJ, Harris WS, et al. Creating the Future of Evidence-Based Nutrition Recommendations: Case Studies from Lipid Research. Adv Nutr. 2016;7(4):747-55.
14. Neale EP, Tapsell LC. Perspective: The Evidence-Based Framework in Nutrition and Dietetics: Implementation, Challenges, and Future Directions. Adv Nutr. 2019;10(1):1-8.

15. Trepanowski JF, Ioannidis JPA. Perspective: Limiting Dependence on Nonrandomized Studies and Improving Randomized Trials in Human Nutrition Research: Why and How. Adv Nutr. 2018;9(4):367-77.

16. Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. Res Synth Methods. 2012;3(2):80-97.

17. Jansen JP, Crawford B, Bergman G, Stam W. Bayesian meta-analysis of multiple treatment comparisons: an introduction to mixed treatment comparisons. Value Health. 2008;11(5):956-64.

18. Mills EJ, Bansback N, Ghement I, Thorlund K, Kelly S, Puhan MA, et al. Multiple treatment comparison meta-analyses: a step forward into complexity. Clin Epidemiol. 2011;3:193-202.

19. Schwingshackl L, Buyken A, Chaimani A. Network meta-analysis reaches nutrition research. Eur J Nutr. 2019;58(1):1-3.

20. Petropoulou M, Nikolakopoulou A, Veroniki AA, Rios P, Vafaei A, Zarin W, et al. Bibliographic study showed improving statistical methodology of network meta-analyses published between 1999 and 2015. J Clin Epidemiol. 2017;82:20-8.

21. Salanti G, Nikolakopoulou A, Sutton AJ, Reichenbach S, Trelle S, Naci H, et al. Planning a future randomized clinical trial based on a network of relevant past trials. Trials. 2018;19(1):365.

22. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. Bmj. 2008;336(7650):924-6.

23. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol. 2011;64(4):383-94.

24. Schwingshackl L, Knuppel S, Schwedhelm C, Hoffmann G, Missbach B, Stelmach-Mardas M, et al. Perspective: NutriGrade: A Scoring System to Assess and Judge the Meta-Evidence of Randomized Controlled Trials and Cohort Studies in Nutrition Research. Adv Nutr. 2016;7(6):994-1004.

25. Meerpohl JJ, Naude CE, Garner P, Mustafa RA, Schunemann HJ. Comment on "Perspective: NutriGrade: A Scoring System to Assess and Judge the Meta-Evidence of Randomized Controlled Trials and Cohort Studies in Nutrition Research". Adv Nutr. 2017;8(5):789-90.

26. Agoritsas T, Merglen A, Heen AF, Kristiansen A, Neumann I, Brito JP, et al. UpToDate adherence to GRADE criteria for strong recommendations: an analytical survey. BMJ Open. 2017;7(11):e018593.

27. Brito JP, Domecq JP, Murad MH, Guyatt GH, Montori VM. The Endocrine Society guidelines: when the confidence cart goes before the evidence horse. J Clin Endocrinol Metab. 2013;98(8):3246-52.

28. Alexander PE, Brito JP, Neumann I, Gionfriddo MR, Bero L, Djulbegovic B, et al. World Health Organization strong recommendations based on low-quality evidence (study quality) are frequent and often inconsistent with GRADE guidance. J Clin Epidemiol. 2016;72:98-106.

Chapter 2: The Scientific Basis of Guideline Recommendations on Sugar Intake: A Systematic Review

Jennifer Erickson*, Behnam Sadeghirad*, Lyubov Lytvyn, Joanne Slavin, Bradley C. Johnston

*Co-first authorship

Cited as and reprinted from: Erickson J, Sadeghirad B, Lytvyn L, Slavin J, Johnston BC. The scientific basis of guideline recommendations on sugar intake. A systematic review. Annals of Internal Medicine, 2017; 166: 257-267. [PMID: 27992898] doi:10.7326/M16-2020.

Author Contributions:

Conception and design: L. Lytvyn, J. Slavin, B.C. Johnston.

Analysis and interpretation of the data: J. Erickson, B. Sadeghirad, L. Lytvyn, J. Slavin, B.C. Johnston.

Drafting of the article: J. Erickson, B. Sadeghirad, L. Lytvyn, B.C. Johnston.

Critical revision of the article for important intellectual content: J. Erickson, B. Sadeghirad, L. Lytvyn, J. Slavin, B.C. Johnston.

Final approval of the article: J. Erickson, B. Sadeghirad, L. Lytvyn, J. Slavin, B.C. Johnston.

Provision of study materials or patients: J. Slavin, B.C. Johnston.

Statistical expertise: B. Sadeghirad.

Collection and assembly of data: J. Erickson, B. Sadeghirad, L. Lytvyn, B.C. Johnston.

Annals of Internal Medicine REVIEW

The Scientific Basis of Guideline Recommendations on Sugar Intake

A Systematic Review

Jennifer Erickson, RD*; Behnam Sadeghirad, PharmD, MPH*; Lyubov Lytvyn, MSc; Joanne Slavin, PhD, RD; and Bradley C. Johnston, PhD

Background: The relationship between sugar and health is affected by energy balance, macronutrient substitutions, and diet and lifestyle patterns. Several authoritative organizations have issued public health guidelines addressing dietary sugars.

Purpose: To systematically review guidelines on sugar intake and assess consistency of recommendations, methodological quality of guidelines, and the quality of evidence supporting each recommendation.

Data Sources: MEDLINE, EMBASE, and Web of Science (1995 to September 2016); guideline registries; and gray literature (bibliographies, Google, and experts).

Study Selection: Guidelines addressing sugar intake that reported their methods of development and were published in English between 1995 and 2016.

Data Extraction: Three reviewers independently assessed guideline quality using the Appraisal of Guidelines for Research and Evaluation, 2nd edition (AGREE II), instrument. To assess evidence quality, articles supporting recommendations were independently reviewed and their quality was determined by using GRADE (Grading of Recommendations Assessment, Development and Evaluation) methods.

Data Synthesis: The search identified 9 guidelines that offered 12 recommendations. Each of the reviewed guidelines indicated a suggested decrease in the consumption of foods containing nonintrinsic sugars. The guidelines scored poorly on AGREE II criteria, specifically in rigor of development, applicability, and editorial independence. Seven recommendations provided nonquantitative guidance; 5 recommended less than 25% to less than 5% of total calories from nonintrinsic sugars. The recommendations were based on various health concerns, including nutrient displacement, dental caries, and weight gain. Quality of evidence supporting recommendations was low to very low.

Limitation: The authors conducted the study independent of the funding source, which is primarily supported by the food and agriculture industry.

Conclusion: Guidelines on dietary sugar do not meet criteria for trustworthy recommendations and are based on low-quality evidence. Public health officials (when promulgating these recommendations) and their public audience (when considering dietary behavior) should be aware of these limitations.

Primary Funding Source: Technical Committee on Dietary Carbohydrates of the North American branch of the International Life Sciences Institute. (PROSPERO: CRD42015029182)

Ann Intern Med. 2017;166:257-267. doi:10.7326/M16-2020 **Annals.org**
For author affiliations, see end of text.
This article was published at Annals.org on 20 December 2016.
* Ms. Erickson and Dr. Sadeghirad contributed equally to this work.

The relationship between sugar and health is complex due to multiple interrelated variables, including state of energy balance, macronutrient substitutions, and underlying diet and lifestyle patterns (1). Existing evidence of a link between sugar intake and adverse health outcomes has been translated into dietary guidance and recommendations for the general public by authoritative health organizations (2). Dietary guidance addresses the types of sugars, especially sources of nonintrinsic sugars, such as added sugars and free sugars (2). Added sugars consist of monosaccharides and disaccharides added during the production and preparation of foods and beverages and do not include sugars naturally found in milk, fruit, and fruit juice. Free sugars comprise sugars added to products as well as sugars naturally found in fruit, honey, and syrup (3).

As research continues to add knowledge, authoritative organizations have issued public health guidance based on the available evidence (2). Recent guidelines have included both qualitative and quantitative recommendations that consistently focus on limiting and reducing sugar consumption, especially sources of nonintrinsic sugars (2). For example, in 2015, the World Health Organization (WHO), the Scientific Advisory Committee on Nutrition (SACN), and the U.S. Department of Agriculture and U.S. Department of Health and Human Services issued public health guidelines (PHGs) with specific recommendations for dietary sugar intake (4–6). Each organization conducted its own review of the available evidence and published its recommendations, including the scientific basis for its conclusions. These organizations have crafted different recommendations with regard to sugar consumption, with various rationales for limiting intake.

When respected organizations issue conflicting recommendations, it can result in confusion and raises concern about the quality of the guidelines and the underlying evidence. We conducted a systematic survey and critical appraisal of authoritative PHGs, including an assessment of the quality of evidence supporting recommendations for dietary sugar intake.

See also:

Editorial comment 305

Web-Only
CME quiz

METHODS

We registered the protocol for this systematic review in the PROSPERO database in November 2015 (registration number CRD42015029182) (7).

Data Sources and Searches

Using a search strategy developed with the help of an experienced librarian, we searched MEDLINE, EMBASE, and Web of Science (1995 to September 2016) using subject terms and keywords. We searched 5 gray literature sources, including Google (Appendix Table 1, available at Annals.org), as well as bibliographies of included studies. We consulted with 3 experts in the field of carbohydrates (Appendix Table 1) to identify additional guidelines we may have missed. Our search was restricted to English-language guidelines.

Study Selection

Our criteria for inclusion were 1) PHGs, defined as documents developed by a nationally recognized committee, a publicly funded institution, or a medical society that provided recommendations for sugar intake in the general population; 2) inclusion of an explicit methodology section, either within the guideline or in supporting documents (for example, definition of the search strategy, evidence quality assessment, and methods used to create recommendations); 3) the most recent version of publications from an organization; and 4) publication between 1995 and 2016.

Our target outcomes of interest were the overall quality of development of the PHGs; the consistency of sugar recommendations, both quantitative and qualitative; the strength of the recommendations; an assessment of the supporting evidence for each recommendation; the use of systematic review methods; explicit links between recommendations and supporting evidence; and the strengths and limitations of the body of evidence.

Data Extraction and Quality Assessment

Two reviewers (B.S. and J.E.) independently screened titles and abstracts, full-text articles, and data extracted from included PHGs by using standardized, pilot-tested forms. We abstracted the following guideline characteristics: title, year, authors, language, organization, whether it was a novel publication or an update, location of development, the recommendations for sugar intake along with the strength of each recommendation, and the authors' assessment of the quality of the supporting evidence. Pairs of reviewers (B.S., J.E., L.L., and B.C.J.) independently identified, extracted, and appraised references to the evidence used to justify each recommendation, including the types of sugars (for example, added, free, or total) referenced in the supporting body of literature. Reviewers resolved disagreements by consensus and, if consensus could not be reached, consulted with senior scientists (B.C.J. and J.S.).

Three reviewers (B.S., J.E., and L.L.) independently appraised guidelines by using the Appraisal of Guidelines for Research and Evaluation, 2nd edition (AGREE II), instrument, comprising 23 items within 6 domains: scope and purpose, stakeholder involvement, rigor of development, clarity of presentation, applicability, and editorial independence (Appendix Table 2, available at Annals.org) (8). In addition, 2 overall assessments were completed for each PHG: a score of 1 to 7, and whether the reviewer would recommend using the guideline (recommended, recommended with modifications, or not recommended). We conducted a calibration exercise using 2 guidelines to ensure consistency and validity and resolved disagreements by consensus. Item rating differences of 3 points or fewer between reviewers were permitted. Senior scientists (B.C.J. and J.S.) were available for discrepancies but were not needed.

Quality Appraisal of Evidence Used in Guidelines

We used the GRADE (Grading of Recommendations Assessment, Development and Evaluation) approach (9) to independently assess the quality of the evidence underlying each recommendation. For each target outcome linked to a recommendation, GRADE assigns the quality of evidence as high, moderate, low, or very low. Systematic reviews of randomized, controlled trials (RCTs) started with high quality of evidence, whereas systematic reviews of observational studies started with low quality. In instances where only single studies for recommendations were cited, RCTs started with moderate-quality evidence and observational studies started with very-low-quality evidence. For each body of evidence (systematic reviews) and for each citation (single studies), where possible, we considered downgrading the quality of evidence on the basis of 5 domains: risk of bias, indirectness, imprecision, inconsistency, and publication bias. Subsequently, we considered rating up on the basis of 3 domains: large effect size, dose–response, and an absence of residual or unmeasured confounding.

Data Synthesis and Analysis

Agreement for the full-text screening was calculated using the κ statistic and its 95% CI (10). For each guideline, we calculated the AGREE II score for each domain as a percentage of the maximum possible score and standardized range. We considered 60% as a threshold of acceptable quality. Interrater agreement was calculated using the intraclass correlation coefficient with corresponding 95% CIs (11). Agreement of 0.01 to 0.20 was considered poor, 0.21 to 0.40 was considered fair, 0.41 to 0.60 was considered moderate, 0.61 to 0.80 was considered substantial, and 0.81 to 1.00 was considered very good (12). For all AGREE II domains across all PHGs, we calculated the median domain score and the interquartile range (IQR). All analyses were conducted using Excel 2013 (Microsoft).

Role of the Funding Source

This study was supported by the Technical Committee on Dietary Carbohydrates of the North American branch of the International Life Sciences Institute (ILSI North America). ILSI North America is a public, nonprofit foundation that provides a forum to advance understanding of scientific issues related to the nutritional quality and safety of the food supply by sponsoring re-

search programs, educational seminars and workshops, and publications. ILSI North America receives 60% of its financial support from its more than 400 industry members. The authors wrote the protocol, which was reviewed for scope clarifications and approved by ILSI. The funding source had no role in the conduct of the review or the interpretation of data, manuscript review, or publication decisions.

RESULTS

A total of 5315 records were screened, 26 records were considered potentially eligible for full-text screening, and 9 PHGs proved eligible (Figure). Eligible guidelines included 1 global guideline (4), 2 international guidelines (13, 14), and 6 national guidelines (5, 6, 15–18). Guidelines were published from 2002 to 2015 by the following agencies: the U.S. Department of Agriculture and the U.S. Department of Health and Human Services (6), WHO (4), SACN and Public Health England (5), the Ministry of Health of Brazil (15), the Australian National Health and Medical Research Council (18), the Nordic Council of Ministers (14), the German Nutrition Society (16), the Food Safety Authority of Ireland (17), and the Institute of Medicine (13) (Table 1).

Recommendation Characteristics

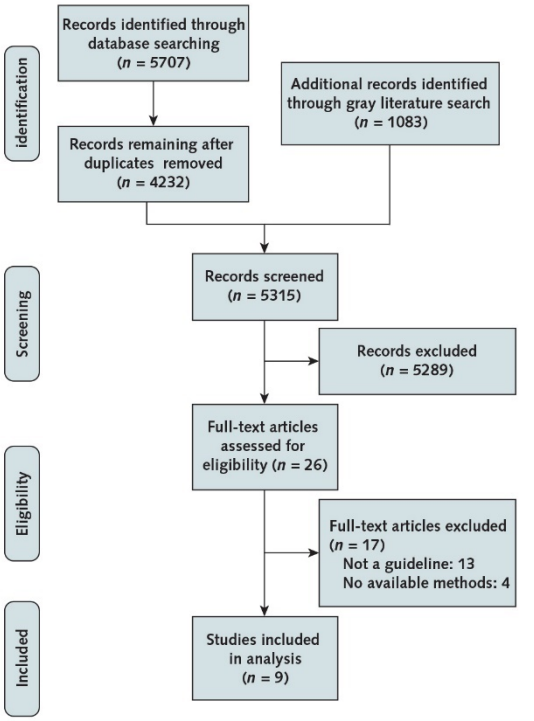
The 9 PHGs provided a total of 12 recommendations on dietary sugar intake. All recommendations advocated for reduced intake of nonintrinsic free or added sugars and/or decreased consumption of foods and beverages high in refined sugars, and 5 recommendations provided specific sugar intake limits (Table 1). Guidelines used variable terminology in sugar recommendations. For example, 2 guidelines used the term "free sugars" (4, 5), 3 used the term "added sugars" (6, 13, 14), 2 made recommendations on sugar-sweetened beverages (SSBs) (5, 16), and 3 referred to food and beverage sources of refined sugars (15, 17, 18). Quantitative recommendations ranged from less than 5% of total energy from free sugars (4, 5) to less than 25% of total energy from added sugars (13). The rationale for decreased sugar intake included nutrient displacement, excess energy intake, dental caries, bone health, weight gain, and obesity. Four guidelines assessed the quality of the evidence and used the assessment to develop their recommendations (4, 5, 16, 18), and 5 did not (6, 13, 15, 17, 19).

Quality Assessment of Guidelines: AGREE II Results

Scope and Purpose

Items in this domain evaluate the overall objectives, related health questions, and the target population of the guideline (20). Across guidelines, the median score for this domain was 81.5% (IQR, 72.2% to 88.0%), indicating that most items were highly rated (Table 2). Eight of the 9 guidelines reached the 60% threshold for reporting. The main limitation across all guidelines was the description of expected benefit, or outcomes, of the guidelines.

Figure. Summary of evidence search and selection.



Stakeholder Involvement

Stakeholder involvement criteria focus on the extent of involvement of appropriate participants in the guideline development process and whether it reflects the views of its intended users (20). The median score for this domain was 63.0% (IQR, 38.9% to 77.8%) (Table 2). Four guidelines scored below 60% in this domain (5, 13, 16, 17). Many guidelines did not describe how they sought the views and preferences of their target population (patients or the public), and those that did were vague about the process.

Rigor of Development

Rigor of development relates to the methods used for gathering and synthesizing the evidence for guideline development, formulation of the recommendations, and the process for updating the guideline (20). The median score for this domain was low, at 47.2% (IQR, 24.0% to 69.4%) (Table 2). Three of the guidelines met the 60% threshold (4, 6, 18). Four guidelines did not use systematic methods to search for evidence (6, 13, 15, 17). Four guidelines assigned strength to their recommendations (4, 6, 16, 18), but only the WHO guideline used the GRADE approach (4). Three of the

Table 2. 1: Identified guidelines and corresponding sugar recommendations

Table 1. Identified Guidelines and Corresponding Sugar Recommendations

Guideline, Year (Reference)	Guideline Title	Funding	Qualitative Recommendation	Quantitative Recommendation*
U.S. Department of Agriculture, U.S. Department of Health and Human Services, 2015 (6)	2015–2020 Dietary Guidelines for Americans	Unclear	–	"Consume less than 10% of calories per day from added sugars"
WHO, 2015 (4)	Sugars Intake for Adults and Children	Ministry of Health, Labour and Welfare of the Government of Japan; Korean Food and Drug Administration; Zhejiang University; and the WHO Regional Office for Europe	"Reduced intake of free sugars throughout the life course"	"In both adults and children, WHO recommends reducing the intake of free sugars to less than 10% of total energy intake"; "WHO suggests further reduction of the intake of free sugars to below 5% of total energy intake"
Public Health England/SACN, 2015 (5)	Carbohydrates and Health	Unclear	"The consumption of sugars-sweetened beverages should be minimised in both children and adults"	"The population average intake of free sugars should not exceed 5% of total dietary energy for age groups from 2 years upwards"
Ministry of Health of Brazil, Secretariat of Health Care, Primary Health Care Department, 2014 (15)	Dietary Guidelines for the Brazilian Population	Unclear	"Use oils, fats, salt, and sugar in small amounts for seasoning and cooking foods and to create culinary preparations"	–
National Health and Medical Research Council, 2013 (18)	Australian Dietary Guidelines	Unclear	"Limit intake of foods and drinks containing added sugars such as confectionary, sugar-sweetened soft drinks and cordials, fruit drinks, vitamin waters, energy and sports drinks"	–
Nordic Council of Ministers, 2012 (14)	Nordic Nutrition Recommendations	Nordic Council of Ministers	–	"Intake of added sugars should be kept below 10% of the energy intake"
German Nutrition Society, 2012 (16)	Evidence-based Guideline of the German Nutrition Society	Unclear	"The consumption of sugar-sweetened beverages should be limited"	–
Food Safety Authority of Ireland, 2011 (17)	Scientific Recommendations for Healthy Eating Guidelines in Ireland	Department of Health and Children	"Healthy eating can be enjoyed with limited amounts of 'other foods' like biscuits, cakes, savoury snacks and confectionery. These foods are rich in calories, fat, sugar and salt so remember—NOT too MUCH and NOT too OFTEN"	–
Institute of Medicine, Food and Nutrition Board, 2002 (13)	Dietary Reference Intakes for Energy, Carbohydrate, Fiber, Fat, Fatty Acids, Cholesterol, Protein and Amino Acids	U.S. Department of Health and Human Services Office of Disease Prevention and Health Promotion, Health Canada, U.S. Food and Drug Administration, National Institutes of Health, Centers for Disease Control and Prevention, U.S. Department of Agriculture, U.S. Department of Defense, Institute of Medicine, and Dietary Reference Intakes Private Foundation Fund and Corporate Donors Fund, including the Dannon Institute, International Life Sciences Institute, Roche Vitamins Inc., Mead Johnson Nutrition Group, and M&M Mars	–	"A maximal intake level of 25% or less of energy is suggested to prevent the displacement of foods that are major sources of essential micronutrients"

Table 1–Continued

Basis for Recommendation	Methods Used to Determine Recommendations	Sugar Definition in Public Health Guidelines	Types of Sugar in Relevant Evidence
Nutrient displacement	Systematic review, diet modeling, and national intake data	Added sugars include syrups, brown sugar, corn sweetener, corn syrup, dextrose, fructose, glucose, high-fructose corn syrup, honey, invert sugar, lactose, malt syrup, maltose, molasses, raw sugar, sucrose, trehalose, and turbinado sugar	Not applicable; diet modeling used for evidence
Dental caries and weight gain	Systematic review	Free sugars include monosaccharides and disaccharides added to foods and beverages by the manufacturer, cook, or consumer and sugars naturally present in honey, syrups, fruit juices, and fruit juice concentrates	Sucrose, added sugar, total sugars, free sugars, SSBs, fructose, and sweet foods
Excess energy intake	Systematic review	All monosaccharides and disaccharides added to foods by the manufacturer, cook, or consumer plus sugars naturally present in honey, syrups, and unsweetened fruit juices	Total sugars, individual sugars, SSBs, sweet food, fruit juice, and nonmilk extrinsic sugars
Excess energy intake	Consensus	No definition	Not applicable; did not assess the literature
Weight gain, dental caries, and bone health	Systematic review and diet modeling	Not applicable	SSBs, energy-dense snack foods, fruit juice, sucrose, and total sugar
Dental caries, obesity, and nutrient displacement	Systematic review	Added sugars include sucrose, fructose, glucose, starch hydrolysates (glucose syrup and high-fructose syrup), and other isolated sugar preparations used as such or added during food preparation and manufacturing	SSBs, dietary sugars, fructose, sucrose, sweet foods, added sugar, and fruit juice
Excess energy intake	Systematic review	Not applicable	Sweets, SSBs, fructose, and glucose
Dental caries	Diet modeling	Not applicable	Total sugar
Nutrient displacement	Literature review	Not applicable	Total sugar, added sugar, and nonmilk extrinsic sugars

SACN = Scientific Advisory Committee on Nutrition; SSB = sugar-sweetened beverage; UK = United Kingdom; WHO = World Health Organization.
* Although scientific reports were commissioned, including systematic reviews, quantitative recommendations were based on modeling and intake data.

Table 2. 2: Public health guideline domain scores on the AGREE II instrument

Table 2. Public Health Guideline Domain Scores on the AGREE II Instrument

Guideline (Reference)	Intraclass Correlation Coefficient*	Score, %						Combined Overall Rating	Systematic Method†
		Scope and Purpose	Stakeholder Involvement	Rigor of Development	Clarity of Presentation	Applicability	Editorial Independence		
Carbohydrates and Health (5)	0.966	81.5	37.0	47.2	48.1	0	0	3.7	Yes
Sugars Intake for Adults and Children (4)	0.887	88.9	77.8	81.3‡	59.3	36.1	83.3‡	4.3	Yes
Nordic Nutrition Recommendations (14)	0.913	83.3	63.0	50.0	53.7	15.3	33.3	4.7	Yes
Dietary Guidelines for the Brazilian Population (15)	0.873	53.7	74.1	16.7	50.0	34.7	33.3	3.7	No
Evidence-based Guideline of the German Nutrition Society (16)	0.941	74.1	18.5	41.0	38.9	6.9	13.9	3.3	Yes
Scientific Recommendations for Healthy Eating Guidelines in Ireland (17)	0.964	70.4	40.7	10.4	72.2	58.3	0	4.0	No
Australian Dietary Guidelines (18)	0.870	92.6‡	77.8	69.4	66.7	61.1‡	77.8	5.3‡	Yes
Dietary Reference Intakes (13)	0.935	75.9	46.3	31.3	70.4	18.1	52.8	3.7	No
2015-2020 Dietary Guidelines for Americans (6)	0.873	87.0	87.0‡	69.4	79.6‡	41.7	30.6	5.0	No

AGREE II = Appraisal of Guidelines for Research and Evaluation, 2nd edition.
* Agreement among reviewers for inclusion of guideline.
† Denotes whether systematic review methods (for example, systematic search and selection of criteria and quality assessment of studies) were used in the development of the guideline.
‡ Highest-rated guideline in this domain.

guidelines discussed external review by experts before publication (4, 6, 18). Two guidelines appropriately described the process for updating recommendations (4, 6).

Clarity of Presentation

Clarity of presentation relates to whether key recommendations are unambiguous and easily identifiable in the guideline (20). The median score for this domain was 59.3% (IQR, 49.1% to 71.3%), with 4 guidelines meeting the 60% threshold (6, 13, 17, 18) (Table 2). The main limitation in this domain was that the different options for management of the health issue (for example, ways to limit sugar intake) were not clearly presented.

Applicability

Items in the applicability domain focus on the likely barriers to and facilitators of implementation, strategies to improve uptake, and resource implications of applying the guideline (20). The median score for this domain was low, at 34.7% (IQR, 11.1% to 50.0%) (Table 2). Only 1 guideline met the 60% threshold (18). The most common issue was failing to discuss the facilitators and barriers to the guideline's application and failing to address the resource implications of applying the recommendations. Only 1 guideline (4) presented monitoring and auditing criteria.

Editorial Independence

Editorial independence relates to unbiased formulation of recommendations and competing interests (20). This domain had the lowest median score (33.3% [IQR, 6.9% to 65.3%]), with only 2 guidelines meeting the 60% threshold (Table 2). Most of the guidelines either did not provide a statement about funding and its influence in the process of guideline development or failed to state conflicts of interest of authors or the guideline panel (Appendix Table 3, available at Annals .org).

Overall Assessment

Overall guideline quality was moderate (median score, 4.0 [IQR, 3.7 to 4.8]), with only the Australian guideline meeting the 60% threshold for all 6 domains. Scores ranged from 3.3 (German guideline [16]) to 5.3 (Australian guideline [18]) (Table 2). All of the guidelines were categorized as "recommended with modifications."

Quality Assessment of Supporting Evidence for Recommendations: GRADE Results

There were a total of 66 unique publications across 9 eligible guidelines supporting the 12 dietary sugar recommendations. Evidence included systematic reviews; RCTs; nonrandomized, controlled trials; prospective cohort studies; case–control studies; national surveys; and cross-sectional studies (**Appendix Table 4**, available at Annals.org). The Dietary Guidelines for the Brazilian Population and the 2015–2020 Dietary Guidelines for Americans did not cite any previously published studies as evidence for their recommendations (6, 15), and Public Health England conducted its own systematic reviews for its Carbohydrates and Health report that have not been published in a peer-reviewed journal but were publicly available (5).

Sixteen systematic reviews were used to inform 7 recommendations across 5 guidelines (4, 5, 14, 16, 18) (**Appendix Table 5**, available at Annals.org). Evidence was low to very low for each systematic review. Fourteen reviews (87.5%) were downgraded for inconsistency, 11 (68.8%) were downgraded for imprecision, 2 (14%) were downgraded for publication bias, and 2 (12.5%) were downgraded for indirectness.

Two large RCTs (21, 22), both on SSBs and body weight, informed 2 recommendations from the German and Australian guidelines (16, 18) (**Appendix Table 5**). Our independent review indicated that the evidence was of very low quality for both and was downgraded for imprecision (wide CIs and trivial treatment effects based on the lower bound of the 95% CI) and indirectness. Eight small RCTs (<300 events for dichotomous outcomes or <400 participants for continuous outcomes) started at moderate quality and were all downgraded to very low quality due to imprecision and indirectness.

Eight large cohort studies (**Appendix Table 5**), all on SSBs and health outcomes (such as type 2 diabetes and body weight), informed 3 recommendations across the Nordic, German, and Australian guidelines (14, 16, 18). Evidence was considered very low quality for 6 studies (75%) (23–28) and low quality for 2 studies (25%) (29, 30). Three studies were downgraded for indirectness (37.5%), and 2 were downgraded for imprecision (25%). Two studies were rated up for a dose-response (25%) (29, 30). Twenty-eight small cohort studies started at very low quality, and we did not rate up given their imprecision and indirectness.

Although a Dietary Guidelines Advisory Committee drafted an extensive scientific report (31) to inform the 2015–2020 Dietary Guidelines for Americans (6), the guidelines cited food pattern modeling and U.S. national caloric intake data from added sugars to inform recommendations. We planned to use GRADE to evaluate the quality of the evidence used in the model components as well as the accuracy of the modeling procedure; however, these details were not publicly available, and we were unable to assess the quality of the evidence for the recommendations.

The WHO guideline was the only one to use the GRADE approach (9). The WHO conducted 2 systematic reviews, one of which included observational studies evaluating effects of free sugars on dental caries (assessed as moderate-quality by the WHO and graded up for large effect size) and the other including RCTs and observational studies evaluating effects of free sugars on body weight (assessed as moderate-quality by the WHO and downgraded for publication bias). Although the WHO guideline recommendations are for free sugars, included studies among both systematic reviews used various forms of sugar, including sucrose, added sugars, and total sugars for the dental caries review (32) and free sugars, SSBs, fructose, sucrose, sweet foods, and added sugars for the body weight review (33). Similar discrepancies were found in 5 additional guidelines (**Table 1**).

We independently reviewed the WHO evidence profiles and deemed the quality of evidence on sugars and body weight to be low (with additional downgrading for inconsistency). We also reasoned that the evidence on sugar and dental caries was low (unlike WHO's rationale, we did not rate up for a large effect size). The WHO issued a strong recommendation to reduce free sugars to less than 10% of daily caloric intake based on 5 cohort studies (1200 children) assessing the risk for dental caries and a weak recommendation to reduce free sugars to less than 5% of daily caloric intake based on 3 ecological studies on the risk for dental caries.

DISCUSSION

We identified 9 PHGs containing 12 dietary sugar recommendations. The quality of development of the guidelines (assessed using the AGREE II instrument) was moderate, with 3 of 6 AGREE II domains (rigor of development, applicability, and editorial independence) having major limitations. Seven recommendations were qualitative, whereas 5 were quantitative, ranging from less than 5% to less than 25% of total calories from nonintrinsic sugars per day. The rationale for the varied sugar intake recommendations was based primarily on nutrient displacement, dental caries, and weight gain.

Using the GRADE approach, we found that the overall quality of evidence to support recommendations was low to very low. Optimal guidelines should be developed with increased rigor, and recommendations should be specific (population, exposure, comparator group, and outcomes critically important to the general public) and transparent (including explicit conflicts of interest and how the body of evidence was considered for developing each recommendation) and should follow GRADE guidance as intended (weak recommendations if the quality of evidence is low, with few exceptions [34]).

A PubMed search for reviews of dietary sugar guidelines done within the past 5 years identified only 1 other review. Although Hess and colleagues (2) reviewed dietary sugar recommendations around the world, the search was not systematic and the review did not assess the quality of the guidelines or the support-

ing evidence. The authors concluded that no clear link exists between added sugar intake and health outcomes.

The included guidelines examined the potential health effects of sugars and risk for dental caries, obesity, type 2 diabetes, and cardiovascular disease. The WHO and SACN suggested that a strong correlation exists between overall free sugars and health outcomes (4, 5). In both guidelines, most of the cited evidence examined SSB consumption and health outcomes rather than the consumption of free sugars from various foods.

Our review had limitations. This project was funded by ILSI, an organization that is funded primarily by the food and agriculture industry. The authors, having expertise in study methodology (particularly in the development of practice guidelines), wrote the protocol and conducted the study independent of the funding body. However, given our funding source, our study team has a financial conflict of interest and readers should consider our results carefully.

We initially sought to assess the quality of the evidence underlying the recommendations by using the Oxford Levels of Evidence, as indicated in our publicly available protocol. Post hoc, we chose to use the GRADE approach, wherein a body of evidence is categorized using intuitive language (high, moderate, low, or very low quality) and each category is accompanied by an explicit definition. In contrast, the Oxford Levels of Evidence uses numbers associated with specific study designs based on the traditional hierarchy of evidence. We believe that the Oxford Levels of Evidence gives a false impression of the evidence (for example, a systematic review of RCTs rated as level 1 evidence despite potentially serious limitations when comprehensively assessed using the GRADE approach). With GRADE methods, the evidence can be rated up or down on the basis of a set of criteria (such as precision, risk of bias, and publication bias). The criteria are applied using a systematic and explicit approach that includes extensive instructions and transparency with respect to the quality assessment. We believe that the use of GRADE reduces the likelihood of mislabeling the overall certainty of evidence.

Only 9 guidelines that explicitly reported their methods were included in this review. Given our focused eligibility criteria, this was not a review of all available dietary sugar recommendations that may influence the beliefs and actions of the public, regulators, and health care practitioners. For example, we identified 4 publications (35–38) containing dietary sugar recommendations written by influential organizations (American Academy of Pediatrics, European Food Safety Authority, American Heart Association, and India National Institute of Nutrition) that were excluded because they lacked a written methodology section. We did not include these reports because a comprehensive understanding of the methods used to develop a PHG is essential to assessing the quality of the development of a guideline and the quality of evidence for recommendations. We also excluded PHGs that were not published in English. Although our review included guidelines from around the world, it was not a comprehensive review of all potentially available guidelines.

Our review also had several strengths. A priori, we documented our eligibility criteria, objectives, and planned methods of analysis as publicly registered on PROSPERO (7). We independently assessed the quality of development of dietary guidelines by using AGREE II and the certainty of evidence for sugar recommendations by using the GRADE framework, which has been endorsed by more than 90 health organizations worldwide (39). On the basis of our methodological analysis of PHGs, we believe the range of various recommendations and the evidence that supports these recommendations can be better interpreted by health care professionals and consumers trying to design effective programs and provide guidance to the public about sugar intake.

All of the reviewed guidelines suggested a decrease in consumption of nonintrinsic sugars. Although the overall direction was consistent, the rationale and evidence used to make each recommendation were inconsistent. This lack of evidentiary consistency, with various health concerns cited, creates confusion for practitioners and the public about the role that sugar plays in health.

Quantitative limits on sugar intake were recommended in 5 of the 9 PHGs (4–6, 13, 14). Each of the quantitative sugar recommendations (except the WHO recommendation) was based on an estimate of how much sugar could be consumed while maintaining a "healthy diet." For example, the Dietary Reference Intakes and the 2015–2020 Dietary Guidelines for Americans set limits of less than 25% and less than 10% of energy from added sugars, respectively (6, 13), based on diet modeling and intake data. Similarly, the SACN recommendation was based on the desired energy reduction of 100 calories per day for effective population-wide weight loss. An approximated 100 calories of free sugars was subtracted from the previous sugar recommendation to obtain this 100-calorie deficit, resulting in the specified maximal intake of 5% of total energy from free sugars (5). The method by which the Nordic Council of Ministers determined a limit of 10% of energy from added sugars was not explained in its PHG (14). In contrast, the WHO used 5 cohort studies (moderate quality) and 3 ecological studies (very low quality) on the risk for dental caries to set the limit of intake of free sugars to below 10% and 5% of total energy intake (4).

The quality of available evidence to link sugar with health outcomes was generally rated as low to very low. The prevailing concerns with high sugar intake are directed toward excessive calorie consumption and nutrient displacement. Sugar added to products adds considerable calories without any nutritional benefits and may take the place of other nutrient-dense foods in the diet. From a practical standpoint, added sugars are a source of calories that many public health authorities believe can be easily reduced. Doing so at a population level may result in a reduction in caloric intake and a subsequent decrease in the rate of overweight and

obesity. At present, there seems to be no reliable evidence indicating that any of the recommended daily caloric thresholds for sugar intake are strongly associated with negative health effects. The results from this review should be used to promote improvement in the development of trustworthy guidelines on sugar intake (40).

From University of Minnesota, St. Paul, Minnesota; McMaster University, Hamilton, Ontario, Canada; Kerman University of Medical Sciences, Kerman, Iran; and The Hospital for Sick Children Research Institute and University of Toronto, Toronto, Ontario, Canada.

Note: As the guarantors of the study, Drs. Johnston and Slavin take full responsibility for the work as a whole, including the study design, access to data, and the decision to submit and publish the manuscript.

Acknowledgment: The authors thank Tamsin Adams-Webber, clinical librarian, for helping to develop our search strategy and Gordon Guyatt, GRADE Working Group Co-Chair, for advice on applying the GRADE methods and for reviewing the manuscript.

Financial Support: This project was funded by the Technical Committee on Dietary Carbohydrates of ILSI North America. The authors wrote the protocol, the scope of which was reviewed and approved by ILSI; and conducted the study independently from ILSI.

Disclosures: Dr. Slavin served on the 2010 Dietary Guidelines for Americans Advisory Committee (DGAC) where she chaired the carbohydrate committee that reviewed the relationships between added sugar intake and health outcomes. The results of that review were published in the 2010 DGACs, in *Nutrition Reviews* (Slavin J. Beverages and body weight: challenges in the evidence-based review process of the Carbohydrate Subcommittee from the 2010 Dietary Guidelines Advisory Committee. Nutr Rev. 2012;70 Suppl 2:S111-20.). She has presented widely on her work as chair of the carbohydrate and protein committees for the 2010 DGAC. Most of her research is in the areas of dietary fiber and gut health. As a dietitian, she is interested in dietary patterns and whole foods. Her research funding in the area of dietary sugars is summarized below. She received a grant from ILSI-NA Carbohydrate Committee in 2010 to examine sugar recommendations. That work was published in 2012 (Hess J, Latulippe ME, Ayoob K, Slavin J. The confusing world of dietary sugars: definitions, intakes, food sources and international dietary recommendations. Food Funct. 2012;3:477-86.). One of the coauthors of that paper was an employee of ILSI-NA at the time. That information is disclosed in the paper. Dr. Slavin and the University of Minnesota received the grant from ILSI-NA to support the current project. Besides ILSI-NA, Dr. Slavin thanks the following organizations for providing research funds for her laboratory the past 3 years: Minnesota Beef Council (satiety), Minnesota Cultivated Wild Rice Council (literature review), Barilla (snacking), Novartis Consumer Health (GSK) (fiber), American Pulse Association (satiety), MNDrive Global Food Ventures (nutrients in spinach), United States Department of Agriculture (fiber), The Mushroom Council (gut health), Pepsico (oatmeal), Welch's (FODMAPs), Nestle Health Sciences (FODMAPs), and DSM (fiber). Her laboratory also has received contracts for services for analytical services in the areas of dietary fiber, whole grains, legumes, FODMAPS, digestive health, protein needs, carbohydrate needs, and snacking: Besides the companies listed, the laboratory has received funds in the past 3 years from Danone (snacking) and Coca-Cola (fiber). When the work is published, the funding source for all work in the laboratory will be disclosed as outlined by the journal. Dr. Slavin speaks widely on a range of human nutrition topics. Some talks on the topic of interest in this paper: "Fluid Consumption: Caloric Contribution to Weight Gain/Loss and Health: Factors That Influence Satiety" (Second International Conference on hydration and Health, sponsored by the ILSI North America Committee on Hydration, November 2011); "The Confusing World of Dietary Sugars: Views From the 2010 Dietary Guidelines Scientific Advisory Committee" (2012 Nutrition News Forecast, Academy of Nutrition and Dietetics, April 2012); "Food Is Not a Talisman: Reflections on the Science and Practice of Nutrition" (WO Atwater Lecture at Experimental Biology, April 2015). For full financial disclosure: ILSI meetings do not pay speakers; other scientific meetings also typically do not pay speakers if you are a member of that society. Dr. Slavin serves on the scientific advisory board for Tate and Lyle, Kerry Ingredients, Atkins Nutritionals, and Midwest Dairy Association. She also owns one-third share of the Slavin Sisters Farm LLC, a 119-acre farm in Walworth, Wisconsin, that is currently rented. Crops in 2016 included corn, soybeans, and pumpkins. Dr. Johnston is a member of GRADE, a working group that has developed a common, sensible, and transparent approach to grading quality of evidence and strength of recommendations. In addition to being a methods consultant to ILSI for this project, over the last 5 years, he has held investigator-initiated grants unrelated to the topic of sugar from BioK+ (a probiotic manufacturer), Genzyme (a manufacturer of enzyme replacement therapy for patients with rare lysosomal storage diseases), and a joint grant funded by Nestle and MITACS Accelerate (a provincially and federally supported not-for-profit organization that works with Canadian universities and companies to build partnerships that support industrial and social innovation in Canada) to assess probiotics for preventing necrotizing enterocolitis. Authors not named here have disclosed no conflicts of interest. Disclosures can also be viewed at www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M16-2020.

Reproducible Research Statement: *Study protocol:* Available at www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42015029182. *Statistical code:* Not applicable. *Data set:* See tables and appendices for all relevant data.

Requests for Single Reprints: Bradley C. Johnston, PhD, Prevention Lab, Child Health Evaluative Sciences, The Hospital for Sick Children Research Institute, Peter Gilgan Centre for Research and Learning, 686 Bay Street, Room 11.9859 West, Toronto, Ontario M5G 0A4, Canada; e-mail, bradley.johnston@sickkids.ca.

Current author addresses and author contributions are available at Annals.org.

References

References

1. **Ruxton CH, Gardner EJ, McNulty HM.** Is sugar consumption detrimental to health? A review of the evidence 1995–2006. Crit Rev Food Sci Nutr. 2010;50:1-19. [PMID: 20047137] doi:10.1080/10408390802248569
2. **Hess J, Latulippe ME, Ayoob K, Slavin J.** The confusing world of dietary sugars: definitions, intakes, food sources and international dietary recommendations. Food Funct. 2012;3:477-86. [PMID: 22402777] doi:10.1039/c2fo10250a
3. **Erickson J, Slavin J.** Total, added, and free sugars: are restrictive guidelines science-based or achievable? Nutrients. 2015;7:2866-78. [PMID: 25884659] doi:10.3390/nu7042866
4. **World Health Organization.** Guideline: Sugars Intake for Adults and Children. Geneva: World Health Organization; 2015.
5. **Public Health England, Scientific Advisory Committee on Nutrition.** Carbohydrates and Health. London: Public Health England; 2015.
6. **U.S. Department of Agriculture, U.S. Department of Health and Human Services.** Dietary Guidelines for Americans, 2015–2020. 8th ed. Washington, DC: US Gov Pr Off; 2015.
7. **Sadeghirad B, Erickson J, Lytvyn L, Webber-Adams T, Slavin J, Johnston B.** Scientific basis for recommendations on sugars from authoritative health organizations: a systematic review of public health guidelines. PROSPERO: CRD42015029182. Accessed at www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42015029182 on 16 November 2016.
8. **AGREE Collaboration.** Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. Qual Saf Health Care. 2003;12:18-23. [PMID: 12571340]
9. **Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al; GRADE Working Group.** GRADE: an emerging consensus on rating quality of evidence and strength of recommendations.BMJ.2008;336:924-6.[PMID:18436948]doi:10.1136/bmj.39489.470347.AD
10. **Cohen J.** Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull. 1968;70:213-20. [PMID: 19673146]
11. **Koch G.** Intraclass correlation coefficient. In: Kotz S, Read C, Balakrishnan N, Vidakovic B, eds. Encyclopedia of Statistical Sciences. 4th ed. New York: J Wiley; 1982:213-7.
12. **Kramer MS, Feinstein AR.** Clinical biostatistics. LIV. The biostatistics of concordance. Clin Pharmacol Ther. 1981;29:111-23. [PMID: 7460469]
13. **Institute of Medicine, Food and Nutrition Board.** Dietary Reference Intakes for Energy, Carbohydrate, Fiber, Fat, Fatty Acids, Cholesterol, Protein and Amino Acids. Washington, DC: National Academies Pr; 2002.
14. **Nordic Council of Ministers.** Nordic Nutrition Recommendations 2012: Integrating Nutrition and Physical Activity. Copenhagen: Nordisk Ministerråd; 2012.
15. **Ministry of Health of Brazil, Secretariat of Health Care, Primary Health Care Department.** Dietary Guidelines for the Brazilian Population. Brasilia, DF, Brazil: Ministry of Health of Brazil; 2014.
16. **Hauner H, Bechthold A, Boeing H, Brönstrup A, Buyken A, Leschik-Bonnet E, et al; German Nutrition Society.** Evidence-based guideline of the German Nutrition Society: carbohydrate intake and prevention of nutrition-related diseases. Ann Nutr Metab. 2012;60 Suppl 1:1-58. [PMID: 22286913] doi:10.1159/000335326
17. **Food Safety Authority of Ireland.** Scientific Recommendations for Healthy Eating Guidelines in Ireland. Dublin: Food Safety Authority of Ireland; 2011.
18. **National Health and Medical Research Council.** Australian Dietary Guidelines. Canberra: National Health and Medical Research Council; 2013. Accessed at www.nhmrc.gov.au/guidelines-publications/n55 on 16 November 2016.
19. **Becker W.** Nordic Nutrition Recommendations 2004, based on scientific evidence. Scandinavian Journal of Nutrition. 2005;49:68-71.
20. **Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, et al; AGREE Next Steps Consortium.** AGREE II: advancing guideline development, reporting and evaluation in health care. CMAJ. 2010;182:E839-42. [PMID: 20603348] doi:10.1503/cmaj.090449
21. **Chen L, Appel LJ, Loria C, Lin PH, Champagne CM, Elmer PJ, et al.** Reduction in consumption of sugar-sweetened beverages is associated with weight loss: the PREMIER trial. Am J Clin Nutr. 2009;89:1299-306. [PMID: 19339405] doi:10.3945/ajcn.2008.27240
22. **Sichieri R, Paula Trotte A, de Souza RA, Veiga GV.** School randomised trial on prevention of excessive weight gain by discouraging students from drinking sodas. Public Health Nutr. 2009;12:197-202. [PMID: 18559131] doi:10.1017/S1368980008002644
23. **Tucker KL, Morita K, Qiao N, Hannan MT, Cupples LA, Kiel DP.** Colas, but not other carbonated beverages, are associated with low bone mineral density in older women: The Framingham Osteoporosis Study. Am J Clin Nutr. 2006;84:936-42. [PMID: 17023723]
24. **Duffey KJ, Gordon-Larsen P, Steffen LM, Jacobs DR Jr, Popkin BM.** Drinking caloric beverages increases the risk of adverse cardiometabolic outcomes in the Coronary Artery Risk Development in Young Adults (CARDIA) Study. Am J Clin Nutr. 2010;92:954-9. [PMID: 20702604] doi:10.3945/ajcn.2010.29478
25. **Cohen L, Curhan G, Forman J.** Association of sweetened beverage intake with incident hypertension. J Gen Intern Med. 2012;27:1127-34. [PMID: 22539069] doi:10.1007/s11606-012-2069-6
26. **Nissinen K, Mikkilä V, Männistö S, Lahti-Koski M, Räsänen L, Viikari J, et al.** Sweets and sugar-sweetened soft drink intake in childhood in relation to adult BMI and overweight. The Cardiovascular Risk in Young Finns Study. Public Health Nutr. 2009;12:2018-26. [PMID: 19476678] doi:10.1017/S1368980009005849
27. **Dhingra R, Sullivan L, Jacques PF, Wang TJ, Fox CS, Meigs JB, et al.** Soft drink consumption and risk of developing cardiometabolic risk factors and the metabolic syndrome in middle-aged adults in the community. Circulation. 2007;116:480-8. [PMID: 17646581]
28. **Paynter NP, Yeh HC, Voutilainen S, Schmidt MI, Heiss G, Folsom AR, et al.** Coffee and sweetened beverage consumption and the risk of type 2 diabetes mellitus: the Atherosclerosis Risk in Communities Study. Am J Epidemiol. 2006;164:1075-84. [PMID: 16982672]
29. **Palmer JR, Boggs DA, Krishnan S, Hu FB, Singer M, Rosenberg L.** Sugar-sweetened beverages and incidence of type 2 diabetes mellitus in African American women. Arch Intern Med. 2008;168:1487-92. [PMID: 18663160] doi:10.1001/archinte.168.14.1487
30. **Schulze MB, Manson JE, Ludwig DS, Colditz GA, Stampfer MJ, Willett WC, et al.** Sugar-sweetened beverages, weight gain, and incidence of type 2 diabetes in young and middle-aged women. JAMA. 2004;292:927-34. [PMID: 15328324]
31. **U.S. Department of Agriculture, U.S. Department of Health and Human Services.** Scientific Report of the 2015 Dietary Guidelines Advisory Committee. 2015.
32. **Moynihan PJ, Kelly SA.** Effect on caries of restricting sugars intake: systematic review to inform WHO guidelines. J Dent Res. 2014;93:8-18. [PMID: 24323509] doi:10.1177/0022034513508954
33. **Te Morenga L, Mallard S, Mann J.** Dietary sugars and body weight: systematic review and meta-analyses of randomised controlled trials and cohort studies. BMJ. 2012;346:e7492. [PMID: 23321486] doi:10.1136/bmj.e7492
34. **Andrews JC, Schünemann HJ, Oxman AD, Pottie K, Meerpohl JJ, Coello PA, et al.** GRADE guidelines: 15. Going from evidence to recommendation-determinants of a recommendation's direction and strength. J Clin Epidemiol. 2013;66:726-35. [PMID: 23570745] doi:10.1016/j.jclinepi.2013.02.003
35. **Council on School Health.** Snacks, sweetened beverages, added sugars, and schools. Pediatrics. 2015;135:575-83. [PMID: 25713277] doi:10.1542/peds.2014-3902
36. **EFSA Panel on Dietetic Products, Nutrition, and Allergies.** Scientific opinion on dietary reference values for carbohydrates and dietary fibre. EFSA Journal. 2010;8:1462.

37. **Johnson RK, Appel LJ, Brands M, Howard BV, Lefevre M, Lustig RH, et al; American Heart Association Nutrition Committee of the Council on Nutrition, Physical Activity, and Metabolism and the Council on Epidemiology and Prevention.** Dietary sugars intake and cardiovascular health: a scientific statement from the American Heart Association. Circulation. 2009;120:1011-20. [PMID: 19704096] doi: 10.1161/CIRCULATIONAHA.109.192627
38. **National Institute of Nutrition.** Dietary Guidelines for Indians: A Manual. 2nd ed. Hyderabad, India: National Institute of Nutrition; 2011.
39. GRADE Working Group Web site. Accessed at http://gradeworkinggroup.org on 19 October 2016.
40. **Greenfield S, Steinberg E, Auerbach A, Avorn J, Galvin R, Gibbons R, et al.** Clinical Practice Guidelines We Can Trust. Washington, DC: Institute of Medicine; 2011.

Current Author Addresses: Ms. Erickson and Dr. Slavin: Department of Food Science and Nutrition, University of Minnesota, 1334 Eckles Avenue, St. Paul, MN 55108.
Dr. Sadeghirad: Department of Clinical Epidemiology & Biostatistics, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4K1, Canada.
Ms. Lytvyn and Dr. Johnston: Child Health Evaluative Sciences, The Hospital for Sick Children Research Institute, Peter Gilgan Centre for Research and Learning, 686 Bay Street, Room 11.9859 West, Toronto, Ontario M5G 0A4, Canada.

Author Contributions: Conception and design: L. Lytvyn, J. Slavin, B.C. Johnston.
Analysis and interpretation of the data: J. Erickson, B. Sadeghirad, L. Lytvyn, J. Slavin, B.C. Johnston.
Drafting of the article: J. Erickson, B. Sadeghirad, L. Lytvyn, B.C. Johnston.
Critical revision of the article for important intellectual content: J. Erickson, B. Sadeghirad, L. Lytvyn, J. Slavin, B.C. Johnston.
Final approval of the article: J. Erickson, B. Sadeghirad, L. Lytvyn, J. Slavin, B.C. Johnston.
Provision of study materials or patients: J. Slavin, B.C. Johnston.
Statistical expertise: B. Sadeghirad.
Obtaining of funding: J. Slavin, B.C. Johnston.
Administrative, technical, or logistic support: L. Lytvyn, J. Slavin, B.C. Johnston.
Collection and assembly of data: J. Erickson, B. Sadeghirad, L. Lytvyn, B.C. Johnston.

Web-Only References

41. **Forshee RA, Anderson PA, Storey ML.** Sugar-sweetened beverages and body mass index in children and adolescents: a meta-analysis. Am J Clin Nutr. 2008;87:1662-71. [PMID: 18541554]
42. **Gibson S.** Sugar-sweetened soft drinks and obesity: a systematic review of the evidence from observational studies and interventions. Nutr Res Rev. 2008;21:134-47. [PMID: 19087367] doi:10.1017/S0954422408110976
43. **Malik VS, Schulze MB, Hu FB.** Intake of sugar-sweetened beverages and weight gain: a systematic review. Am J Clin Nutr. 2006;84:274-88. [PMID: 16895873]
44. **Vartanian LR, Schwartz MB, Brownell KD.** Effects of soft drink consumption on nutrition and health: a systematic review and meta-analysis. Am J Public Health. 2007;97:667-75. [PMID: 17329656]
45. **Wolff E, Dansinger ML.** Soft drinks and weight gain: how strong is the link? Medscape J Med. 2008;10:189. [PMID: 18924641]
46. **Anderson CA, Curzon ME, Van Loveren C, Tatsi C, Duggal MS.** Sucrose and dental caries: a review of the evidence. Obes Rev. 2009;10 Suppl 1:41-54. [PMID: 19207535] doi:10.1111/j.1467-789X.2008.00564.x
47. **Sonestedt E, Overby NC, Laaksonen DE, Birgisdottir BE.** Does high sugar consumption exacerbate cardiometabolic risk factors and increase the risk of type 2 diabetes and cardiovascular disease? Food Nutr Res. 2012;56.
48. **Zhang YH, An T, Zhang RC, Zhou Q, Huang Y, Zhang J.** Very high fructose intake increases serum LDL-cholesterol and total cholesterol: a meta-analysis of controlled feeding trials. J Nutr. 2013;143:1391-8. [PMID: 23825185] doi:10.3945/jn.113.175323
49. **Malik VS, Popkin BM, Bray GA, Després JP, Willett WC, Hu FB.** Sugar-sweetened beverages and risk of metabolic syndrome and type 2 diabetes: a meta-analysis. Diabetes Care. 2010;33:2477-83. [PMID: 20693348] doi:10.2337/dc10-1079
50. **Fogelholm M, Anderssen S, Gunnarsdottir I, Lahti-Koski M.** Dietary macronutrients and food consumption as determinants of long-term weight change in adult populations: a systematic literature review. Food Nutr Res. 2012;56.
51. **Burt BA, Pai S.** Sugar consumption and caries risk: a systematic review. J Dent Educ. 2001;65:1017-23. [PMID: 11699972]
52. **Mattes RD, Shikany JM, Kaiser KA, Allison DB.** Nutritively sweetened beverage consumption and body weight: a systematic review and meta-analysis of randomized experiments. Obes Rev. 2011;12:346-65. [PMID: 20524996] doi:10.1111/j.1467-789X.2010.00755.x
53. **Nutritional Epidemiology Group, University of Leeds.** A systematic review of the evidence of the benefits and risks of different dietary carbohydrates on cardio-metabolic health and disease. July 2012. Accessed at www.nutritionsociety.org/sites/www.nutritionsociety.org/files/02%20-%20Cardiometabolic%20Health%20Report%20Introduction.pdf on 16 November 2016.

Appendices

Appendix 2. 1. Grey literature sources

1. National Guidelines Clearinghouse
2. National Institute for Health and Care Excellence
3. Scottish Intercollegiate Guidelines Network (SIGN)
4. Guidelines International Network
5. Google internet search engine (terms searched: “sugar guidelines” or “recommend* daily sugar”; limited to sites ending in “.gov” or “.org”; limited to the first 20 pages)

Experts in carbohydrates contacted in search for public health guidelines

Dr. John L. Sievenpiper, MD, PhD, FRCPC

Associate Professor, Department of Nutritional Sciences, University of Toronto; Scientist, Li Ka Shing Knowledge Institute, St. Michael's Hospital; Consultant Physician, Division of Endocrinology & Metabolism, St. Michael's Hospital

Dr. Julie Miller Jones, PhD, CNS, LN, Fellow of AACCI and ICC

Distinguished Scholar and Professor Emerita, Foods and Nutrition, St. Catherine University

Dr. Keith-Thomas Ayoob, EdD, RD, FAND

Associate Clinical Professor, Department of Pediatrics, Albert Einstein College of Medicine

Appendix 2. 2. Appraisal of Guidelines Research and Evaluation (AGREE) II instrument

AGREE II Instrument.	
Domain	**Item**
Scope and purpose	1. The overall objective(s) of the guideline is (are) specifically described.
	2. The health questions covered by the guideline are specifically described.
	3. The population (patients, public, etc.) to whom the guideline is meant to apply is specifically described.
Stakeholder involvement	4. The guideline development group includes individuals from all the relevant professional groups.
	5. The views and preferences of the target population (patients, public, etc.) have been sought.
	6. The target users of the guideline are clearly defined.
Rigor of development	7. Systematic methods were used to search for evidence.
	8. The criteria for selecting the evidence are clearly described.
	9. The strengths and limitations of the body of evidence are clearly described.
	10. The methods for formulating the recommendations are clearly described.
	11. The health benefits, side effects and risks have been considered in formulating the recommendations.
	12. There is explicit link between recommendations and supporting evidence.
	13. The guideline has been externally reviewed by experts prior to publication.
	14. A procedure for updating the guideline is provided.
Clarity of presentation	15. The recommendations are specific and unambiguous.
	16. The different options for management of the condition or health issue are clearly presented.
	17. Key recommendations are easily identifiable.
Applicability	18. The guideline describes facilitators and barriers to its application.
	19. The guideline provides advice and/or tools on how the recommendations can be put into practice.
	20. The potential resource implications of applying the recommendations have been considered.
	21. The guideline presents monitoring and/ or auditing criteria.
Editorial independence	22. The views of funding body have not influenced content of the guideline.
	23. Competing interests of guideline development group members have been recorded and addressed.
Overall Guideline Assessment	1. Rate the overall quality of this guideline.
Overall Guideline Assessment	2. I would recommend this guideline for use.

Appendix 2. 3. Conflicts of interest reporting across guidelines

Appendix Table 3. COI Reporting Across Guidelines

Guideline	COI Process Reporting	Groups Requiring COIs (Number of Members)	COI Reporting		
			Affiliation	Financial	Intellectual
Australian Dietary Guidelines	Unclear	Dietary guidelines working committee (11)	Yes	No	Marginally–lists their research focuses
		National Health and Medical Research Council project team (4), Department of Health and Ageing Project Team (5), contractors (8), expert reviewers (5)	Yes	No	No
		Public consultation contributors; 2 rounds	No	No	No
Dietary Guidelines for the Brazilian Population	No	Listening workshop (59), evaluation workshop (29), working group for consideration of public consultation (10)	Yes	No	No
		Public consultation contributors	No	No	No
Nordic Nutrition Recommendations	No	Working group (11), topic experts ("over 100"; for carbohydrates = 4), topic peer reviewers (unspecified; for carbohydrates = 2), reference group of senior experts (9), steering group with representatives from each national authority (5), librarians (5)	No (country only)	No	No
		Public consultation contributors	No	No	No
Scientific Recommendations for Healthy Eating Guidelines in Ireland	No	Steering committee (11)	Yes (except for contract researcher)	No	No
		Research team (11), Irish Nutrition and Dietetic Institute (unspecified), consultation day contributors (e.g., dietitians, nutritionists, Irish Nutrition and Dietetic Institute members, Irish Heart Foundation; unspecified)	No	No	No
		Nutrition and novel foods subcommittee	Yes	No	No
Evidence-based Guideline of the German Nutrition Society	No	Authors of publication	Yes	No	No
DRI/Institute of Medicine	No	Panel on DRI for macronutrients (21), panel on the definition of dietary fiber (7), subcommittee on upper reference levels of nutrients (10), subcommittee on interpretation and uses of DRI (8)	Yes	Marginally–lists some industry work	Marginally–lists their research focuses
		Staff macronutrient panel (8), staff fiber panel (7), staff upper reference levels panel (2), staff interpretation/use panel (3), staff standing committee (8), staff food and nutrition board (5), individuals who provided input (31 and some "unnamed"), federal DRI working committee (23)	No	No	No
		Consultants (2), standing committee on the scientific evaluations of DRI (9), technical advisor to the DRI projects (1), U.S. government liaison (1), Canadian government liaison (1), food and nutrition board (15), independent reviewers (18), independent reviewers (18)	Yes	No	No
		Organizations, including industry (15)	NA (these are organizations, not people)	No	No

Continued on following page

Appendix Table 3–Continued

Guideline	COI Process Reporting	Groups Requiring COIs (Number of Members)	COI Reporting		
			Affiliation	Financial	Intellectual
2015–2020 Dietary Guidelines for Americans	Unclear	Federal advisory committee, divided into subcommittees for each chapter (14)	Yes	No	Marginally–lists their research focuses
		Consultant subcommittee members (3)	Yes	No*	Marginally–lists their research focuses
		Co-executive secretaries (4), policy officials (5), dietary guidelines management team (17), nutrition evidence library team (13), data analysis team (18), science writer/editor (1), public consultation contributors throughout commentary period up to December 2014 (918 [number of comments accounted for; comments may be by person or group or organization]), public consultation contributors 2014 meeting (53), public consultation 2015 meeting (73)	Yes	No	No
		Invited expert speakers (32), staff, contract, and/or technical support (20), national service volunteer evidence abstractors (28)	No	No	No
Carbohydrates and Health (SACN)	No	Membership of Scientific Advisory Committee on Nutrition: Carbohydrates Working Group (12), observers (4), observers carbohydrates working group (1), external consultants carb working group (5)	Yes	No	No
		Membership of Scientific Advisory Committee on Nutrition (16)	Yes (but unclear affiliation for 2 members)	No	No
		Secretariat nutrition committee (15), secretariat carb working group (12), public consultation contributors (unspecified)	No	No	No
Sugars Intake for Adults and Children (WHO)	Unclear	WHO secretariat headquarters (11), WHO secretariat regional offices (11), members of the WHO Steering Committee for Nutrition Guideline Development 2012–2014 (17), public consultation contributors planning stage (18), public consultation contributors draft stage (173)	Yes	No	No
		Members of the guideline development group Nutrition Guidance Expert Advisory Group Subgroup on Diet and Health (15), external resource persons 2012–2014 (10), external peer review group (6)	Yes	Yes	Yes

COI = conflict of interest; DRI = Dietary Reference Intakes; NA = not applicable; SACN = Scientific Advisory Committee on Nutrition; WHO = World Health Organization.
* Dietary Guidelines for Americans do, however, state, "Per Federal Advisory Committee Act rules, Advisory Committee members were thoroughly vetted for conflicts of interest before they were appointed to their positions and were required to submit a financial disclosure form annually."

Appendix 2. 4. Assessment of the supporting evidence for each recommendation (GRADE)

Appendix Table 4. Assessment of the Supporting Evidence for Each Recommendation (GRADE)

Guideline Title	Overall Recommendation	Specific Recommendations, Including Strength (if Reported)	Citations Supporting Recommendation, *n*	Study Design	GRADE Evidence Quality (Certainty in Estimates of Effect)
Sugars Intake for Adults and Children (WHO)*	–	"Reduced intake of free sugars throughout the life course–Strong Recommendation"	0	–	NA
		"In both adults and children, WHO recommends reducing the intake of free sugars to less than 10% of total energy intake–Strong Recommendation"	1	Systematic review	Low†
		"WHO suggests further reduction of the intake of free sugars to below 5% of total energy intake–Conditional Recommendation"	1	Systematic review	Very low
Carbohydrates and Health (Public Health England)‡	"The population average intake of free sugars should not exceed 5% of total dietary energy for age groups from 2 years upwards" and "The consumption of sugars-sweetened beverages should be minimised, in both children and adults."	"Greater sugar intake is associated with increased energy intake–Adequate Evidence" and "Sugar sweetened beverage intake is associated with risk of type-2 diabetes–Moderate Evidence"	1	Systematic review	Very low
		"Sugar consumption is associated with increased risk of dental caries–Moderate Evidence" and "Amount and frequency of SSB consumption is associated with dental caries–Adequate Evidence" and "Greater SSB consumption is associated with increased BMI–Limited Evidence"	1	Systematic review	Very low
Australian Dietary Guidelines	"Limit intake of foods and drinks containing added sugars such as confectionary, sugar-sweetened soft drinks and cordials, fruit drinks, vitamin waters, energy and sports drinks"	"Consumption of sugar-sweetened beverages is associated with increased risk of weight gain in adults and children–Grade B"	15	Systematic review; randomized, controlled trial; observational study	Low, very low
		"High or frequent consumption of added sugars, particularly for infants and young children, is associated with increased risk of dental caries–Grade C"	1	Observational study	Very low
		"Consumption of soft drinks is associated with increased risk of dental caries in children–Grade C"	1	Observational study	Very low
		"Consumption of soft drinks is associated with increased risk of reduced bone strength–Grade C"	3	Randomized, controlled trial; observational study	Very low
Nordic Nutrition Recommendations	"Intake of added sugars should be kept below 10% of the energy intake"	–	14	Systematic review; observational study	Low, very low
Evidence-based Guideline of the German Nutrition Society: Carbohydrate Intake and the Prevention of Nutrition-Related Diseases	"The consumption of sugar-sweetened beverages should be limited, because they increase the risk of obesity and diabetes"	"The available cohort and intervention studies regarding adults mainly show that a higher consumption of SSB is accompanied by an increased risk of obesity–Probable"	6	Systematic review; randomized, controlled trial; observational study	Low, very low
		"The majority of prospective cohort studies and meta analysis indicate an increased risk of type 2 diabetes with regular consumption of sugar sweetened beverages–Probable"	5	Systematic review; observational study	Low, very low

Appendix Table 4–Continued

Guideline Title	Overall Recommendation	Specific Recommendations, Including Strength (if Reported)	Citations Supporting Recommendation, *n*	Study Design	GRADE Evidence Quality (Certainty in Estimates of Effect)
Scientific Recommendations for Healthy Eating Guidelines in Ireland	"Healthy eating can be enjoyed with limited amounts of `other foods' like biscuits, cakes, savoury snacks and confectionery. These foods are rich in calories, fat, sugar and salt so remember–NOT too MUCH and NOT too OFTEN"	–	6	Randomized, controlled trial; narrative review or report	Very low
Dietary Reference Intakes for Energy, Carbohydrate, Fiber, Fat, Fatty Acids, Cholesterol, Protein and Amino Acids	A maximal intake level of 25% or less of energy is suggested to prevent the displacement of foods that are major sources of essential micronutrients.	–	7	Observational study	Very low
2015-2020 Dietary Guidelines for Americans§	"Consume less than 10% of calories per day from added sugars"	–	0	–	NA
Dietary Guidelines for the Brazilian Population	"Use oils, fats, salt, and sugar in small amounts for seasoning and cooking foods and to create culinary preparations"	–	0	–	NA

BMI = body mass index; GRADE = Grading of Recommendations Assessment, Development and Evaluation; NA = not applicable; SSB = sugar-sweetened beverage; WHO = World Health Organization.
* A systematic review on sugars and weight was conducted and referenced. However, authors did not look specifically at 10% reduction; only the effect of sugar on dental caries was cited for the final 2 of 3 recommendations.
† The WHO rated the quality of evidence as "moderate"; however, in our independent assessment, we considered WHO's reasoning for rating up from low to be inappropriate.
‡ Public Health England conducted its own systematic reviews that were unpublished.
§ A rigorous scientific report of unpublished systematic reviews was conducted but was not used to make recommendation.

Appendix 2. 5. Assessment of individual studies supporting recommendations (GRADE)

Appendix Table 5. Assessment of Individual Studies Supporting Recommendations (GRADE)

Study, Year (Reference)	Guidelines That Included the Study	GRADE	Reasons for Rating Up or Down
Systematic reviews			
Forshee et al, 2008 (41)	Australia 2013, Nordic 2012	Very low	Inconsistency, imprecision, publication bias
Gibson, 2008 (42)	Australia 2013	Very low	Inconsistency, imprecision
Malik et al, 2006 (43)	Australia 2013	Very low	Inconsistency, imprecision
Vartanian et al, 2007 (44)	Australia 2013, Germany 2012, Nordic 2012, Germany 2012	Low	Inconsistency, imprecision
Wolff and Dansinger, 2008 (45)	Australia 2013	Very low	Inconsistency, imprecision
Anderson et al, 2009 (46)	Australia 2013, Nordic 2012	Very low	Inconsistency
Sonestedt et al, 2012 (47)	Nordic 2012	Low	Inconsistency
Te Morenga et al, 2012 (33)	WHO 2015, Nordic 2012	Low	Inconsistency, publication bias
Zhang et al, 2013 (48)	Nordic 2012	Low	Indirectness, imprecision
Malik et al, 2010 (49)	Nordic 2012	Very low	Inconsistency
Fogelholm et al, 2012 (50)	Nordic 2012	Very low	Imprecision
Burt and Pai, 2001 (51)	Nordic 2012	Very low	Inconsistency
Moynihan and Kelly, 2014 (32)	WHO 2015	Very low	Inconsistency, imprecision
Mattes et al, 2011 (52)	Germany 2012	Low	Inconsistency, imprecision
Nutritional Epidemiology Group, 2012 (53)	SACN 2015	Very low	Inconsistency, indirectness, imprecision
SACN, 2011 (unpublished)	SACN 2015	Very low	Inconsistency, imprecision
Randomized, controlled trials			
Sichieri et al, 2009 (22)	Australia 2013, Germany 2012	Very low	Imprecision, indirectness
Chen et al, 2009 (21)	Germany 2012	Very low	Imprecision, indirectness
Cohort studies			
Tucker et al, 2006 (23)	Australia 2013	Very low	Indirectness, imprecision
Duffey et al, 2010 (24)	Nordic 2012	Very low	Imprecision
Cohen et al, 2012 (25)	Nordic 2012	Very low	Indirectness
Nissinen et al, 2009 (26)	Germany 2012	Very low	Indirectness
Dhingra et al, 2007 (27)	Germany 2012	Very low	None
Schulze et al, 2004 (30)	Germany 2012	Low	Dose-response
Palmer et al, 2008 (29)	Germany 2012	Low	Dose-response
Paynter et al, 2006 (28)	Germany 2012	Very low	None

GRADE = Grading of Recommendations Assessment, Development and Evaluation; SACN = Scientific Advisory Committee on Nutrition; WHO = World Health Organization.

Chapter 3: Influence of unhealthy food and beverage marketing on children's dietary intake and preference: a systematic review and meta-analysis of randomized trials

Behnam Sadeghirad, Tara Duhaney, Shahrzad Motaghipisheh, Norm R.C. Campbell,

Bradley C. Johnston

This Chapter was accepted for publication on 31 May, 2016 and was available online on 18 July 2016. In this dissertation, we present the revised submitted version. the final published paper is available as: Sadeghirad B, Duhaney T, Motaghipisheh S, Campbell NR, Johnston BC. Influence of unhealthy food and beverage marketing on children's dietary intake and preference: a systematic review and meta-analysis of randomized trials. Obesity Reviews, 2016; 17(10):945-59. [PMID: 27427474] doi: 10.1111/obr.12445.

Author Contributions:

Study concept and design: NRCC and BCJ.

Acquisition, analysis, or interpretation of data: TD, BS, SM, and BCJ.

Statistical analysis: BS and BCJ.

Drafting of the manuscript: BS and TD.

Critical revision of the manuscript for important intellectual content: All authors.

Study supervision: BCJ and NRCC.

Influence of unhealthy food and beverage marketing on children's dietary intake and preference: a systematic review and meta-analysis of randomized trials

Behnam Sadeghirad [1,2,3]; Tara Duhaney [4]; Shahrzad Motaghipisheh [5]; Norm R.C. Campbell [4,6.7]; Bradley C. Johnston [1,2,8,9]

Authors' affiliations

1. Systematic Overviews through advancing Research Technology (SORT), Child Health Evaluative Sciences, The Hospital for Sick Children Research Institute, Toronto, Ontario, Canada
2. Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada
3. Regional Knowledge Hub and WHO Collaborating Centre for HIV Surveillance, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran
4. Libin Cardiovascular Institute of Alberta, University of Calgary, Calgary, Alberta, Canada
5. Research Center for Modeling in Health, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran
6. Departments of Medicine, Physiology and Pharmacology and Community Health Sciences, University of Calgary, Calgary, Alberta, Canada

7. O'Brien Institute of Public Health, University of Calgary, Calgary, Alberta, Canada
8. Department of Anesthesia and Pain Medicine, The Hospital for Sick Children, Univeristy of Toronto, Toronto, Ontario
9. Institute for Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, Canada

Keywords: meta-analysis; unhealthy, food and beverage marketing; dietary intake; dietary preference; randomized trial

Acknowledgements

We would like to thank Gordon Guyatt, MD, Distinguished University Professor in the Departments of Clinical Epidemiology & Biostatistics and Medicine at McMaster University for his methodological advice.

Corresponding author: Bradley C. Johnston

Systematic Overviews through advancing Research Technology (SORT), Child Health Evaluative Sciences, The Hospital for Sick Children Research Institute

Peter Gilgan Centre for Research and Learning

686 Bay Street, Rm 11.9859 West

Toronto, Ontario, Canada, M5G 0A4

Email: bradley.johnston@sickkids.ca

Conflict of interest

Authors declare no conflicts of interest.

Abstract

Media exposure and marketing of foods and beverages high in fat, sugar, and salt are suggested to be major contributors to poor dietary behaviors in children and related dietary diseases. Our objective was to assess the effects of unhealthy food and beverage marketing on dietary intake and dietary preference among children 2 to 18 years of age. We searched MEDLINE, EMBASE and PsycINFO up to January 2015 for terms related to advertising, unhealthy foods or beverages among children. Randomized trials that assessed the effects of unhealthy food and beverage marketing compared to non-dietary advertisement or no advertisement in children were considered eligible. Two authors independently extracted information on study characteristics and outcomes of interest and assessed risk of bias and the overall quality of evidence using GRADE methodology. Meta-analysis was conducted separately for dietary intake and preference using a random effects model. We identified 29 eligible studies, of which 17 were included for meta-analysis of dietary preference and 9 for meta-analysis of dietary intake. Almost half of the studies were at high risk of bias. Our meta-analysis showed that in children exposed to unhealthy dietary marketing, dietary intake significantly increased (MD= 30.4 kcals, 95% CIs: 2.9 to 57.9, and MD= 4.8 grams, 95%CI: 0.8 to 8.8) during or shortly after exposure to advertisements. Similarly, children exposed to the unhealthy dietary marketing had a higher risk of selecting the advertised foods or beverages (RR= 1.1, 95% CIs: 1.0 to 1.2; p = 0.052). The evidence indicates that unhealthy food and beverage marketing increases dietary intake (moderate certainty) and preference (moderate to low certainty) for energy-dense, low nutrition food and beverage. Unhealthy food and beverage marketing increased dietary intake and influenced dietary preference in children during or shortly after exposure to advertisements.

Introduction

The rates of overweight and obesity among children are rising worldwide.[1, 2] Obesity is one of the major predisposing factors of most non-communicable diseases and it is associated with a lower life expectancy.[3, 4] Unhealthy diet and the food and beverage environment which perpetuates poor dietary behaviors are suggested to play major roles in the global obesity epidemic.[5, 6] In 2010, unhealthy diet was the leading risk factor for death and disability globally.[7, 8]

There is increasing and consistent evidence that unhealthy food and beverage marketing directed at children negatively impacts their eating behaviors.[9] The increasing prevalence of obesity seems to further coincide with marked increases in the food and beverage industry's budget for marketing aimed at children and youth[10], with data showing that energy-dense, low-nutrient foods and beverages make up the majority of commercially marketed products.[9, 10]

Regulating bodies and international health organizations have concluded the advertising of unhealthy foods/beverages impact children's eating habits and may be associated with the concurrent rise in childhood obesity;[11, 12] nevertheless governments in North America remain committed to industry self-regulation as the primary approach to reduce child-directed marketing of energy dense, low nutrient products, which, to date, has not been effective.[13, 14]

Several systematic and narrative reviews on the effects of child-oriented food and beverage promotion on diet, dietary determinants and health have been published.[11, 15-17]

However, these reviews have mostly reviewed observational or non-randomized experimental studies, and none have focused specifically on randomized controlled trials (RCTs). We aimed to systematically review all RCTs involving children aged 2 to 18 years that evaluated the impact of unhealthy food and beverage marketing compared to non-active control (e.g. TV programs or movies with toys or non-food advertising) on dietary intake and preference.

Methods

Search strategy

In January 2015, we searched MEDLINE, EMBASE and PsycINFO to identify studies published in English with the following criteria: (i) the population (children and adolescents 2-18 years of age); (ii) the intervention (unhealthy food or non-alcoholic beverages advertising delivered through TV/movie commercials, advergames [electronic games to advertise a product and might be played online or offline], or use of branded logos, packaging with licensed characters, or booklet/magazine advertisements); (iii) comparison (TV programs or movies with toys or non-dietary advertising, unbranded logos, plain packaging, watching regular TV programs or a movies without advertising); (iv) the outcomes (dietary intake or preference), and (v) methodology (randomized trials, according to Cochrane definition and criteria[18]). An a priori protocol for this study was not published. No substantive changes were made to the study design after inception. The search terms and strategies are available in **Appendix 3.1**. We also reviewed reference

lists and bibliographies of all included studies and related reviews for additional studies of relevance.

Study selection

Two reviewers independently screened the titles and/or abstracts of all identified studies and excluded those that were clearly not relevant. Subsequently, the full-text of the identified articles were collected and independently read to determine if they met our eligibility criteria. Discrepancies were resolved by consensus, or, if needed, by arbitration from a senior author. We used the eligibility criteria listed above. We excluded studies or study arm(s) that exclusively focused on healthy foods and beverages (fruits, vegetables) marketing. If in the article marketed foods/beverages were only named but not categorized as healthy or unhealthy, we used the WHO definition of 'unhealthy foods/beverages' as products high in energy, added fat, added sugar or, sodium.[19]

Data abstraction and risk of bias assessment

Data were extracted independently and in duplicate. We extracted the following data: (i) general study information (author's name, publication year, and study location), (ii) study population details (sample size, age, and ratio of males versus females), (iii) details on the intervention and comparison (e.g. marketing method including TV/movie advertisement, advergames, branded foods/beverages), duration of exposure to the marketed foods/beverages (eating opportunity and duration of advertising), test foods/beverages and type of foods/beverages provided for children to consume during or after the intervention (e.g. potato chips, candy, soda pop), and (iv) dietary intake in grams or kilo-

calories (kcals), and foods/beverages preference score or percentage of participants who selected specific foods/beverages.

Risk of bias was assessed using the Cochrane risk of bias instrument.[20] Among eligible studies, two reviewers independently assessed the following risk of bias issues: random sequence generation, allocation concealment, blinding of study participants, blinding of outcome assessors, incomplete outcome data, and other potential sources of bias. Studies were considered at high risk of bias when at least three items were assessed as high risk of bias.

GRADE principles were applied to independently assess the certainty (quality) of our pooled estimates using the following criteria: risk of bias, consistency, directness, imprecision, publication bias.[21] GRADE profiler software (version 3.6) was used to prepare the summary of finding table and to evaluate the quality of the evidence. Any discrepancies in data extraction, risk of bias or quality of evidence were resolved by consensus and a third researcher was consulted for advice when necessary.

Data synthesis and statistical methods

To compare the effects of unhealthy dietary marketing on dietary intake and dietary preference, three measures of effect were used: mean difference, standardized mean difference (SMD) and relative risk (RR). We calculated the mean difference and its corresponding 95% confidence intervals (CIs) for dietary intake, reported as grams or kcals of foods/beverages consumed during or after the experiments. To assess the dietary preferences, we calculated the SMD and its corresponding 95%CIs. Dietary preference

was reported as the percentage of children who preferred the experimental foods/beverages under study (all those included in the dietary preference measure). We treated this as a dichotomous variable (yes/no) and pooled eligible trials using the RR and the corresponding 95%CIs.

Heterogeneity was determined using the Q statistic and I^2. A significance level of $P < 0.10$ for Cochran's Q test or $I^2 > 40\%$ were considered as clinically important heterogeneity. [18, 22] We used the DerSimonian–Laird random-effects model for meta-analysis. Regardless of the observed statistical heterogeneity, we conducted the following subgroup analyses to explain any observed heterogeneity: age (8 years or less vs. > 8 years of age), assuming a larger dietary intake in older children, sex (boys vs. girls), assuming a larger intake in boys; type of foods/beverages provided for children (healthy vs. less healthy/unhealthy –foods/beverages high in fat, sugar, or salt), assuming a larger intake of less healthy/unhealthy foods/beverages; type of advertisement (TV advertisement vs. advergames vs. branded logos/packaging with licensed characters), assuming a larger intake of foods/beverages using TV advertisements. For any observed or theoretical heterogeneity in pooled estimates of dietary intake, we also considered two more probable explanations: duration of exposure to advertisements (5 minutes or less vs. > 5 minutes), assuming a larger intake in children with > 5 minutes exposure to advertisements; duration of exposure to experimental foods/beverages for consumption (eating opportunity) during and/or after advertisement (15 minutes or less vs. > 15 minutes), assuming an increased intake in children with > 15 minutes to consume the provided foods/beverages.

For subgroup analysis, we tested for interaction using a chi-square significance test.[23] For subgroups with more than 2 variables and seven observations, we performed meta-regression. If 10 or more studies were included in the meta-analysis, publication bias was examined by funnel plots and Begg's and Mazumdar's adjusted rank correlation test.[24] Data were analyzed in STATA software version 11.0, Texas, US.

Results

Description of included studies

Our literature search identified 2468 titles and/or abstracts, 108 full-texts were retrieved and screened. Of these, 79 studies were excluded after applying our eligibility criteria. The main reasons for exclusion included: 1) not being a randomized trial (n=34), 2) no child directed marketing (n=17), and 3) nonfood/beverage marketing such as advertising toys or cosmetics (n=11). The stages of evaluation and exclusion of the identified studies are presented in **Figure 3.1**.

We identified 29 eligible RCTs enrolling a total of 5,814 children. Of these, three studies were excluded from our meta-analyses due to insufficient outcome data (no data provided on number of children randomized, or no data on measures of variability) and different duration of intervention (repeated exposure to advertisements for more than a week). [25-27] The majority of studies were performed in North America (n=20). Studies most frequently examined the impact of TV advertising (12 studies), followed by licensed characters/logos (9 studies) and advergames (6 studies). Two studies looked at the effects of advertising in magazines/booklets. The median for the mean age of participants in the

included studies was 8.2 years (IQR = 5.6 to 9.5). The median sample size among included studies was 105 participants (IQR = 65 to 261). A detailed description of included studies is presented in **Table 3.1**.

We identified 17 studies eligible for meta-analysis on dietary preference and 9 for meta-analysis on dietary intake. The quality of reporting among the included studies was poor, with half of the included studies assessed as high risk of bias. Based on the full-text of included studies, the main reasons for assessing studies as high risk of bias included poor reporting with respect to allocation concealment and blinding of participants and data assessors. Four of the nine studies on dietary intake were rated as high risk of bias, while 9 of 17 studies reporting dietary preference were rated at high risk of bias (**Table 3.1**).

Effects of unhealthy food/beverage marketing on dietary intake

Dietary intake (kilocalories)

Of nine studies included in our meta-analysis on dietary intake, four studies reported dietary intake in grams[28-31] while five studies reported intake data in kcals.[32-36] We were able to convert food intake in grams to kcals of food intake in only one study.[28] Among the 6 studies (665 participants) providing data on dietary intake in kcals, the average time children were exposed to marketing was 3.8 minutes (median: 3.8 minutes) and the average time they were given to consume the food was 17.3 minutes (median: 17.5 minutes). Among the six included studies, the pooled estimate showed a significant increase of 30.4 kcals (95%CI: 2.9 to 57.9) favoring exposure to unhealthy dietary advertising versus non-dietary advertising (I^2 = 72.0%) (**Figure 3.2**).

The results for our seven subgroups are as follows. Among the six identified studies evaluating dietary caloric intake, in one study the intervention was TV advertisements, three used advergames, and in the remaining two studies, familiar licensed-characters or logos were used as the intervention. The pooled estimate of dietary intake reported as kcals was not significantly different among the three categories (**Table 3.2**). Our subgroup analysis for risk of bias revealed that the difference between studies at high risk of bias (n=3) and low risk (n=3) was significant (P=0.016) (MD=46.4 kcals, 95% CI: 11.0 to 81.7 and MD = -7.9kcals, 95%CI: -34.6 to 18.8, respectively), indicating that more methodologically sound studies found a stronger effect of advertising on caloric intake.

Children exposed to unhealthy dietary advertisements for > 5 minutes (n=222) had less caloric intake than those who were exposed ≤ 5 minutes (n =265) (MD=6.5 kcals, 95%CI: -25.8 to 38.8; I^2 = 77.0%; and MD=64.4 kcals, 95% CI: 39.8 to 89.0; I^2 = 0.0%). The test of interaction showed the difference between two estimates was statistically significant (P=0.005). Our subgroup analysis for duration of exposure (eating opportunity) to unhealthy foods/beverages showed that participants given < 15 minutes (2 studies) for eating/drinking had more caloric intake than those given ≥ 15 minutes to eat/drink (4 studies) (**Table 3.2**), and the difference between two estimates was significant (P=0.001).

In our subgroup analysis on type of foods/beverages provided, we found that that when children were exposed to unhealthy advertisements they consumed more unhealthy calories (n=487; MD=30.3 kcals, 95%CI: 7.8 to 52.9, I^2 = 82.1%) than healthy calories (n=236; MD = -2.7 kcals, 95%CI: -27.9 to 22.6; I^2 = 75.7%) and the difference between the two estimates was statistically significant (P = 0.051). With respect to baseline

characteristics, the mean difference of dietary intake as reported in kcals among boys (n=128) was 94.8 kcals (95% CI: 77.0 to 112.5; I^2 = 0.0%), while in girls (n=160) it was -8.8 kcals (95% CI: -77.6 to 60.1; I^2 = 60.1%) (**Table 3.2**), and the difference was significant (P=0.004). Results for our subgroup analysis on age were not significantly different (≤ 8 years MD=43.0 kcals; 95%CI: 1.4 to 84.7; >8 years MD=27.5 kcals; 95%CI: -7.8 to 62.7, P for test of interaction = 0.58).

Dietary intake (grams)

Among the four studies (395 participants) assessing dietary intake in grams, the average time children were exposed to the marketing was 6.9 minutes (median: 5.25 minutes) and average time they were given for eating was 19.3 minutes (median: 22.5 minutes). Our meta-analysis showed a significant increase of 4.8 grams (95%CI: 0.8-8.8) among those exposed to unhealthy dietary advertising (I^2 = 31.6%) (**Appendix 3.2**).

All four studies included in our pooled estimate for dietary intake as grams employed TV advertisements as the intervention. Results of subgroup analysis were similar based on risk of bias, duration of exposure to advertisements, and duration of exposure (eating opportunity) to unhealthy foods (**Table 3.2**). We had insufficient data to assess subgroups based on quality of calories (healthy vs. less healthy/unhealthy), sex or age.

For dietary intake reported as either kcals or grams there were too few studies to assess the risk of publication bias. The overall quality of evidence for dietary intake for both estimates was moderate. We rated the quality of evidence down from high to moderate due to indirect evidence (dietary intake is a surrogate for more patient-important

outcomes such as weight-gain and obesity). Details of the overall quality of evidence are summarized in the GRADE summary of findings **Table 3.5**.

Effects of unhealthy dietary marketing on dietary preference

Dietary preference scores

Of the 17 included studies on dietary preference, 12 trials reported a food or taste preference score. Our meta-analysis showed a small non-significant increased effect favoring preference for unhealthy foods/beverages when accompanied by advertising (SMD=0.23, 95%CI: -0.04 to 0.5; I^2 = 87.6%) (**Appendix 3.3**). Results of the subgroup analysis showed that dietary preference was not influenced by type of advertisement, risk of bias, and type of foods/beverages provided to children (**Table 3.3**). The mean age of participants in the 8 RCTs (879 children) was ≤8 years and their preference for unhealthy foods/beverages showed a small to moderate effect size (SMD=0.46; 95%CI: 0.21 to 0.72; I^2 = 72.7%), whereas in the 4 RCTs (n=1174) including participants >8 years their dietary preference for unhealthy foods/beverages showed a small non-significant effect size (SMD=-0.28; 95%CI: -0.72 to 0.16; I^2 = 19.5%). The test for interaction was significant (z = 2.85, P=0.004).

Food preference percentage

Of the 17 included studies on foods/beverages preference, 8 trials reported the percentage of children who preferred specific foods/beverages. Children exposed to unhealthy foods/beverages marketing had a higher risk of selecting the advertised products that were associated with a familiar licensed-character/logo (RR = 1.1, 95%CI: 1.0 to 1.2; P=0.052,

I^2 = 27.6%) (**Appendix 3.4**). Subgroup analysis based on types of advertising demonstrated no significant difference, whereas studies with higher risk of bias and studies performed on children less than 8 years of age showed significantly increased risk of selecting the advertised products. However, the test of interaction for all three subgroups was non-significant (**Table 4.3**). We had insufficient data to assess subgroups based on advertisement time, type of food (unhealthy versus healthy) and sex.

The funnel plot and the Begg's and Mazumdar's adjusted rank correlation test for 12 studies reporting dietary preference scores did not indicate evidence of publication bias (**Appendix 3.5**). We did not test for publication bias among studies that reported dietary preference as a percentage as only eight studies were included. The overall quality of evidence for dietary preference scores was low. We rated the quality of evidence down due to risk of bias and unexplained heterogeneity. The overall quality of evidence for dietary preference reported as a percentage was moderate. We rated down from high to moderate based on risk of bias issues (**Table 3.5**)

Discussion

We identified 29 randomized trials evaluating the effects of unhealthy food and beverage marketing involving almost 6000 children aged 2-18 years. We found that exposure to unhealthy food and beverage marketing increased children's dietary intake and influenced children's dietary behaviors during or shortly after exposure to advertisements. Our findings were consistent across studies. That is, in 18 of 26 studies amenable for meta-

analysis, the mean dietary intake or preference was greater for the marketed dietary products than non-marketed products.

Using GRADE methodology, the overall quality of evidence for food intake in kcals (665 children) and food intake in grams (395 children) was moderate, meaning the true effect is likely to be close to the estimate of the effect but there is a possibility that it is different. Considering the short average time children were exposed to the adverts (approximately 5 minutes), and the nearly 30 kcals (4.5 grams) increase in dietary intake over an average of 15 minutes, an association between exposure to energy-dense, low nutrition food and beverage advertising and weight gain, obesity and other dietary related non-communicable diseases is plausible. Although results were non-significant with respect to food and beverage preferences, among 1648 children exposed to energy-dense, low-nutrient products marketing, we found an increased risk of selecting advertised foods or beverages that were associated with a familiar licensed-character or logos (moderate quality evidence). Similarly, the food and beverage preference score among 2053 children showed a non-significant increased risk (low quality evidence).

Our findings suggest that younger children (≤ 8 years of age) might be more susceptible to the impact of food and beverage marketing in terms of quantity and quality of calories consumed. In the subgroup analyses, the most consistent finding suggested that younger children have increased caloric intake, preference scores and often selected unhealthy foods and beverages as compared to older children. However, only preference scores were significant, demonstrating that those ≤ 8 years of age had higher preference scores than those > 8 years. While children at the age of two or three are able to recognize

familiar characters and identify food and beverage products, they are less able to understand the intention behind advertising and differentiate between program content and advertisements until the age of seven or eight. .[10,41] Thus, younger children might be more vulnerable to the influence of advertisements and associate the marketed products with positive features of commercials and subsequently try to imitate the behaviors they see.

Although we were only able to conduct a subgroup analysis based on sex for one of our four outcomes, our findings further suggest that boys might be more susceptible to the impact of food and beverage marketing in terms of caloric intake. Girls may have a higher tendency towards dieting practices possibly as result of maternal encouragements to be thin [35,36] that may have suppressed their natural response.[37] It has also been suggested that boys may be more vulnerable when exposed to external cues for food and beverage advertisements and therefore may consume more than girls.[37,38] Another explanation for the observed difference between boys and girls might be that child and adolescent targeted food and beverage advertisements tend to focus on boys perhaps because they are more susceptible to external cues of food advertisements.[39,40]

It is important to note that advergames differ from TV advertising in several key ways (active vs. passive reception, low vs. high interactivity while exposed to the brand, exposure time).[37, 38] In comparing the subgroups (TV advertisement, advergames, and using familiar characters/logo) our analysis showed no significant difference in children's dietary intake or preference among different types of marketing. This might be due to the small number of included studies. In addition, none of the identified trials directly

compared the effects of advergames with TV advertising in terms of dietary intake or preference.

Subgroup analysis of included studies according to the time children were exposed to the advertisements (≤ 5 minutes vs. > 5 minutes) and the time they were given to eat (< 15 minutes vs. ≥ 15 minutes) showed that those exposed to less marketing and those who had less time to consume, had higher intakes. These findings were counter-intuitive; however, studies that exposed children > 5 minutes of advertisements or provided ≥ 15 minutes to consume tended to have higher risk of bias and were more likely to provide more energy dense foods.[28, 34-36] These findings may also be due to chance given the sparse number of studies included in the analysis, or the fact that children may have gorged the energy dense snacks at the beginning of each study given that they had limited time.

Four systematic reviews have investigated the effects of food and beverage marketing to children, three of these being technical reports from authoritative bodies such as World Health Organization (WHO).[14-16,28] While largely based on evidence from observational studies, each review concluded that the marketing and promotion of foods and beverages high in fat, sugar and/or salt have a negative impact on children's nutrition preferences, purchase behavior, consumption patterns and diet-related health. A recent meta-analysis showed acute exposure to food and beverage advertising is associated with greater food intake in children;[39] however, they combined randomized and non-randomized trials and did not assess risk of bias or the quality in evidence using the GRADE approach. Further, Boyland at al 2016 included only 13 studies in their meta-analysis, while we included 26 RCTs. While not conclusive, the findings from this review contribute to the growing body

of research suggesting that the marketing of energy-dense, low nutrition foods and beverages to children contribute to unhealthy dietary choices, which puts children at risk for diet-related diseases later in life.

This paper has a number of noted limitations. First, using the GRADE approach the overall quality of evidence for the effects of food and beverage advertising on dietary intake and preference was low to moderate quality. The quality of evidence was impacted primarily because of lack of reporting of allocation concealment, blinding of outcome assessors and participants, and the unavailability of study protocols were substantial among included articles, limiting the overall certainty in evidence. Second, the included studies examined responses to acute advertising exposure only and the collective effects of continued exposure to food and beverage marketing that occurs in real life and over a lifetime may differ. Third, the designs of these interventions (being conducted in laboratory setting rather than real life situations) may be different from the typical daily exposure to advertising children are subjected to.

Implications for public health policy

A recent global study spanning 13 countries revealed that children are exposed to an average of five food advertisements per hour with unhealthy ‘non-core’ foods accounting for greater than 80% of all televised food advertisements in Canada, the United States and Germany.[40] Collectively, the evidence linking children’s exposure to unhealthy food and beverage marketing to poor dietary behaviors and increased risk of overweight and obesity has sparked global debate. Results of a recent modeling study suggested that a

ban on television advertising of foods high in fat, sugar and/or salt could reduce overweight and obesity in childhood by 18% and 2.5%, respectively.[41, 42] Given the potential impact on children's health, in 2010 the WHO released a set of recommendations urging member states to restrict the marketing of foods and beverages high in saturated fats, trans-fats, added sugar and salt to children.[12] Voluntary self-monitoring by industry and inadequate nutritional standards for defining healthy/unhealthy dietary products, and the lack of government monitoring and oversight remain key flaws to recent initiatives and likely account for the lack of reduction in child-targeted marketing for unhealthy foods and beverages.[13, 43-45]

Conclusions

The evidence indicates that unhealthy food and beverage marketing increases dietary intake and preference for energy-dense, low nutrition products in children during or shortly after exposure to advertisements. Further research is needed to evaluate the impact of unhealthy food and beverage advertising on daily and weekly dietary intake and choices. Overall, our analyses support the need for a review of public policy on child targeted unhealthy food and beverage marketing.

References

1. Ng M, Fleming T, Robinson M, *et al.* Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 2014; 384: 766-81.
2. Lobstein T, Jackson-Leach R, Moodie ML, *et al.* Child and adolescent obesity: part of a bigger picture. *Lancet* 2015; 385: 2510-20.
3. Wang Y, Lobstein T. Worldwide trends in childhood overweight and obesity. *Int J Pediatr Obes* 2006; 1: 11-25.
4. Weiss R, Kaufman FR. Metabolic Complications of Childhood Obesity Identifying and mitigating the risk. *Diabetes Care* 2008; 31: S310-S16.
5. Drewnowski A, Darmon N. The economics of obesity: dietary energy density and energy cost. *Soc Sci Med* 2005; 82: 265S-73S.
6. Harris JL, Bargh JA. Television viewing and unhealthy diet: implications for children and media interventions. *Health Commun* 2009; 24: 660-73.
7. Cecchini M, Sassi F, Lauer JA, Lee YY, Guajardo-Barron V, Chisholm D. Tackling of unhealthy diets, physical inactivity, and obesity: health effects and cost-effectiveness. *Lancet* 2010; 376: 1775-84.
8. U.S. Burden of Disease Collaborators. The state of us health, 1990-2010: Burden of diseases, injuries, and risk factors. *JAMA* 2013; 310: 591-606.
9. Boyland EJ, Halford JCG. Television advertising and branding. Effects on eating behaviour and food preferences in children. *Appetite* 2013; 62: 236-41.
10. Linn S, Novosat CL. Calories for sale: food marketing to children in the twenty-first century. *Ann Am Acad Pol Soc Sci* 2008; 615: 133-55.
11. Cairns G, Angus K, Hastings G. The extent, nature and effects of food promotion to children A review of the evidence to December 2008. www.who.int/dietphysicalactivity/Evidence_Update_2009.pdf: World Health Organization, Geneva 2009.
12. World Health Organization. Set of recommendations on the marketing of foods and non-alcoholic beverages to children. Geneva: WHO 2010.
13. Potvin Kent M, Dubois L, Wanless A. Self-regulation by industry of food marketing is having little impact during children's preferred television. *Int J Pediatr Obes* 2011; 6: 401-08.
14. Powell LM, Schermbeck RM, Szczypka G, Chaloupka FJ, Braunschweig CL. Trends in the nutritional content of television food advertisements seen by children in the united states: Analyses by age, food categories, and companies. *Arch Pediatr Adolesc Med* 2011; 165: 1078-86.

15. Cairns G, Angus K, Hastings G, Caraher M. Systematic reviews of the evidence on the nature, extent and effects of food marketing to children. A retrospective summary. *Appetite* 2013; 62: 209-15.

16. Hastings G, McDermott L, Angus K, Stead M, Thomson S. The extent, nature and effects of food promotion to children. A review of the evidence technical paper prepared for the world health organization. http://www.whqlibdoc.who.int/publications/2007/9789241595247_eng.pdf: World Health Organization, Geneva 2006.

17. Jenkin G, Madhvani N, Signal L, Bowers S. A systematic review of persuasive marketing techniques to promote food to children on television. *Obes Rev* 2014; 15: 281-93.

18. Higgins JP, Green S. Cochrane Handbook for Systematic Reviews of Interventions, version [5.1.0] (updated March 2011). The Cochrane Collaboration 2011.

19. World Health Organization. Healthy diet. Fact sheet No. 394. http://www.who.int/nutrition/publications/nutrientrequirements/healthydiet_factsheet394.pdf : Updated May 2015.

20. Higgins JPT, Altman DG, Gøtzsche PC, *et al.* The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011; 343: d5928.

21. Guyatt GH, Oxman AD, Vist GE, *et al.* GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008; 336: 924-26.

22. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; 327: 557-60.

23. Altman DG, Bland JM. Interaction revisited: the difference between two estimates. *BMJ* 2003; 326: 219.

24. Sterne JA, Sutton AJ, Ioannidis J, *et al.* Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* 2011; 343: d4002.

25. Chernin A. The effects of food marketing on children's preferences: testing the moderating roles of age and gender. *Ann Am Acad Pol Soc Sci* 2008; 615: 102-18.

26. Galst JP. Television food commercials and pro-nutritional public service announcements as determinants of young children's snack choices. *Child Dev* 1980; 51: 935-38.

27. Gorn GJ, Goldberg ME. Behavioral evidence of the effects of televised food messages on children. *J Consum Res* 1982; 9: 200-05.

28. Anschutz DJ, Engels RC, Van Strien T. Side effects of television food commercials on concurrent nonadvertised sweet snack food intakes in young children. *Am J Clin Nutr* 2009; 89: 1328-33.

29. Anschutz DJ, Engels RCME, Van Strien T. Maternal encouragement to be thin moderates the effect of commercials on children's snack food intake. *Appetite* 2010; 55: 117-23.

30. Gorn GJ, Goldberg ME. Children's responses to repetitive television commercials. *J Consum Res* 1980; 6: 421-24.

31. Harris JL, Bargh JA, Brownell KD. Priming effects of television food advertising on eating behavior. *Health Psychol* 2009; 28: 404-13.

32. Folkvord F, Anschütz D, Buijzen M, Valkenburg P. The effect of playing advergames that promote energy-dense snacks or fruit on actual food intake among children. *Am J Clin Nutr* 2013; 97: 239-45.

33. Folkvord F, Anschütz DJ, Nederkoorn C, Westerik H, Buijzen M. Impulsivity, "Advergames," and Food Intake. *Pediatrics* 2014; 133: 1007-12.

34. Forman J, Halford JC, Summe H, MacDougall M, Keller KL. Food branding influences ad libitum intake differently in children depending on weight status. Results of a pilot study. *Appetite* 2009; 53: 76-83.

35. Harris JL, Speers SE, Schwartz MB, Brownell KD. Us food company branded advergames on the internet: Children's exposure and effects on snack consumption. *J Child Media* 2012; 6: 51-68.

36. Keller KL, Kuilema LG, Lee N, *et al.* The impact of food branding on children's eating behavior and obesity. *Physiol Behav* 2012; 106: 379-86.

37. Waiguny MK, Nelson MR, Terlutter R. The Relationship of Persuasion Knowledge, Identification of Commercial Intent and Persuasion Outcomes in Advergames—the Role of Media Context and Presence. *J Consum Policy* 2014; 37: 257-77.

38. Bellman S, Kemp A, Haddad H, Varan D. The effectiveness of advergames compared to television commercials and interactive commercials featuring advergames. *Comput Human Behav* 2014; 32: 276-83.

39. Boyland EJ, Nolan S, Kelly B, *et al.* Advertising as a cue to consume: a systematic review and meta-analysis of the effects of acute exposure to unhealthy food and nonalcoholic beverage advertising on intake in children and adults. *Am J Clin Nutr* 2016; 103: 519-33.

40. Kelly B, Halford JCG, Boyland EJ, *et al.* Television food advertising to children: a global perspective. *Am J Public Health* 2010; 100: 1730-36.

41. Chou SY, Rashad I, Grossman M. Fast-food restaurant advertising on television and its influence on childhood obesity. *J Law Econ* 2008; 51: 599-618.

42. Veerman JL, Van Beeck EF, Barendregt JJ, Mackenbach JP. By how much would limiting TV food advertising reduce childhood obesity? *Eur J Public Health* 2009; 19: 365–69.

43. Potvin Kent M, Dubois L, Wanless A. A nutritional comparison of foods and beverages marketed to children in two advertising policy environments. *Obesity* 2012; 20: 1829-37.

44. Potvin Kent M, Wanless A. The influence of the Children's Food and Beverage Advertising Initiative: change in children's exposure to food advertising on television in Canada between 2006-2009. *Int J Obes* 2014; 38: 558-62.

45. Powell LM, Harris JL, Fox T. Food marketing expenditures aimed at youth: putting the numbers in context. *Am J Prev Med* 2013; 45: 453-61.

46. Borzekowski DLG, Robinson TN. The 30-second effect: an experiment revealing the impact of television commercials on food preferences of preschoolers. *J Am Diet Assoc* 2001; 101: 42-46.

47. Dawson BL, Jeffrey DB, Walsh JA. Television food commericals' effect on children's resistance to temptation. *J Appl Soc Psychol* 1988; 18: 1353-60.

48. de Droog SM, Valkenburg PM, Buijzen M. Using brand characters to promote young children's liking of and purchase requests for fruit. *J Health Commun* 2010; 16: 79-89.

49. Dixon HG, Scully ML, Wakefield MA, White VM, Crawford DA. The effects of television advertisements for junk food versus nutritious food on children's food attitudes and preferences. *Soc Sci Med* 2007; 65: 1311-23.

50. Dixon H, Scully M, Niven P, *et al.* Effects of nutrient content claims, sports celebrity endorsements and premium offers on pre-adolescent children's food preferences: Experimental research. *Pediatr Obes* 2014; 9: e47-e57.

51. Elliott CD, Den Hoed RC, Conlon MJ. Food branding and young children's taste preferences: a reassessment. *Can J Public Health* 2013; 104: e364-e68.

52. Goldberg ME, Gorn GJ, Gibson W. TV messages for snack and breakfast foods: Do they influence children's preferences? *J Consum Res* 1978; 5: 73-81.

53. Jones SC, Kervin L. An experimental study on the effects of exposure to magazine advertising on children's food choices. *Public Health Nutr* 2011; 14: 1337-44.

54. King L, Hill AJ. Magazine adverts for healthy and less healthy foods: Effects on recall but not hunger or food choice by pre-adolescent children. *Appetite* 2008; 51: 194-97.

55. Kotler JA, Schiffman JM, Hanson KG. The influence of media characters on children's food choices. *J Health Commun* 2012; 17: 886-98.

56. Lapierre MA, Vaala SE, Linebarger DL. Influence of licensed spokescharacters and health cues on children's ratings of cereal taste. *Arch Pediatr Adolesc Med* 2011; 165: 229-34.

57. Mallinckrodt V, Mizerski D. The effects of playing an advergame on young children's perceptions, preferences, and requests. *J Advert* 2007; 36: 87-100.

58. Pempek TA, Calvert SL. Tipping the balance: use of advergames to promote consumption of nutritious foods and beverages by low-income African American children. *Arch Pediatr Adolesc Med* 2009; 163: 633-37.

59. Rifon NJ, Taylor Quilliam E, Paek HJ, Weatherspoon LJ, Kim SK, Smreker KC. Age-dependent effects of food advergame brand integration and interactivity. *Int J Advert* 2014; 33: 475-508.

60. Roberto CA, Baik J, Harris JL, Brownell KD. Influence of licensed characters on children's taste and snack preferences. *Pediatrics* 2010; 126: 88-93.

61. Robinson TN, Borzekowski DL, Matheson DM, Kraemer HC. Effects of fast food branding on young children's taste preferences. *Arch Pediatr Adolesc Med* 2007; 161: 792-97.

62. Toomey DA, Francis AL. Branded product placement and pre-teenaged consumers: influence on brand preference and choice. *Young Consumers* 2013; 14: 180-92.

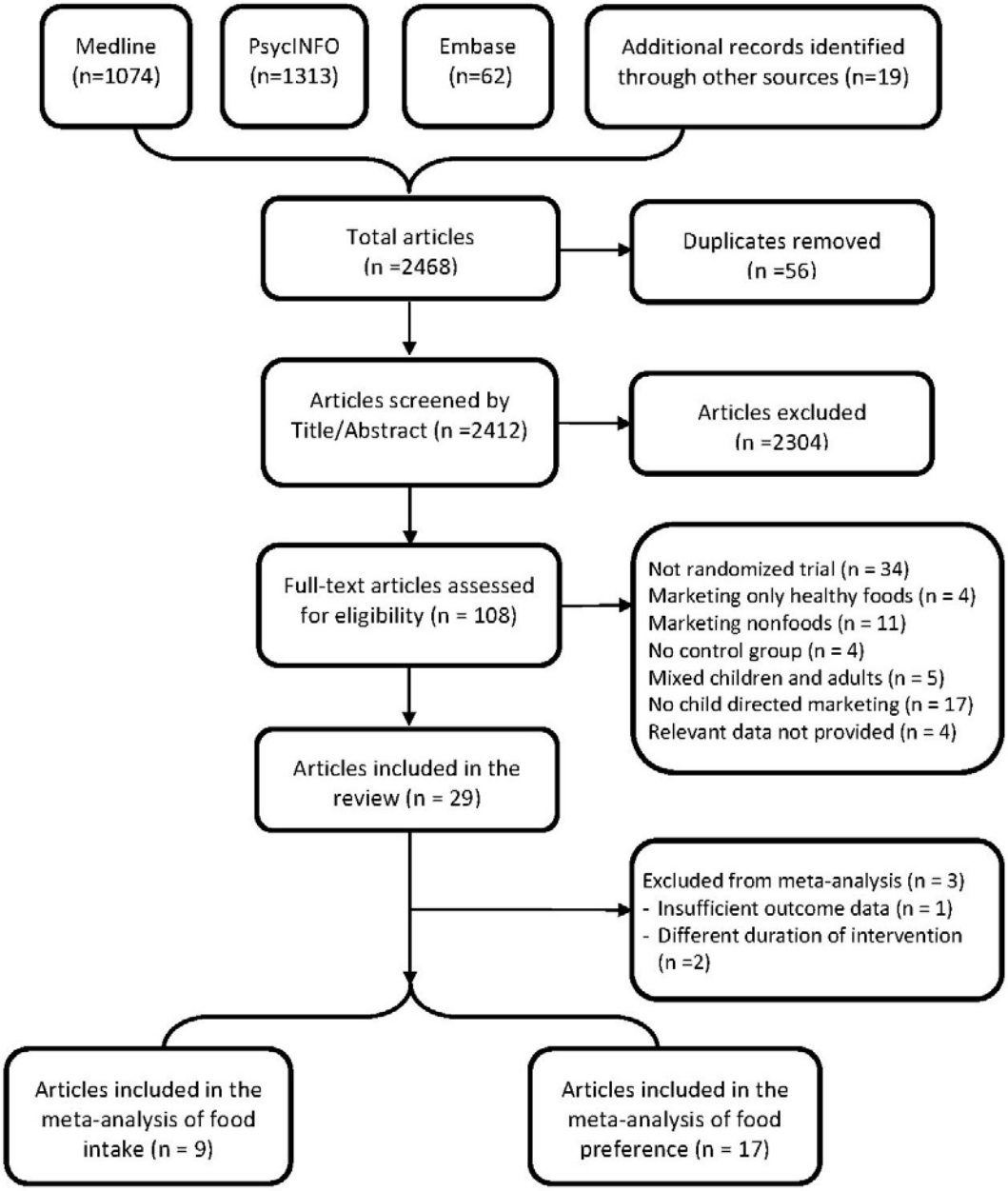


Figure 3. 1: Flow diagram of database searches and articles included in the systematic review and meta-analysis

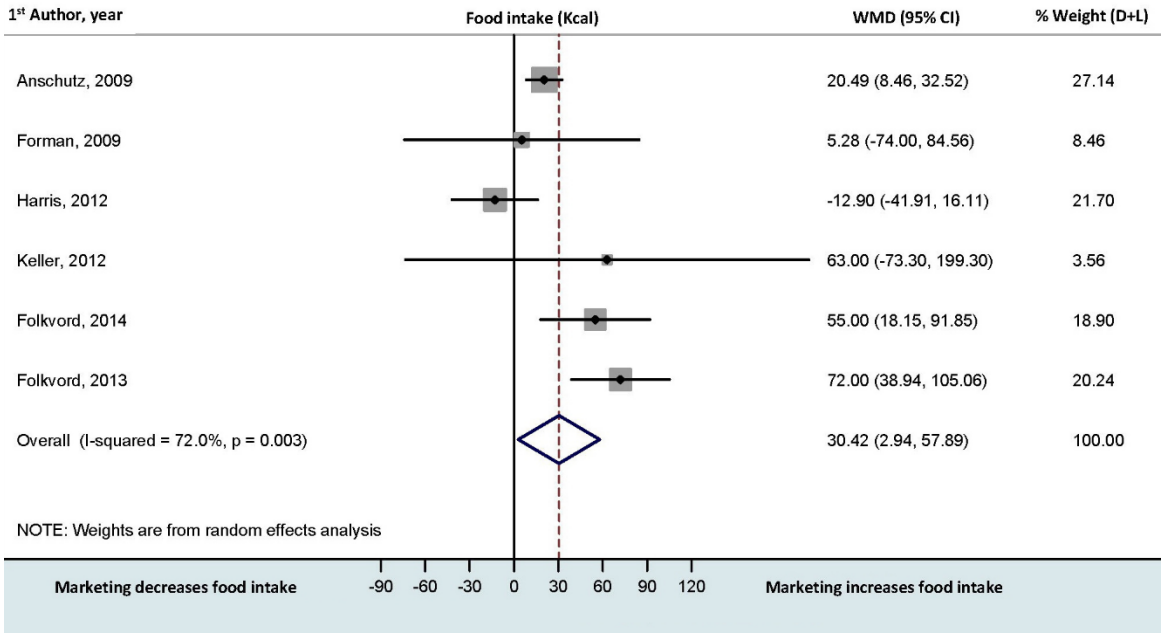


Figure 3. 2: Forest plot showing the weighted mean difference in food intake (kcal) between unhealthy food and nonfood marketing groups.

Horizontal bars denote 95% CIs. Studies are represented as squares centered on the point estimate of the result of each study. The area of the square represents the weight given to the study in the meta-analysis by STATA software. The pooled mean difference was calculated by a random-effects model. The diamond represents the overall estimated effect and its 95% CIs in total (center line of diamond, dashed line). The solid vertical line is the line of no effect.

Table 3. 1: characteristics of studies included in the systematic review

Author (date) ref # Country	N †	Mean age [range]	% male	Intervention	Comparison(s)	Test food/beverage	Outcome & results	Risk of bias
Anschutz (2009)[28] Netherlands	120	9.8 [NR]	46.7	20min movie with three food and two neutral commercials for 5min	Same movie as intervention with five neutral commercials	Freely eat from a pre-weighed bowl containing M&M chocolate-coated peanuts during the movie	significant interaction between commercial type and sex of the child	Low
Anschutz (2010)[29] Netherlands	120	9.6 [8-12]	-	20min movie with four food and one neutral commercials for 5min	Same movie as intervention with five neutral commercials	Freely eat from a pre-weighed bowl containing M&M chocolate-coated peanuts during the movie	No significant effect on food intake	Low
Borzekowski (2001)[46] USA	39	4.0 [2-6]	52.0	26min videotape of two animated children programs with seven food commercials mixed with one neutral commercial for about 4min	the same animated videotape as intervention with no commercial	choose foods similar to the advertised item	Significant preference toward advertised items	High
Chernin (2008)[25] ‡ USA	133	8.2 [5-11]	39.8	13min segment of an animated children program with 1 food commercial	same program with one food commercial (different from intervention item)	choose advertised food product among three alternatives in the same product category	Significant preference toward advertised items	Low
Dawson (1988)[47] USA	80	6.2 [NR]	-	two 30-sec low-nutritional TV commercials with one repetition	two 30-sec toy commercials played for two times	self-report desire to transgress using VAS§	No significant effect on temptation to transgress toward low-nutrient foods	High
de Droog (2010)[48] Netherlands	210	NR [4-6]	50.0	food with a familiar (cartoon) character on the package	food with a unfamiliar character /no character on the package	purchase request intent for banana candy using VAS§	No significant effect on purchase request intent	Low
Dixon (2007)[49] Australia	919	10.8 [NR]	-	four 30-sec junk food commercials mixed with four 30-sec neutral commercials played twice during a video program	same video as intervention with eight 30-sec neutral commercials	intention to eat four junk foods using a 5 point Likert scale	No significant effect on intention to eat	Low

Author (date) ref # Country	N †	Mean age [range]	% male	Intervention	Comparison(s)	Test food/beverage	Outcome & results	Risk of bias
Dixon (2014)[50] Australia	1302	11.0 [10-12]	48.7	Foods with sports celebrity endorsements on packages	Foods with nutrient content claims or no promotion on packages	Percentage of children choosing EDNP products and rating the likelihood of asking to buy	Significant lower rating in likelihood of asking to buy (all) and selecting control products (only boys)	Low
Elliot (2013)[51] Canada	65	3.8 [3-5]	44.6	Food with the logo of a popular fast food restaurant packaging	Food in plain, colorful, or Starbucks wrapping	Taste preference score range from -1 to +1	Significant taste preference of branded over plain packaging	Low
Folkvord (2013)[32] Netherlands	270	8.9 [8-10]	51.5	5min of memory game promoting energy-dense snacks on cards	Same game as intervention promoting toys or no game	Freely eat from two pre-weighed bowls of energy-dense snacks and two bowls of fruits	Children who played advergame promoting food ate significantly more than control and toy advergame	Low
Folkvord (2014)[33] Netherlands	261	7.7 [7-10]	50.2	5min of online memory game promoting energy-dense snacks on cards	Same game as intervention promoting toys	Freely eat from pre-weighed bowls of snacks	Significant effect of advergame promoting food on caloric intake	Low
Forman (2009)[34] USA	43	5.9 [4-6]	39.8	In two visits children were exposed to branded foods (McDonalds, Coca Cola, Trix)	In two visits children were exposed to unbranded foods	Eat ad libitum for 30min from their respective dinner	No difference in intake of the branded vs. unbranded food conditions	High
Galst (1980)[26‡] USA	65	NR [3-6]	55.4	Groups of children watched two different short cartoons each day for four weeks with nine 30-seconds commercials for food products	Same cartoons as intervention without commercials	Children were allowed to select a daily snack containing added sugar or no added sugar	Significant lower request for sugared snack in control. No difference in request for no added sugar snacks between groups	High
Goldberg (1978)[52] USA	80	NR [4-6]	-	Groups of children were exposed to a minimum of 4.5min of sugared snack commercials embedded in a 24min cartoon program	Children not exposed to any program	Preference was assessed based on selection of the snack foods on a series of boards	Significant more sugared food was selected by children in intervention group than controls	High
Gorn (1980)[30] Canada	77	NR [8-10]	100	Five 30-seconds ice cream commercials embedded in a 30min cartoon	Same cartoons as intervention without commercials	Eat ad libitum for 15min from his choice of ice cream	No significant difference in the consumption of ice cream between groups	High

Author (date) ref # Country	N †	Mean age [range]	% male	Intervention	Comparison(s)	Test food/beverage	Outcome & results	Risk of bias
Gorn (1982)[27‡] Canada	288	NR [5-8]	-	14 different 30-minute shows with 4.5min candy commercials were played for two weeks during a summer camp	Same procedure as intervention with 4.5min public announcements or no commercial	Children were allowed to choose one of two beverages and two of four food choices (fruits and candy bars) each afternoon	Those in candy commercial condition picked significantly less healthy foods and beverages	Low
Harris (2009)[31] USA	118	8.8 [7-11]	52.5	14min of an animated children program with four 30-seconds food commercials	Same program as intervention with four 30seconds non-food commercials	Freely eat from a pre-weighed bowl containing goldfish crackers during the program	Significant more crackers were eaten by those who watched food commercials	Low
Harris (2012)[35] USA	149	9.4 [7-12]	52.6	12min playing with an online game featuring foods	12min playing with a nonfood online game	Freely eat from a pre-weighed snack bowls for 20min	Significant effect of advergame promoting healthy/unhealthy food on food consumption	High
Jones (2011)[53] Australia	47	8.7 [5-12]	-	15min to read the magazine with food advertisements	magazine with no food advertisements	given two vouchers for their choice of snack foods from items advertised/not advertised brands	Participants were equally likely to select a healthy but not an unhealthy food item	Low
Keller (2012)[36] USA	41	8.4 [7-9]	51.2	A multi-item test-meal that was branded with the logo of a popular fast food restaurant	Same test-meal as intervention served in plain white packaging	Eat ad libitum for 30min from the test foods	Non-significant increase in branded food intake	High
King (2008)[54] UK	309	9.7 [9-10]	51.1	Children received a booklet with food adverts as a media literacy exercise	Same booklet as intervention with non-food adverts	After the intervention children exchanged food choice coupons for raisins or confectionery	No significant effect of advert group on food choice	Low
Kotler (2012a)[55] USA	343	4.1 [2-6]	-	children were given food pairs with a familiar and an unknown character on the first and the second of each of the nine pairs	Children were given food pairs with no character stickers associated with the foods	They were asked to pick one food from each of nine pictured pairs they would like to eat	Significant effect of character on food preference	High

Author (date) ref # Country	N †	Mean age [range]	% male	Intervention	Comparison(s)	Test food/beverage	Outcome & results	Risk of bias
Kotler (2012b)[55] USA	207	NR [3-6]	-	A familiar and an unknown character were placed in front of bowls with small pieces of food	no character stickers were associated with the food bowls	Choice of foods children ate were recorded	Significant more foods associated with a familiar character were eaten	High
Lappier (2011)[56] USA	80	5.6 [4-6]	45.0	children were given a small cup dry serving of a cereal with a familiar character on the box	Cereal box with no character on the box	Five-point rating scale was used for taste preference	significant effect for character presence on children tastes	Low
Mallinckrodt (2007)[57] Australia	294	NR [5-8]	40.0	5min playing with an advergame featuring foods	not exposed to the advergame	Preference for advertised cereal over other cereal options	Significant effect of advertisement on children's preference	High
Pempek (2009)[58] USA	30	9.5 [9-10]	46.7	5min playing an advergame before selection of a snack and beverage	5min playing an advergame after snack and beverage selection	A summary score ranged from 0-2 for selection of a snack and beverages same advertised items	Significant effect of advertisement on preference score	High
Rifon (2014)[59] USA	92	7.3 [5-10]	43.8	Playing an advergame with a branded cereal box as the game token	Same advergame as intervention with unbranded cereal box	Attitude towards the brand was measured using two items on a 5-point scale	No significant effect of playing branded advergame on brand attitude	Low
Roberto (2010)[60] USA	40	5.0 [4-6]	65.0	Three pairs of identical foods presented in packages with a popular cartoon character	Same pairs of foods as ntervention in packages with no character	Children selected which food items they prefer to eat for a snack, and which taste better	Significant effect of licensed-character on both preference measures	High
Robinson (2007)[61] USA	63	4.6 [3-6]	47.6	Pairs of identical foods presented in packages with the logo of a popular fast food restaurant	Same pairs of foods as intervention in plain packaging	Taste preference score range from -1 to +1	Significant taste preference of branded over plain packaging	High
Toomey (2013)[62] USA	69	9.8 [8-12]	-	Product placement was implemented using a soft drink brand within a 4min video	Same video as intervention with an unbranded soft drink	Preference and choice were assessed two weeks after the experiment	No significant effect on preference of branded foods	High

† Trial sample size (number randomized)

‡ Excluded from meta-analysis

§ Visual analogue scale

Table 3. 2: Results of the meta-analysis and subgroup analysis of randomized trials investigating the effect of unhealthy food/beverage marketing on dietary intake

		No. of trials	Mean difference	95% CI		No. of participants		P value for difference	I^2	P value for interaction
				Lower	Upper	Control	Intervention			
Dietary intake (kcals)	TV advertisement	1	20.5	8.5	32.5	57	63	0.001	-	-
	Advergame	3	37.4	-16.8	91.6	177	199	0.176	87.7	
	Logo/Brand	2	19.9	-48.7	88.4	89	89	0.570	0.0	
	Low risk of bias	3	46.4	11.0	81.7	184	201	0.010	80.6	0.016
	High risk of bias	3	-7.9	-34.6	18.8	139	141	0.561	0.0	
	Advertisement time ≤ 5min§	2	64.4	39.8	89.0	127	138	<0.001	0.0	0.005
	Advertisement time > 5min§	2	6.5	-25.8	38.8	107	115	0.693	77.0	
	Consumption time < 15min¥	2	64.4	39.8	89.0	127	138	<0.001	0.0	0.001
	Consumption time ≥ 15min¥	4	9.6	-13.4	32.6	196	204	0.413	38.4	
	Healthy	2	-2.7	-27.9	22.6	115	121	0.837	75.7	0.051
	Unhealthy/less healthy	4	30.3	7.8	52.9	234	253	0.008	82.1	
	Boys	3	94.8	77.0	112.5	64	64	<0.001	0.0	0.004
	Girls	3	-8.8	-77.6	60.1	77	83	0.803	60.1	
	≤ 8 years of age†	2	43.0	1.38	84.7	110	117	0.043	79.7	0.578
	> 8 years of age†	4	27.5	-7.8	62.7	213	225	0.127	19.5	
	Total	6	30.4	2.9	57.9	323	342	0.030	72.0	-
Dietary intake (grams)‡	Low risk of bias	3	4.9	0.3	9.5	156	162	0.036	50.7	0.552
	High risk of bias	1	-4.1	-33.4	25.3	40	37	0.785	-	
	Advertisement time ≤ 5min	3	3.6	-6.5	13.6	139	136	0.485	42.0	0.936
	Advertisement time > 5min	1	4.0	1.7	6.4	57	63	0.001	-	
	Consumption time < 15min	1	8.8	3.4	14.3	59	59	0.002	-	0.099
	Consumption time ≥ 15min	3	3.8	1.4	6.1	137	140	0.002	0.0	
	Total	4	4.8	0.8	8.8	196	199	0.018	31.6	-

† Based on the mean age reported in the trial

‡ All trials in this category used TV advertisements as intervention, mean age in all of them was more than 8 years and none reported the intake of healthy vs. unhealthy products.

§ Time participants were exposed to unhealthy food/beverage marketing

¥ The time given to the participants for eating the food/beverage provided by researchers during or after the intervention.

Table 3. 3: Results of the meta-analysis and subgroup analysis of studies investigating the effect of unhealthy food/beverage marketing on dietary preference (preference score)

	Food/beverage preference								
	No. of trials	SMD	95% CI		No. of participants		P value	P value for interaction	I^2
			Lower	Upper	Control	Intervention			
TV advertisement	3	0.29	-0.28	0.86	262	268	0.313	0.772	72.1
Advergame	2	-1.20	-4.44	2.04	56	56	0.467		95.6
Logo/Brand	7	0.34	0.01	0.67	706	705	0.050		89.6
Low risk of bias	7	0.11	-0.10	0.32	904	898	0.315	0.743	77.5
High risk of bias	5	0.23	-0.46	0.91	120	131	0.518		86.8
Healthy	4	-0.01	-0.57	0.56	498	493	0.982	0.071	92.3
Unhealthy/less healthy	6	0.74	0.16	1.33	551	559	0.013		95.1
≤ 8 years of age	8	0.46	0.21	0.72	433	446	0.001	0.004	72.7
> 8 years of age	4	-0.28	-0.72	0.16	591	583	0.212		87.7
Total	12	0.23	-0.04	0.50	1024	1029	0.094	-	87.6

Table 3. 4: Results of the meta-analysis and subgroup analysis of studies investigating the effect of unhealthy food/beverage marketing on dietary preference

	Food/beverage preference (percentage)								
	No. of trials	RR	95% CI		No. of participants		P value	P value for interaction	I^2
			Lower	Upper	Control	Intervention			
TV advertisement	2	1.1	0.74	1.58	54	54	0.688	0.303	0.0
Advergame	2	1.5	0.95	2.37	122	193	0.082		39.4
Logo/Brand	4	1.1	0.96	1.13	612	613	0.282		0.0
Low risk of bias	2	1.0	0.93	1.12	429	430	0.641	0.061	0.0
High risk of bias	6	1.2	1.05	1.36	359	430	0.007		0.0
Healthy	-	-	-	-	-	-	-	-	-
Unhealthy/less healthy	-	-	-	-	-	-	-		-
≤ 8 years of age	4	1.2	1.04	1.37	314	386	0.012	0.223	0.0
> 8 years of age	4	1.1	0.91	1.22	474	474	0.491		34.1
Total	8	1.1	1.0	1.23	788	860	0.052	-	27.6

Table 3. 5: Summary of findings

Patient or population: children 2 to 18 years of age
Intervention: unhealthy foods/beverages marketing
Comparison: non-foods/beverages marketing and/or no marketing

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	# of participants (studies)	Quality of the evidence (GRADE)
	Risk with non-food/beverage marketing	Risk with unhealthy food/beverage marketing			
Dietary intake (kilo-calories) for 2 to 30 minutes during or after exposure to advertisements	The median food intake was 140.6 kcals	The mean food intake in the intervention group was 30.4 kcals higher (2.9 higher to 57.9 higher)	-	665 (6 RCTs)	⊕⊕⊕◯ MODERATE [1 2 3]
Dietary intake (grams) for 2 to 12 minutes during or after exposure to advertisements	The median food intake was 33.1 grams	The mean food intake in the intervention group was 4.8 grams higher (0.8 higher to 8.8 higher)	-	395 (4 RCTs)	⊕⊕⊕◯ MODERATE [3 4]
Dietary preference score after exposure to advertisements	The mean preference score was 0	Although non-significant, the mean preference score in the intervention group was 0.23 standard deviation units higher (-0.04 lower to 0.5 higher)	-	2053 (12 RCTs)	⊕⊕◯◯ LOW [5 6]
Dietary preference as a percentage after exposure to advertisements	**Risk with non-food marketing**		**RR 1.1** (1.0 to 1.2) p = 0.052	1648 (8 RCTs)	⊕⊕⊕◯ MODERATE [7]
	504 per 1000	554 per 1000 (504 to 605)			

Table 3.5. continues

*__The risk in the intervention group__ (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).
CI: Confidence interval; **RR:** Risk ratio; **OR:** Odds ratio

GRADE Working Group grades of evidence
High quality: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

1. 5 of 6 trials had an unclear risk of bias due to lack of allocation concealment and 3 of 6 trial had a high risk of bias due to lack of blinding of participants and/or assessors. However, given that dietary intake is an objective outcome, we did not rate down for risk of bias
2. Substantial heterogeneity (I-squared 72.0%) in the pooled estimate was observed. Results of our subgroup analyses on risk of bias, quality of calories and sex were significant, helping to explain the inconsistency.
3. Considering that the dietary intake is a surrogate outcome for weight gain and other patient-important outcomes, we rated down for indirectness
4. 2 of 4 trials had unclear risk of bias due to lack of allocation concealment and blinding of participants and/or assessors. However, given that dietary intake is an objective outcome, we did not rate down for risk of bias.
5. 6 of 12 trials had an unclear risk of bias due to lack of allocation concealment and 8 of 12 trial had a high risk of bias due to lack of blindness of participants and/or assessors. Dietary preference is a subjective and many of the instruments used were unvalidated so we rated down for risk of bias
6. Given the substantial heterogeneity (I-squared 87.6%) in the pooled estimate that was generally unexplained (3 of 4 subgroups were non-significant; the subgroup on age was significant but both subgroups had I-squared values > 72%), we rated down for inconsistency
7. The majority of trials (6 of 8) had an unclear or high risk of bias due to lack allocation concealment and blinding of participants and/or assessors and given that this is a subjective measure we rated down for risk of bias

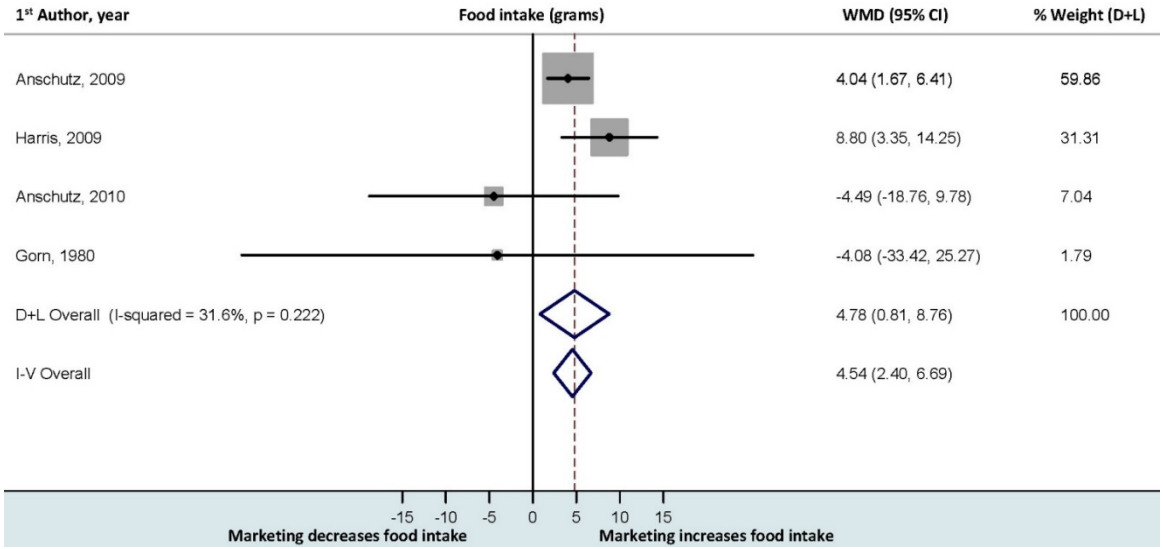
Appendices

Appendix 3. 1. Search strategy

Embase (1980-December 2014) Database Coverage: 1980-present		
#	**Searches**	**Results**
1	(child* or teen* or adolescent* or youth*).mp.	2632337
2	exp parent/	160727
3	parent*.mp.	391290
4	1 or 2 or 3	2865091
5	exp social marketing/ or exp social media/	5783
6	food commercial*.mp.	60
7	exp advertizing/	16,401
8	television advertis*.mp.	337
9	(food promot* or food advertis*).mp.	383
10	advergame*.mp.	13
11	food market*.mp.	582
12	exp recreation/ or television viewing/	42,322
13	5 or 6 or 7 or 8 or 9 or 10 or 11 or 12	64,377
14	food message*.mp.	17
15	exp fast food/	3397
16	food choice*.mp.	3414
17	food preference/	9697
18	((food or eating) adj3 (habit* or behaviour or behavior)).tw.	15809
19	(food or beverage* or snack*).mp.	575366
20	14 or 15 or 16 or 17 or 18 or 19	581038
21	breastfeed*.mp.	17901
22	20 not 21	578624
23	alcohol*.mp.	464598
24	22 not 23	550829
25	(artificial milk or infant formula).mp.	10722
26	24 not 25	547618
27	13 and 26	4372
28	4 and 27	1641
29	limit 28 to (english language and yr="1980 - 2014")	1544
30	limit 29 to ((clinical trial or randomized controlled trial or controlled clinical trial or multicenter study or phase 1 clinical trial or phase 2 clinical trial or phase 3 clinical trial or phase 4 clinical trial) and journal and (infant <to one year> or child <unspecified age> or preschool child <1 to 6 years> or school child <7 to 12 years> or adolescent <13 to 17 years>))	62

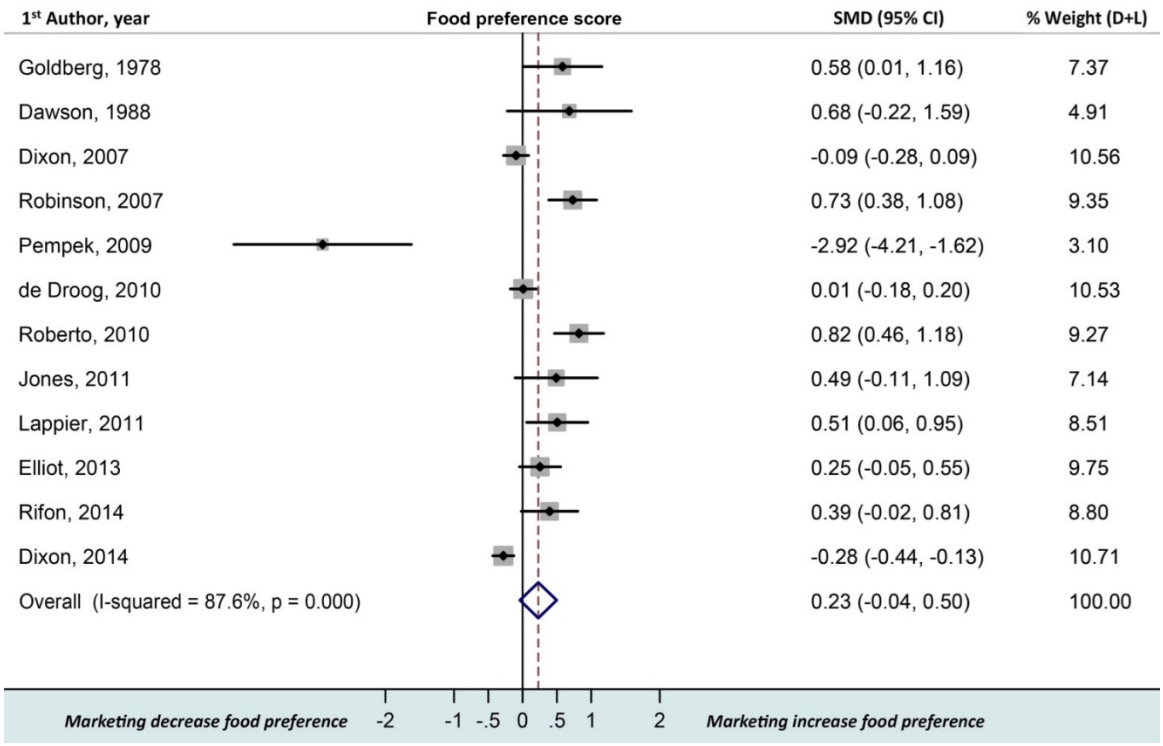
MEDLINE(R): 1966-December 2014 Database coverage: 1946-present		
#	**Searches**	**Results**
1	(child* or teen* or adolescen*).mp.	2765151
2	(youth or teen*).mp.	60756
3	exp Parents/ or parents.mp.	161902
4	1 or 2 or 3	2808168
5	exp marketing/ or exp advertising as topic/ or exp social marketing/	30373
6	exp internet/	51850
7	exp Mass Media/	39492
8	exp "play and playthings"/ or video games/	9217
9	advergame*.tw.	12
10	brand*.mp.	13695
11	(food packag* or product packag* or food label*).tw.	1785
12	Cartoons as Topic/	468
13	(food commercial* or television commercial*).tw.	121
14	5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13	140087
15	fast food.mp.	1687
16	cereal.tw.	7925
17	snack*.tw.	4521
18	exp beverages/ or exp food/	1096564
19	nutrition.tw	103975
20	15 or 16 or 17 or 18 or 19	1179808
21	Child Behavior/px [Psychology]	3371
22	exp Food Preferences/px [Psychology]	1713
23	exp feeding behavior/ or food habits/ or exp habits/	128224
24	21 or 22 or 23	131424
25	4 and 14 and 24	1403
26	alcohol*.tw. or alcohol*.mp.	313256
27	25 not 26	1372
28	breastfeeding.mp. or breastfeeding.tw.	14415
29	27 not 28	1246
30	limit 29 to (english language and humans and yr="1966 - 2014")	1165
31	limit 30 to (("all infant (birth to 23 months)" or "all child (0 to 18 years)" or "infant (1 to 23 months)" or "preschool child (2 to 5 years)" or "child (6 to 12 years)" or "adolescent (13 to 18 years)") and (clinical trial, all or clinical trial or comparative study or controlled clinical trial or journal article or randomized controlled trial or research support, nih, extramural or research support, nih, intramural or research support, non us gov't or research support, us gov't, non phs or research support, us gov't, phs))	1074

PsycINFO : 1966-December 2014, Search run January Database Coverage: 1806-present		
#	**Searches**	**Results**
1	(child* or teen* or adolescent*).mp.	695026
2	(marketing or market or advertise or advertising or promote or promoting or promotion).tw.	147375
3	food marketing.tw.	135
4	food advertis*.tw.	194
5	exp Television Advertising/	1882
6	television advertising.tw.	457
7	fast food*.tw.	908
8	(television or internet or web or brand or product or game or character or label*).tw.	188099
9	2 or 3 or 4 or 5 or 6 or 7 or 8	315993
10	exp food/	9938
11	food preferences.mp. or exp food preferences/	3900
12	exp diets/	9356
13	consumer behavior/	19885
14	((food or eating) adj3 (habit* or behaviour or behavior)).tw.	9191
15	nutrition.mp. or exp nutrition/	14513
16	beverages.tw.	2456
17	exp food intake/	12182
18	10 or 11 or 12 or 13 or 14 or 15 or 16 or 17	68179
19	breastfeeding.mp. or breastfeeding.tw.	2574
20	18 not 19	67891
21	alcohol*.tw. or alcohol*.mp.	108001
22	20 not 21	64482
23	9 and 22	18232
24	1 and 23	2446
25	limit 24 to (human and english language and yr="1966 - 2014")	2351
26	limit 25 to ((childhood <birth to 12 years> or adolescence <13 to 17 years>) and (100 childhood <birth to age 12 yrs> or 140 infancy <2 to 23 mo> or 160 preschool age <age 2to 5 yrs> or 180 school age <age 6 to 12 yrs> or 200 adolescence <age 13 to 17 yrs>) and ("0100 journal" or "0110 peer-reviewed journal" or "0120 non-peer-reviewed journal" or "0130 peer-reviewed status unknown") and (journal article or reprint))	1313



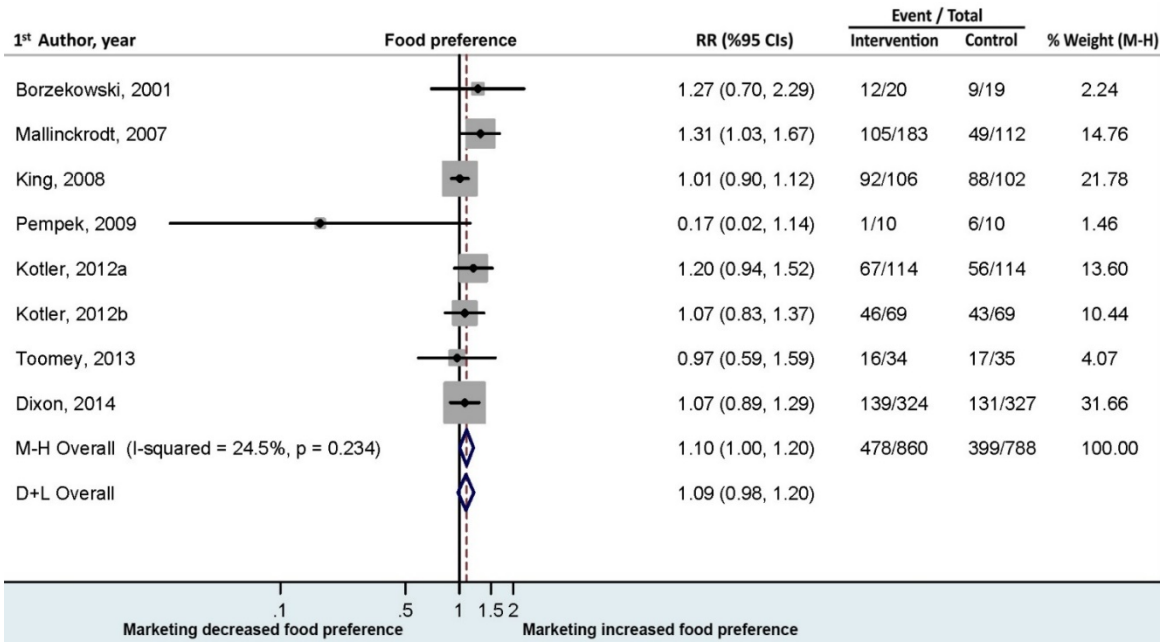
Appendix 3. 2. Forest plot showing the weighted mean difference in food intake (grams) between unhealthy food and nonfood marketing groups.

Horizontal bars denote 95% CIs. Studies are represented as squares centered on the point estimate of the result of each study. The area of the square represents the weight given to the study in the meta-analysis by STATA software. The pooled mean difference was calculated by DerSimonian–Laird (D+L) random-effects model inverse variance (I-V) fixed-effects model. The diamond represents the overall estimated effect and its 95% CIs (center line of diamond, dashed line). The solid vertical line is the line of no effect.



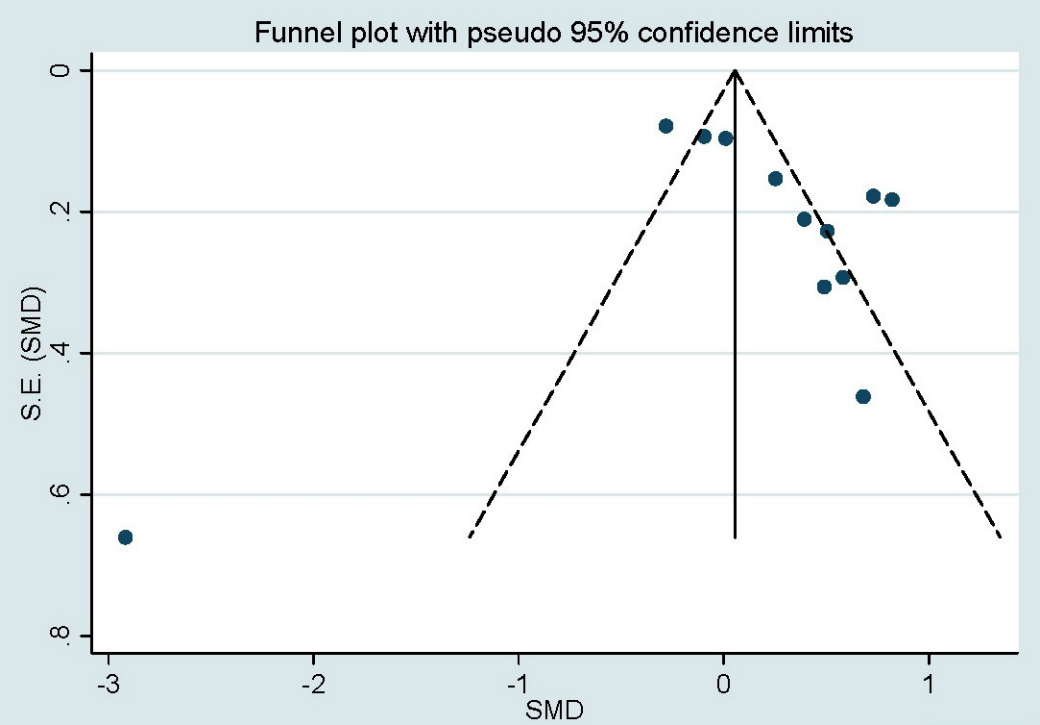
Appendix 3. 3. Forest plot showing the standardized mean difference (SMD) in food/taste preference between unhealthy food marketing and control groups.

Horizontal bars denote 95% CIs. Studies are represented as squares centered on the point estimate of the result of each study. The area of the square represents the weight given to the study in the meta-analysis by STATA software. The pooled standardized mean difference was calculated by a random-effects model. The diamond represents the overall estimated effect and its 95% CIs in each subgroup and in total (centerline of diamond, dashed line). The solid vertical line is the line of no effect.



Appendix 3. 4. Forest plot showing relative risk (RR) for unhealthy food marketing vs. control groups.

Horizontal bars denote 95% CIs. Studies are represented as squares centered on the point estimate of the result of each study. The area of the square represents the weight given to the study in the meta-analysis by STATA software. The pooled RR was calculated by DerSimonian–Laird (D+L) random-effects model Mantel-Hansel (M-H) fixed-effects model. The diamond represents the overall estimated effect and its 95% CIs in total (center line of diamond, dashed line). The solid vertical line is the line of no effect.



Appendix 3. 5. Funnel plot of studies included in the meta-analysis food preference. The standardized mean difference (SMD) is plotted on the x axis, and the Standard Error (SE) of the SMD is plotted on the y axis. Each point in the plot represents a study and its effect estimate; the shape of a symmetrical funnel suggests the absence of publication bias

Chapter 4: Human and bovine colostrum for prevention of necrotizing enterocolitis in preterm infants: A systematic review and meta-analysis of randomized trials

Behnam Sadeghirad, Rebecca L. Morgan, Dena Zeraatkar, Adriana M. Zea, Rachel Couban, Bradley C. Johnston, Ivan D. Florez

This Chapter was accepted for publication on 10 July, 2018 and was available online on 01 August 2018. In this dissertation, we present the revised submitted version. the final published paper is available as: Sadeghirad B, Morgan RL, Zeraatkar D, et al. Human and Bovine Colostrum for Prevention of Necrotizing Enterocolitis: A Meta-analysis. Pediatrics, 2018; 142 (2): e20180767. [PMID: 29991526] doi: 10.1542/peds.2018-0767.

Author Contributions:

BS and IDF conceived the study idea. BS, RLM, and IDF coordinated the systematic review. BS and IDF wrote the first draft of the manuscript. BS and RC designed the search strategy. BS, RLM, and IDF screened abstracts and full texts. BS, DZ, AMZ, and RLM acquired the data and judged risk of bias in the studies. BS, RLM, BCJ assess the certainty of evidence. BS performed the data analysis. All authors interpreted the data analysis and critically revised the manuscript.

Abstract

Context: To date there has not been a systematic review on the effects of human and bovine colostrum on mortality and morbidity associated necrotizing enterocolitis (NEC).

Objective: To determine the effectiveness and safety of bovine and human colostrum administration for reducing NEC, mortality, culture-proven sepsis, and enteral feeding tolerance in preterm infants. **Data Sources:** Search of MEDLINE, Embase, CINAHL, CENTRAL, and grey literature. **Study Selection:** Randomized controlled trials comparing human or bovine colostrum to placebo. **Data Extraction:** Two reviewers independently did screening, review, and extraction**.**

Results: A total of 8 studies (385 infants) proved eligible. In comparison to placebo, bovine and human colostrum showed no effect on incidence of NEC stage II or more (RR: 0.99; 95%CI 0.48 to 2.02, I^2=2.2%; moderate certainty of evidence), all-cause mortality (RR: 0.88; 95%CI 0.39 to 1.82, I^2=0%; moderate certainty), culture proven sepsis (RR: 0.78; 95%CI 0.53 to 1.14, I^2=0%; moderate certainty), and feed intolerance (RR: 0.97; 95%CI 0.37 to 2.56, I^2=55%; low certainty). Colostrum showed a significant effect on mean days to reach full enteral feed (MD: -3.55; 95%CI 0.33 to 6.77, I^2=41.1%; moderate certainty). The indirect comparison of bovine and human colostrum showed no difference in any of the outcomes. **Limitations:** The number of patients was modest while the number of NEC-related events was low.

Conclusion: Bovine or human colostrum has no effect on severe NEC, mortality, culture-proven sepsis, feed intolerance, or length of stay. Further research focused on the impact on enteral feeding may be needed to confirm the findings on this outcome.

Introduction

Preterm birth is defined by the World Health Organization (WHO) as live births before 37 weeks of pregnancy. Preterm birth complications are the leading cause of death among children under five years of age, and are responsible for approximately one million deaths in 2015.[1] Extremely premature (birth weight < 1,250 g) newborns have substantial mortality and morbidity, often resulting from infectious morbidities including late-onset sepsis and necrotizing enterocolitis (NEC).[2]

NEC is a multifactorial and life-threatening inflammation of the gastrointestinal tract and the most frequent surgical emergency in neonates. The mechanism of NEC is poorly understood, but research suggests that factors such as bowel hypo-perfusion, use of antibiotics and the delay to start enteral feeding seem to promote intestinal atrophy and abnormal bacterial intestinal colonization which are crucial features of the disease.[3] Despite the significant advances in neonatal care, morbidity and mortality related to NEC have remained unchanged for decades. NEC-related mortality is reported to be 20% to 30%, while in infants in need of surgery could be up to 50%.[4-6] Furthermore, infants recovering from NEC are at increased risk for microcephaly, short-bowel syndrome, serious neurodevelopmental delays, and functional disabilities.[7, 8]

Mother's milk has many immune and trophic factors (such as growth factors, cytokines, lactoferrin, lysozymes and immunoglobulins)[9, 10] that may protect newborns from infection and might have an effect on the gastrointestinal tract maturation. Mother's milk feedings have been linked with a reduced incidence of several prematurity-specific

morbidities including NEC, bacteremia, and enteral feed intolerance for premature infants.[11] Colostrum is the first milk produced by the mammary when the tight junctions in the mammary epithelium are open.[12] It has been found that the immune protective factors are more highly concentrated in the colostrum of mothers delivering premature infants than in those who give birth at term.[13, 14] This, in turn, suggests that immune components in colostrum may provide infants with protection against infection.[15]

Method of colostrum administration has been studied in two different ways: oropharyngeal and enteral. Human colostrum has been administered in small volumes directly into the buccal cavity of intubated premature infants.[16] Likewise, commercially available bovine colostrum has been administered via enteral along with the enteral feeding. Bovine colostrum also contains protective factors, which have substantial homology to their human counterparts.[17]

Randomized trials of both bovine and human colostrum in comparison to placebo have been performed in preterm infants to assess their potentially protective effects. To date, this evidence has not been systematically summarized. We have therefore conducted a systematic review to determine the effectiveness and safety of human and bovine colostrum in preterm infants for decreasing NEC related outcomes, including mortality and morbidities.

Methods

Protocol registration

The protocol for this systematic review is registered with PROSPERO: CRD 42018085566.

Data sources

We searched MEDLINE, EMBASE, the Cumulative Index to Nursing and Allied Health Literature (CINAHL), and the Cochrane Central Register of Controlled Trials for relevant published RCTs (search strategy is provided in the **Appendix 4.1**). We did not apply language or publication status restrictions. We reviewed reference lists from eligible trials and related reviews for additional eligible RCTs, and searched ClinicalTrials.gov and the WHO International Clinical Trials Registry Platform for ongoing or unpublished trials.

Study selection

Reviewers (BS, IDF, and RLM) independently screened the titles and abstracts of all identified studies using *a priori* selection criteria. Subsequently, reviewers independently assessed eligibility of the full-texts of potentially eligible studies. Reviewers resolved discrepancies through discussion.

We included RCTs that compared oropharyngeal or enteral administration of human or bovine colostrum to preterm infants (gestational age < 37 weeks) within the first week of life irrespective of when enteral feeding was initiated, the type of milk used for enteral feeding or the feed advancement regimen with placebo, standard clinical care or, standard clinical care plus placebo. Standard clinical care typically includes parenteral nutrition or feeding of own mother's milk, donor's milk, or preterm formula milk.

Our outcomes of interest were as follows: (i) NEC – stage II or more based on Bell's criteria;[18, 19] (ii) NEC-related mortality; (iii) all-cause mortality; (iv) culture-proven sepsis; (v) patent ductus arteriosus; (vi) intraventricular hemorrhage; (vii) duration of hospitalization; (viii) weight gain; and (ix) incidence of adverse events (as reported by authors).

Data abstraction and risk of bias assessment

Reviewers (BS, IDF, RLM, AMZ, DZ) extracted the following data, independently and in duplicate: (i) general study information (author's name, publication year, study design, and number of arms), (ii) population-related information (birth weight, gestational age, APGAR score at 1- and 5-minutes, percentage of cesarean-section deliveries, and percentage of infants small for gestational age), (iii) feeding details (feeding protocol and percentage of infants receiving mother's or formula milk), (iv) details on the intervention and comparison (type of colostrum, time of initiation, dose, duration of therapy, and type of control group), and (v) outcomes as listed above.

Two reviewers independently assessed risk of bias using a modified Cochrane Risk of Bias instrument for RCTs[20, 21] that addresses the following issues: random sequence generation, allocation concealment, blinding of study participants, healthcare providers, and outcome assessors/adjudicators, incomplete outcome data, and other potential sources of bias.

To assess the certainty of evidence (CoE), we used the GRADE (Grading of Recommendations, Assessment, Development, and Evaluation) approach for evidence

assessment that classifies evidence as high, moderate, low, or very low quality based on considerations of risk of bias, inconsistency, indirectness, imprecision, and publication bias.[22] We resolved disagreements between reviewers in data extraction, and assessments of risk of bias or certainty of evidence by discussion and, if needed, by third party adjudication. We used the GRADE profiler (GRADEpro GDT; https://gradepro.org/) to generate the GRADE Summary of Findings table.

Data synthesis and statistical methods

For dichotomous outcomes, we calculated the relative risk (RR) and its corresponding 95% CIs and calculated the absolute effect by multiplying the RR and its CIs with the estimated baseline risk. The median of the placebo group of included RCTs provided the baseline risk. For continuous outcomes, we calculated the mean difference and its corresponding 95% CIs.

Statistical heterogeneity was determined using the Q statistic and I^2. We used the DerSimonian–Laird random-effects model for the meta-analysis of all outcomes. Regardless of the observed statistical heterogeneity, we conducted the following pre-specified subgroup analyses: birth weight, assuming larger effects for infants with larger birth weights; gestational age, assuming larger effects for infant with higher gestational age; type of colostrum, assuming a larger effect for infants receiving human colostrum; and risk of bias, assuming larger effects for studies at high risk of bias. For subgroup analysis, we tested for interaction using a chi-square significance test, when each subgroup was represented by at least two studies.[23] We performed univariate and

multivariate meta-regression to assess the effects of birth weight, gestational age, duration of therapy, APGAR score at 1 minute and 5 minutes, percentages of C-section deliveries, and publication year on the treatment effect. We planned to examine publication bias using funnel plots for outcomes in which 10 or more studies were available.[24] We performed indirect meta-analysis using frequentist approach to compare the effect of human versus bovine colostrum. Conventional meta-analysis combines effect estimates from direct comparisons of interventions (i.e. evidence from trials with head-to-head comparison of interventions). Indirect comparisons are made by looking at the impact of the interventions of interest versus a third intervention – a common comparator (in this case, inferring the effect of bovine vs human colostrum through trials of bovine colostrum vs placebo, and human colostrum vs placebo). Indirect meta-analysis is a relatively new technique and is intended for situations where there is no direct evidence and comparisons are made pairwise. More details on the statistical methods can be found in Miladinovic et al.[25] Data were analyzed using STATA (Version 14.2, Texas, USA).

Results

Description of included studies

We identified 1,075 titles and abstracts through our literature search, of which 26 proved potentially eligible for full-text evaluations and 18 were excluded for the following reasons: (i) not randomized trials (n = 8), (ii) colostrum was not used as the intervention (n = 5), (iii) not preterm infants (n = 2), and no relevant outcome was reported (n = 3). **Figure 4.1** provides the details of study selection process.

We included eight RCTs that proved eligible enrolling 394 individuals. The intervention in six studies[13, 16, 26-29] was human colostrum, and in two studies bovine colostrum.[30, 31] Two studies enrolled preterm infants with birth weight ≤ 1.0 kg or gestational age < 28 weeks,[13, 27] five studies with birth weight ≤ 1.5 kg or gestational age < 32 weeks,[16, 26, 28-30] and one study birth weight between 1.0 kg to 1.8 kg and gestational age ≥ 28 weeks.[31] **Table 4.1** presents details of included trials.

Among the included studies, five out of eight RCTs demonstrated concerns for high risk of bias due to allocation concealment, blinding, and outcome reporting.[16, 26, 28, 29, 31] One study had issues in incomplete outcome reporting, [29] four studies had issues in blinding of participants and/or outcome assessors,[16, 26, 28, 31] and two studies had issues in concealing the treatment allocation.[28, 29] **Appendices 4.2 and 4.3** provides the summary of risk of bias assessments.

NEC and NEC-related mortality

Meta-analysis from seven studies that reported the incidence of NEC stage II or more[13, 16, 26-30] showed no difference among infants who received colostrum versus those who received placebo/usual care group (RR 0.99, 95% CI: 0.48 to 2.02; I^2 = 2.2%; moderate CoE; **Figure 4.2** and **Table 4.2**). One study reporting this outcome used bovine colostrum. Tests of interaction showed no evidence of any subgroup effect (**Table 4.3**). The univariate meta-regression confirmed the results of subgroup analysis (**Appendix 4.6**). The indirect comparison of human and bovine colostrum showed no difference (**Appendix 4.7**).

In the four RCTs that reported NEC-related mortality,[13, 16, 26, 27] no infant died as a result of developing NEC in colostrum or placebo/usual care group (moderate CoE; **Table 4.2**).

All-cause mortality

All-cause mortality was reported in seven RCTs,[13, 16, 26-28, 30, 31] and the results of meta-analysis showed no effect for colostrum compared to placebo/usual care (RR 0.84, 95%CI: 0.39 to 1.82; I^2=0%; moderate CoE; **Figure 4.3** and **Table 4.2**). There was no evidence of any subgroup effect (**Table 4.3** and **Appendix 4.6**). Two studies reporting this outcome used bovine colostrum. The indirect comparison of human and bovine colostrum showed no difference (**Appendix 4.7**).

Culture proven sepsis

In the eight studies that reported on culture proven sepsis,[13, 16, 26-31] for infants who received colostrum the risk of developing sepsis was 22% less than those who received placebo/usual care (RR 0.78, 95% CI: 0.53 to 1.14; I^2 = 0.0%; moderate CoE, **Figure 4.4, Table 4.2**). We found no evidence of subgroup effect for this outcome (**Table 4.3** and **Appendix 4.6**). Two studies reporting this outcome used bovine colostrum. The indirect comparison of human and bovine colostrum showed no difference (**Appendix 4.7**).

Feed intolerance and time to reach full feed

Of the two studies that reported feeding intolerance, one used human colostrum[26] and the other used bovine colostrum.[31] None of these studies reported a benefit for using

colostrum and the pooled estimate was not significant (RR 0.97, 95% CI: 0.37 to 2.56; I^2 = 55.5%; low CoE; **Appendix 4.4**).

Time to reach full enteral feeding was reported in six studies.[13, 26-29, 31] On average, infants receiving colostrum reached full feed 3.5 days earlier (95% CI: −0.33 to −6.77; I^2 = 38.3%; moderate CoE; **Figure 4.5, Table 4.2**). We found no evidence of subgroup effect for this outcome (**Table 4.3** and **Appendix 4.6**). The results of indirect meta-analysis showed larger effect for human colostrum, but the difference was not statistically significant (mean difference -7.1 days, 95% CI: -18.2 to 3.9; **Table 4.3**).

Other outcomes

Three studies reported on duration of hospital stay.[13, 27, 28] The results of meta-analysis didn't show any significant difference between the duration of hospitalization between the infants who received colostrum and those who received placebo/usual care (mean difference 1.26 days, 95% CI: - 13.7 to 16.3; I^2 = 41.1%; low CoE; **Appendix 4.5** and **Table 4.2**). Juhl et al. reported four cases of intraventricular hemorrhage (three grade I and one grade II) in placebo/usual care group (21.1%), while no infants in bovine colostrum group were reported to develop intraventricular hemorrhage.[31] None of the five studies that assessed the occurrence of adverse events reported any serious adverse events associated with the intervention.[13, 26-28, 31]

Discussion

To our knowledge, this the first systematic review and meta-analysis of human and bovine colostrum administration in preterm infants for necrotizing enterocolitis, mortality and related health outcomes. In our review, we synthesized the evidence from eight RCTs, including 394 infants, to describe the effect of bovine and human colostrum administration in preterm infants. Based on low to moderate CoE, we found that colostrum has no effect on mortality or morbidities in preterm infants. Nonetheless, colostrum administration resulted in less time to get full enteral feeding (moderate CoE). We explored and found no evidence of subgroup effect for any of the outcomes and our univariate meta-regression was not significant for any of the covariates (birth weight, gestational age, APGAR score at 1 and 5 minutes, proportion of infants delivered in a C-section procedure, and duration of treatment). These findings show that there is no effect of colostrum on NEC-related outcomes.

Although we did not find differences in culture proven sepsis, there was a trend towards a positive effect. This effect, if present, may be related to the immune effects of colostrum. The lack of effect may be related to the lack of power, given the relative low number of subjects studied. Further clinical trials will increase the number of patients and may change the results for this outcome.

Colostrum contains numerous protective immune and trophic factors that seem to play an important role in the first days of extra-uterine life. [9, 11]. Mother's milk provides the ideal form of administration of colostrum. However, considering that the content of bovine

colostrum has been described as similar in many components to the human colostrum[32], when the latter is not available, bovine colostrum might be considered a good alternative.

Both types of colostrum have been used in different ways. Human colostrum has been administered using an oropharyngeal method, while the bovine colostrum administration has been administered enterally. The rationale behind the oropharyngeal administration is that, due to the gastric tube feeding, preterm infants are not being exposed to the effect of protective bio-factors on the the oropharyngeal associated lymphoid tissue to obtain an effect on their immune system.[13] Thus, the administration of small amounts of colostrum in the oral mucosa aims to provide that exposure and to produce a positive impact on the immune system and therefore on the incidence of NEC-related health outcomes.

In contrast, the bovine colostrum has been administered via an enteral route.[17, 30] For example, Juhl et al. administered colostrum as a reconstituted colostrum powder to reach the required energy density[31], while Balachandran et al. used small amounts of a different product powder that was mixed with expressed human milk and given four times per day.[30] In this case, the aim was to produce an effect on the maturation of the gut as has been described on infant piglets.[33, 34]

Although they have been administered through different routes, in our review, we considered both types of colostrum. We hypothesized that the beneficial effects from colostrum contents could be similar; however, acknowledging the potential differences among both interventions, we conducted a between-study subgroup analyses and an

indirect comparison to determine potential differences based on colostrum type. In the direct comparisons of bovine and human colostrum, we did not find any difference.

The only outcome in which we found differences was the time to achieve full-enteral feeding. Colostrum administration reduced the time to achieve the enteral feeding by approximately 3.5 days. Although the definition of full enteral feeding provided by authors varied, ranging from 100 to 150 ml/Kg/day, these results may be clinically relevant. In the subgroup analysis, we did not find differences by type of colostrum. However, the results, based on this limited number of trials, trended towards a treatment effect but were non-significant. We await published data from ongoing trials to further assess the potential of colostrum for time to achieve full enteral feeding.

The certainty across the body of evidence was judged to be moderate or low. The reasons for rating down the certainty on the evidence were due to heterogeneity and imprecision. Potential reasons for heterogeneity were explored using both meta-regression and subgroup analyses, and all analyses were non-significant. Imprecision was the reason for rating down due to the lack of significant effects (confidence intervals ranged from values suggesting a substantial benefit, to values suggesting substantial harm) and the modest sample size.[35] Further RCTs with more participants and more events will likely have an impact on the precision of estimates, which in turn will improve our certainty in evidence.

To date there are at least four ongoing RCTs comparing colostrum to placebo registered in clinical trials register platforms (WHO and clinical trials.gov). Three trials are currently comparing oropharyngeal administration of human colostrum with placebo[36-38],

and one is comparing bovine colostrum with infant formula.[39] In total these studies will analyze more than 1,300 patients. Certainly, incorporating results of these trials will be add the precision to the estimates, which in turn will provide higher certainty in our estimates of effect.

Strengths of this review include explicit eligibility criteria; a comprehensive search developed with a research librarian with no language or publication status restriction; duplicate assessment of eligibility, and independent data abstraction, risk of bias and certainty of evidence assessment using the GRADE approach; and summarizing evidence for both human and bovine colostrum; consideration of possible subgroup effects. Currently, there is a protocol for a Cochrane review aiming to summarize the evidence of the administration of oropharyngeal human colostrum on morbidity and mortality in preterm infants; [40] however, this review is not considering bovine colostrum. To our knowledge, this is the first review that synthesizes the evidence from bovine colostrum.

The limitations of our review have to do with the underlying evidence. The total number of patients was modest while the number of NEC-related events was low, which along with the heterogeneity led to low CoE. The inclusion of future ongoing trials will likely lead to more precise estimates and more confidence in the results. Second, we pooled the evidence for both types of colostrum, even though they are different interventions administered through different routes. To explore the potential heterogeneity related to type and administration route of colostrum, we performed subgroup analysis and indirect comparisons to evaluate the differences between both interventions and our results demonstrated non-significant differences.

Conclusion

Moderate to low certainty evidence suggests that human and bovine colostrum have no effect on NEC incidence, mortality, length of stay and culture proven infections among preterm infants. Colostrum may reduce the time for achieving full-enteral feeding. Further studies need to confirm whether the effect on this outcome is similar between both types of colostrum or whether it is limited only to human colostrum. Data from at least four ongoing trials will be useful in providing more patients to improve the precision of estimates for each of our outcomes. Given the interest in this topic, readers should look for review updates.

References

1. World Health Organization. Preterm birth. www.who.int/mediacentre/factsheets/fs363/en/. Published 2017. Accessed March 15th, 2018.
2. Rodriguez NA, Vento M, Claud EC, Wang CE, Caplan MS. Oropharyngeal administration of mother's colostrum, health outcomes of premature infants: study protocol for a randomized controlled trial. *Trials*. 2015;16(1):453.
3. Westerbeek EA, van den Berg A, Lafeber HN, Knol J, Fetter WP, van Elburg RM. The intestinal bacterial colonisation in preterm infants: a review of the literature. *Clinical nutrition*. 2006;25(3):361-368.
4. Thyoka M, de Coppi P, Eaton S, Khoo K, Hall NJ, Curry J, et al. Advanced necrotizing enterocolitis part 1: mortality. *Eur J Pediatr Surg*. 2012;22(1):8-12.
5. Neu J, Walker WA. Necrotizing enterocolitis. *NEJM*. 2011;364(3):255-264.
6. Raval MV, Hall NJ, Pierro A, Moss RL. Evidence-based prevention and surgical treatment of necrotizing enterocolitis—A review of randomized controlled trials. *Semin Pediatr Surg*. 2013;22(2):117-121.
7. Lin PW, Stoll BJ. Necrotising enterocolitis. *The Lancet*. 2006;368(9543):1271-1283.
8. Christensen RD, Gordon PV, Besner GE. Can we cut the incidence of necrotizing enterocolitis in half-today? *Fetal Pediatr Pathol*. 2010;29(4):185-198.
9. Radillo O, Norcio A, Addobbati R, Zauli G. Presence of CTAK/CCL27, MCP-3/CCL7 and LIF in human colostrum and breast milk. *Cytokine*. 2013;61(1):26-28.
10. Hettinga K, van Valenberg H, de Vries S, Boeren S, van Hooijdonk T, van Arendonk J, et al. The Host Defense Proteome of Human and Bovine Milk. *PLOS ONE*. 2011;6(4):e19433.
11. Rodriguez NA, Caplan MS. Oropharyngeal administration of mother's milk to prevent necrotizing enterocolitis in extremely low-birth-weight infants: theoretical perspectives. *The Journal of perinatal & neonatal nursing*. 2015;29(1):81-90.
12. Zhang Y, Ji F, Hu X, Cao Y, Latour JM. Oropharyngeal colostrum administration in very low birth weight infants: A Randomized Controlled Trial. *Pediatric Critical Care Medicine*. 2017;18(9):869-875.
13. Rodriguez NA, Groer MW, Zeller JM, Engstrom JL, Fogg L, Du H, et al. A randomized controlled trial of the oropharyngeal administration of mother's colostrum to extremely low birth weight infants in the first days of life. *Neonatal Intensive Care*. 2011;24(4):31-35.
14. Lee J, Kim H-S, Jung YH, Choi KY, Shin SH, Kim E-K, et al. Oropharyngeal colostrum administration in extremely premature infants: an RCT. *Pediatrics*. 2015;135(2):e357-e366.
15. Ronayne de Ferrer PA, Baroni A, Sambucetti ME, López NE, Ceriani Cernadas JM. Lactoferrin levels in term and preterm milk. *Journal of the American College of Nutrition*. 2000;19(3):370-373.

16. Sohn K, Kalanetra KM, Mills DA, Underwood MA. Buccal administration of human colostrum: impact on the oral microbiota of premature infants. *J Perinatol*. 2016;36(2):106-111.

17. Schams D. Growth factors in milk. *Endocrine regulations*. 1994;28(1):3-8.

18. Bell MJ, Ternberg JL, Feigin RD, Keating JP, Marshall R, Barton L, et al. Neonatal necrotizing enterocolitis. Therapeutic decisions based upon clinical staging. *Ann Surg*. 1978;187(1):1-7.

19. Walsh M, Kliegman R. Necrotizing enterocolitis: treatment based on staging criteria. *Pediatr Clin North Am*. 1986;33(1):179-201.

20. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. 2011;343:d5928.

21. Akl EA, Sun X, Busse JW, Johnston BC, Briel M, Mulla S, et al. Specific instructions for estimating unclearly reported blinding status in randomized trials were reliable and valid. *Journal of clinical epidemiology*. 2012;65(3):262-267.

22. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008;336(7650):924-926.

23. Altman DG, Bland JM. Interaction revisited: the difference between two estimates. *BMJ*. 2003;326(7382):219.

24. Sterne JA, Sutton AJ, Ioannidis J, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*. 2011;343:d4002.

25. Miladinovic B, Hozo I, Chaimani A, Djulbegovic B. Indirect treatment comparison. *Stata Journal*. 2014;14(1):76-86.

26. Glass KM, Greecher CP, Doheny KK. Oropharyngeal Administration of Colostrum Increases Salivary Secretory IgA Levels in Very Low-Birth-Weight Infants. *Am J Perinatol*. 2017;34(14):1389-1395.

27. Lee J, Kim HS, Jung YH, Choi KY, Shin SH, Kim EK, et al. Oropharyngeal colostrum administration in extremely premature infants: an RCT. *Pediatrics*. 2015;135(2):e357-366.

28. Romano-Keeler J, Azcarate-Peril MA, Weitkamp JH, Slaughter JC, McDonald WH, Meng S, et al. Oral colostrum priming shortens hospitalization without changing the immunomicrobial milieu. *J Perinatol*. 2017;37(1):36-41.

29. Zhang Y, Ji F, Hu X, Cao Y, Latour JM. Oropharyngeal colostrum administration in very low birth weight infants: a randomized controlled trial. *Pediatric critical care medicine : a journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies*. 2017;18(9):869-875.

30. Balachandran B, Dutta S, Singh R, Prasad R, Kumar P. Bovine Colostrum in Prevention of Necrotizing Enterocolitis and Sepsis in Very Low Birth Weight Neonates: A Randomized, Double-blind, Placebo-controlled Pilot Trial. *Journal of tropical pediatrics*. 2017;63(1):10-17.

31. Juhl SM, Ye X, Zhou P, Li Y, Iyore EO, Zhang L, et al. Bovine colostrum for preterm infants in the first days of life: a randomized controlled pilot trial. *Journal of pediatric gastroenterology and nutrition*. 2018;66(3):471-478.

32. Davis TA, Nguyen HV, Garcia-Bravo R, Fiorotto ML, Jackson EM, Reeds PJ. Amino acid composition of the milk of some mammalian species changes with stage of lactation. *British Journal of Nutrition*. 1994;72(6):845-853.

33. Jensen ML, Sangild PT, Lykke M, Schmidt M, Boye M, Jensen BB, et al. Similar efficacy of human banked milk and bovine colostrum to decrease incidence of necrotizing enterocolitis in preterm piglets. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*. 2013;305(1):R4-R12.

34. Rasmussen SO, Martin L, Østergaard MV, Rudloff S, Li Y, Roggenbuck M, et al. Bovine colostrum improves neonatal growth, digestive function, and gut immunity relative to donor human milk and infant formula in preterm pigs. *American Journal of Physiology-Gastrointestinal and Liver Physiology*. 2016;311(3):G480-G491.

35. Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, et al. GRADE guidelines 6. Rating the quality of evidence—imprecision. *Journal of Clinical Epidemiology*.64(12):1283-1293.

36. Rodriguez NA, Vento M, Claud EC, Wang CE, Caplan MS. Oropharyngeal administration of mother's colostrum, health outcomes of premature infants: study protocol for a randomized controlled trial. *Trials*. 2015;16:453.

37. GOV CT. Bovine Colostrum for Preterm Newborns (PreColos-RCT). Clinical Trials Gov; 2018.

38. Platform; WICTR. Oral immunotherapy in very low birth weight preterm infants - randomized double-blind, placebo-controlled clinical trial. WHO International Clinical Trials Registry Platform; 2018.

39. Platform; WICTR. Efficacy of oropharyngeal administration of colostrum in reducing morbidity and mortality in very preterm infants: A randomized controlled trial - Colostrum. WHO International Clinical Trials Registry Platform; 2018.

40. Nasuf AW, Ojha S, Dorling J. Oropharyngeal colostrum in preventing mortality and morbidity in preterm infants. *Cochrane Database of Systematic Reviews*. 2015(10).

Table 4. 1: Characteristics of studies included in the systematic review

Study	**Mean BW** (grams)	**Mean GA** (weeks)	**# randomized (intervention/ control)**	**Type of colostrum**	**Duration of therapy/dose**	**Time of initiation**	**method of colostrum administration**
Rodriguez 2011 [13]	842.0	26.3	9 / 6	Human	0.2 ml every 2 hours for 2 days	Within 48 hours of life	Oropharyngeal
Lee 2015 [27]	815.0	26.8	24 / 24	Human	0.2 ml every 3 hours for 3 days	48 to 96 hours after birth	Oropharyngeal
Sohn 2016 [16]	1053.5	27.0	6 / 6	Human	0.2 ml every 2 h for 46 hours	Median age of 39 hours (range 32 to 87)	Oropharyngeal
Balachandran 2017 [30]	1202.9	29.9	43 / 43	Bovine	1.2 g to 2.0 g four time per day for until discharge or death or day 21 of life	In the first 96 hours of life	Oro-gastric tube
Romano-Keeler 2017 [28]	1219.5	25.5	48 / 51	Human	0.2 ml every 6 hours for 5 days	In the first 48 hours of life	Oropharyngeal
Glass 2017 [26]	1109.0	28.4	17 / 13	Human	0.2 ml every 3 hours for 7 days	In the first 48 hours of life	Oropharyngeal
Zhang 2017 [29]	1244.5	30.2	32 / 32	Human	0.2 ml every 4 hours for 7 days	Between day 2 to 4 of life	Oropharyngeal
Juhl 2018 [31]	1487.5	30.5	21 / 19	Bovine	volume limited by a pre-set total protein intake of 4.5 g/kg/day for 10 to 14 days	In the first 48 hours of life	Enteral

BW: birth weight; GA: gestational age; NR: not reported.

Table 4. 2: Summary of findings

Population: Preterm infants (Gestational Age < 37 weeks)
Intervention: Colostrum (human or bovine); **Comparator**: No colostrum

Outcome (studies) № of participants	Relative effect (95% CI)	Anticipated absolute effects (95% CI)			Certainty
		Risk with No colostrum	Risk with colostrum	Difference	
NEC stage II or more (7 RCTs) № of participants: 345	**RR 0.99** (0.48 to 2.02)	9.4%	**9.3%** (4.5 to 18.9)	**0.1% fewer** (4.9 fewer to 9.5 more)	⨁⨁⨁◯ MODERATE [a,b]
Culture proven sepsis (8 RCTs) № of participants: 385	**RR 0.78** (0.53 to 1.14)	21.1%	**16.4%** (11.2 to 24.0)	**4.6% fewer** (9.9 fewer to 2.9 more)	⨁⨁⨁◯ MODERATE [a,b]
All-cause mortality (7 RCTs) № of participants: 330	**RR 0.88** (0.39 to 1.82)	7.4%	**6.5%** (2.9 to 13.5)	**0.9% fewer** (4.5 fewer to 6.1 more)	⨁⨁⨁◯ MODERATE [a,b]
Feed intolerance (2 RCTs) № of participants: 70	**RR 0.97** (0.37 to 2.56)	43.8%	**42.4%** (16.2 to 100.0)	**1.3% fewer** (27.6 fewer to 68.3 more)	⨁⨁◯◯ LOW [a,b,c]
NEC-related mortality (4 RCTs) № of participants: 105	not estimable	0.0%	**0.0%** (0.0 to 0.0)	**0.0% fewer** (0 fewer to 0 fewer)	⨁⨁⨁◯ MODERATE [a,d]
Duration of hospital stay (3 RCTs) № of participants: 160	-	The mean duration of hospital stay was **79.0** Day	-	MD **1.26 Day more** (13.73 fewer to 16.26 more)	⨁⨁◯◯ LOW [a,b,e]
Time to reach full feed (6 RCTs) № of participants: 285	-	The mean time to reach full enteral feed was **22.1** Day	-	MD **3.55 Day fewer** (0.33 fewer to 6.77 fewer)	⨁⨁⨁◯ MODERATE [a,f]

***The risk in the intervention group** (and its 95% CIs) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).
CI: Confidence interval; **RR:** Risk ratio; **MD:** Mean difference

GRADE Working Group grades of evidence
High certainty: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of effect, but it is possible that it is substantially different
Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Explanations
a. While some studies were at risk for bias due to allocation concealment and blinding, subgroup analyses did not suggest that any heterogeneity was introduced due to risk of bias. Thus, we did not rate down the evidence for risk of bias.
b. 95% CI includes values suggesting substantial benefit and substantial harm; thus, we rated down for imprecision.
c. I^2 value is 56%, suggesting some heterogeneity; however, exploratory analyses did not highlight the source. Thus, we rated down for inconsistency due to unexplained heterogeneity.
d. No events were reported for either arm.
e. I^2 value is 41%, demonstrating potential heterogeneity; however, exploratory analyses (meta-regression and subgroup analyses) did not highlight the source. We rated down for inconsistency due to unexplained heterogeneity.
f. I^2 value is 38%, potential heterogeneity; however, exploratory analyses (meta-regression and subgroup analyses) did not highlight the source. Thus, we decided to rate down for inconsistency due to unexplained heterogeneity.

Table 4. 3: Results of the meta-analysis and subgroup analysis of RCTs assessing the effects of colostrum

Outcome/subgroups*		No. of trials	ES	95% CI		No. of participants		I^2	P value for interaction
				Lower	Upper	Intervention	Control		
NEC (stage ≥ II)	Human	5	0.83	0.39	1.75	131	128	0.0	- †
	Bovine	1	4.00	0.47	34.34	43	43	-	
	BW < 1,000 g	2	0.67	0.22	2.07	33	30	-	0.488
	BW ≥ 1,000 g	5	1.26	0.49	3.27	141	141	9.4	
	Low risk of bias	5	1.13	0.51	2.49	99	92	0.0	0.752
	High (due to allocation concealment)	2	0.62	0.06	6.06	75	79	52.5	
	Low risk of bias	3	1.29	0.23	7.30	76	73	54.0	0.871
	High (due to blinding)	4	0.98	0.35	2.70	98	98	3.6	
	Total	**7**	**0.99**	**0.48**	**2.02**	**174**	**141**	**2.2**	**-**
Mortality	Human	5	0.74	0.27	2.06	104	100	0.0	0.765
	Bovine	2	1.00	0.31	3.21	64	62	-	
	BW < 1,000 g	2	0.86	0.15	4.80	33	30	33.7	0.922
	BW ≥ 1,000 g	5	0.99	0.36	2.78	135	132	0.0	
	Low risk of bias	6	0.78	0.35	1.72	120	111	0.0	- †
	High (due to allocation concealment)	1	3.18	0.13	76.31	48	51	-	
	Low risk of bias	3	0.83	0.36	1.88	76	73	0.0	0.947
	High (due to blinding)	4	0.98	0.11	8.91	92	89	1.9	
	Total	**7**	**0.84**	**0.39**	**1.82**	**168**	**162**	**0.0**	**-**
Culture proven sepsis	Human	6	0.79	0.51	1.23	131	128	0.0	0.873
	Bovine	2	0.73	0.33	1.61	64	62	0.0	
	BW < 1,000 g	2	1.22	0.24	6.20	33	30	43.8	0.744
	BW ≥ 1,000 g	6	0.72	0.41	1.24	162	160	0.0	
	Low risk of bias	6	0.83	0.55	1.25	120	111	0.0	0.331
	High (due to allocation concealment)	2	0.47	0.16	1.43	75	79	0.0	
	Low risk of bias	3	0.80	0.50	1.26	76	73	0.0	0.839
	High (due to blinding)	5	0.73	0.36	1.48	119	117	0.0	
	Total	**8**	**0.78**	**0.53**	**1.14**	**195**	**190**	**0.0**	**-**
Time to reach full feed (days)	Human	5	-2.87	-6.02	0.28	123	122	34.3	- †
	Bovine	1	-9.60	-19.46	0.26	21	19	-	
	BW < 1,000 g	2	-4.55	-14.33	5.23	31	30	72.5	0.794
	BW ≥ 1,000 g	4	-3.17	-6.65	0.31	113	111	29.8	

	Low risk of bias	4	-4.19	-9.40	1.03	69	62	48.8	0.854
	High (due to allocation concealment)	2	-3.47	-9.06	2.13	75	79	48.1	
	Low risk of bias	2	-4.55	-14.33	5.23	31	30	72.5	0.794
	High (due to blinding)	4	-3.17	-6.65	0.31	113	111	29.8	
	Total	**6**	**-3.55**	**-6.77**	**-0.33**	**144**	**141**	**38.3**	**-**

RR: relative risk; ES: effect estimate (weighted mean difference for time to reach full enteral feed, and relative risk for the remaining outcomes); BW: mean birth weight as reported by RCTs.

* We did not perform any subgroup analysis for duration of hospital stay and incidence of feeding intolerance as there was 3 or less studies reporting those outcomes.

† Due to small number of trials, we did not perform a statistical test of interaction between the two group.

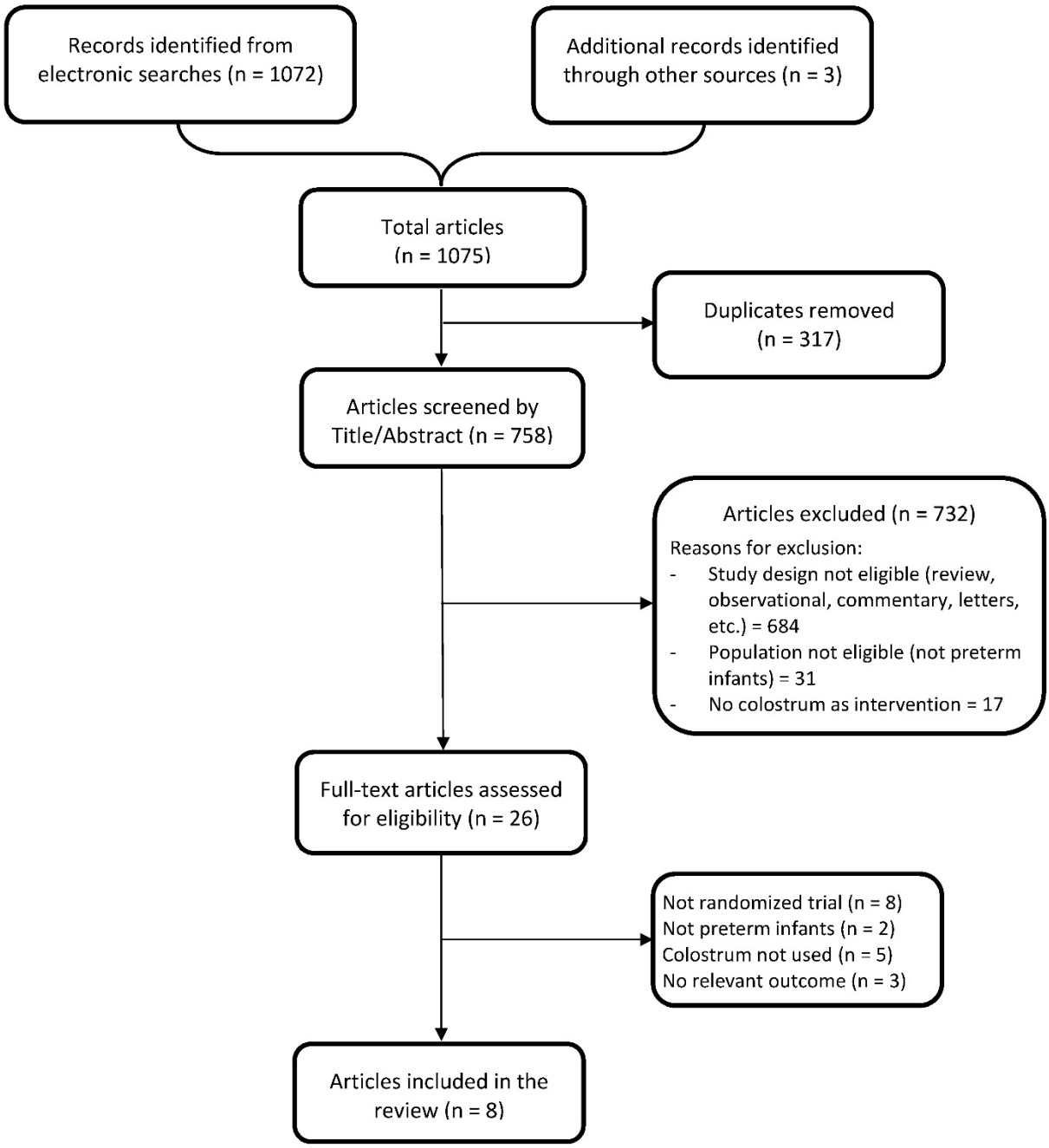


Figure 4. 1: Flow diagram of database searches and articles included in the systematic review and meta-analysis

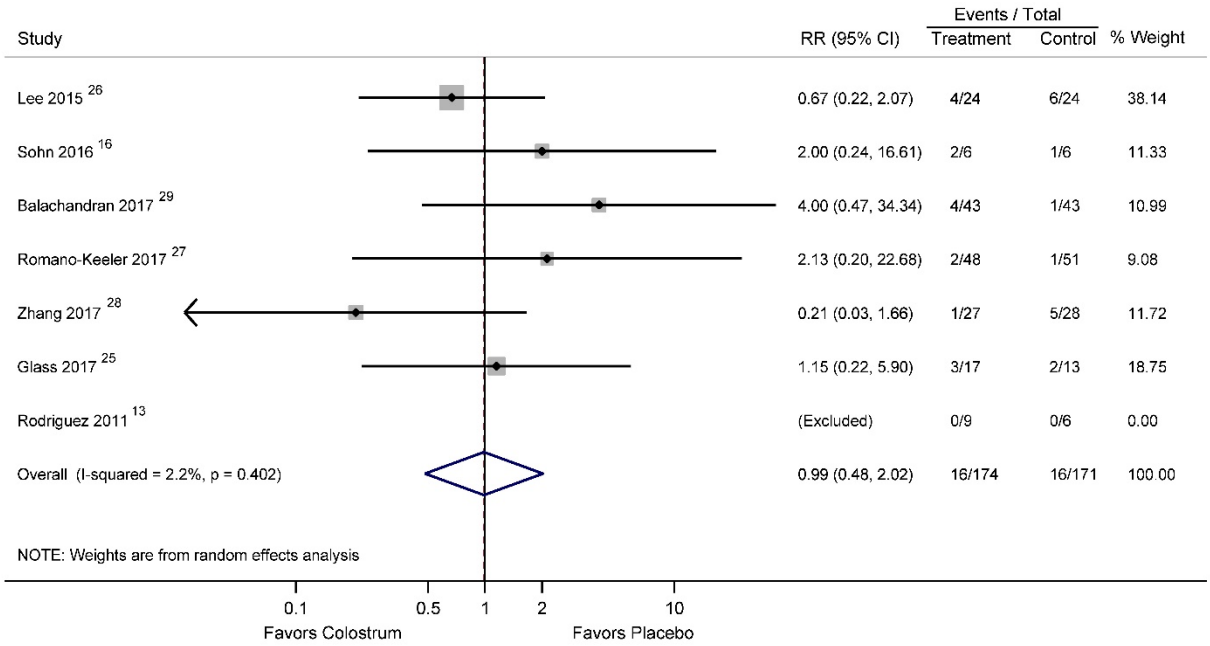


Figure 4. 2: Forest plot revealing RR for NEC stage II or more (based on Bell's criteria) for colostrum versus placebo groups.

Horizontal bars denote 95% CIs. Studies are represented as squares centered on the point estimate of the result of each study. The area of the square represents the weight given to the study in the meta-analysis. The pooled RR was calculated by DerSimonian-Laird random-effects model. The diamond represents the overall estimated effect and its 95% CI in total (center line of diamond, dashed line). The solid vertical line is the line of no effect. Weights are from random effects analysis.

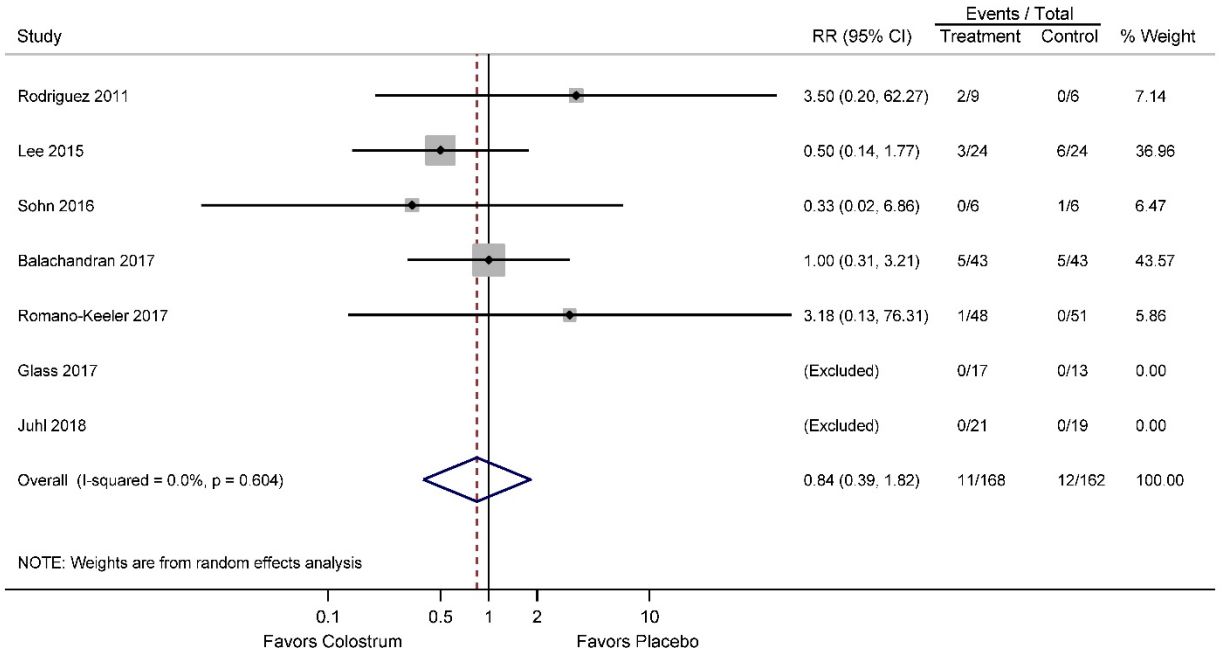


Figure 4. 3: Forest plot revealing RR for mortality for colostrum versus placebo groups.

Horizontal bars denote 95% CIs. Studies are represented as squares centered on the point estimate of the result of each study. The area of the square represents the weight given to the study in the meta-analysis. The pooled RR was calculated by DerSimonian-Laird random-effects model. The diamond represents the overall estimated effect and its 95% CI in total (center line of diamond, dashed line). The solid vertical line is the line of no effect. Weights are from random effects analysis.

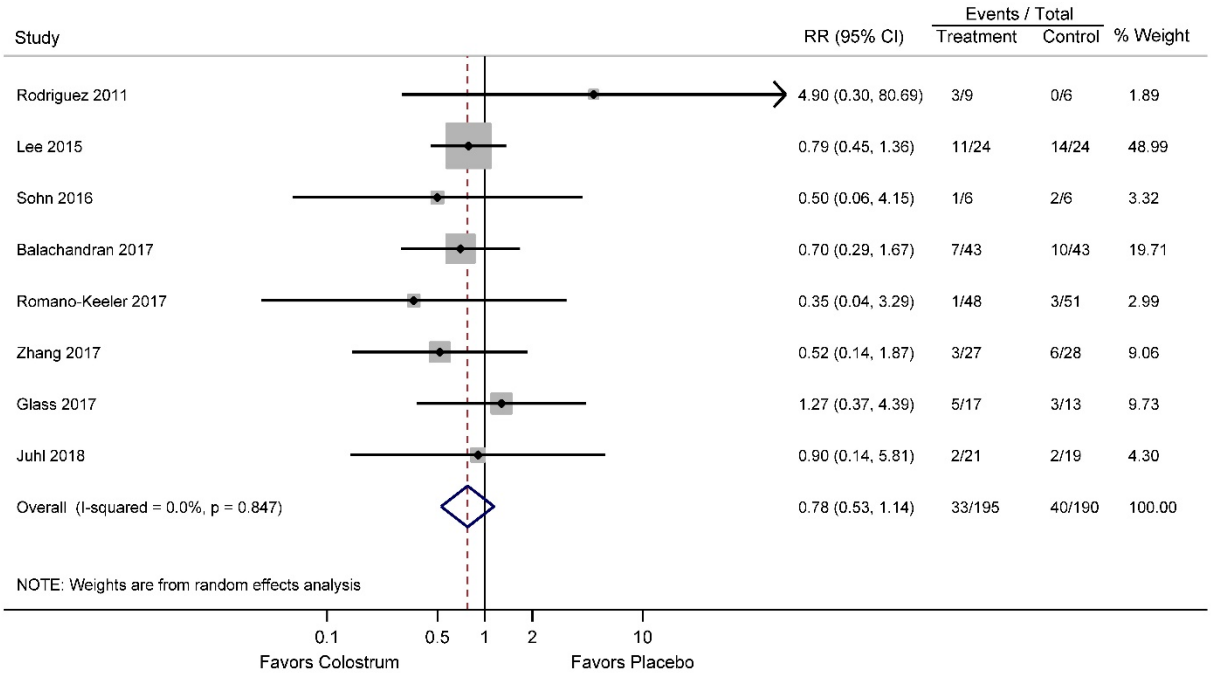


Figure 4. 4: Forest plot revealing RR for culture-proven sepsis for colostrum versus placebo groups.

Horizontal bars denote 95% CIs. Studies are represented as squares centered on the point estimate of the result of each study. The area of the square represents the weight given to the study in the meta-analysis. The pooled RR was calculated by DerSimonian-Laird random-effects model. The diamond represents the overall estimated effect and its 95% CI in total (center line of diamond, dashed line). The solid vertical line is the line of no effect. Weights are from random effects analysis.

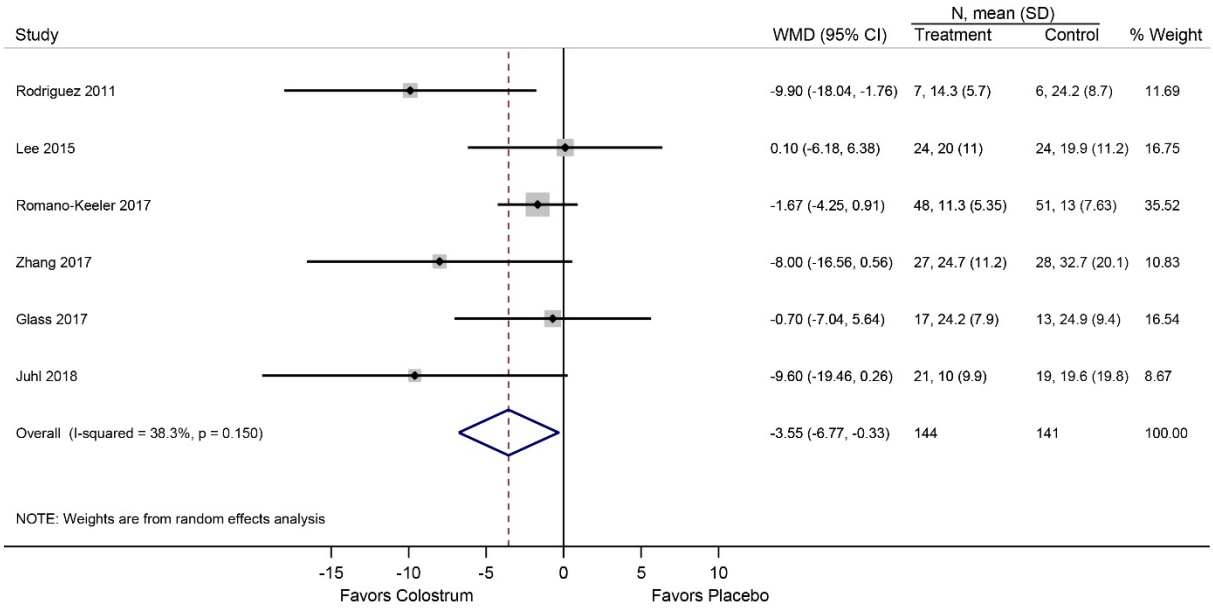


Figure 4. 5: Forest plot revealing the weighted mean difference in mean time to reach full enteral feed for colostrum versus placebo groups.

Horizontal bars denote 95% CIs. Studies are represented as squares centered on the point estimate of the result of each study. The area of the square represents the weight given to the study in the meta-analysis. The pooled mean difference was calculated by DerSimonian-Laird random-effects model. The diamond represents the overall estimated effect and its 95% CI in total (center line of diamond, dashed line). The solid vertical line is the line of no effect. Weights are from random effects analysis. WMD, weighted mean difference.

Appendices

Appendix 4. 1. Search terms and strategies

Database: OVID Medline Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) 1946 to Present
Search Strategy:
---

1 colostrum.mp. or Colostrum/ (8840)
2 exp Infant, Newborn/ (609585)
3 (newborn or neonate or neonatal or premature or low birth weight or VLBW or LBW or infan* or neonat*).mp (1566797)
4 2 or 3 (1566797)
5 1 and 4 (4261)
6 randomized controlled trial.pt. (515870)
7 controlled clinical trial.pt. (101741)
8 randomized.ab. (452787)
9 placebo.ab. (210486)
10 drug therapy.fs. (2198498)
11 randomly.ab. (311971)
12 trial.ab. (477314)
13 groups.ab. (1924550)
14 or/6-13 (4545735)
15 exp animals/ not humans.sh. (4813914)
16 14 not 15 (3932519)
17 5 and 16 (375)
18 Enterocolitis, Necrotizing/ (3040)
19 (necroti?ing enterocolitis or enterocolitis necroticans or nec or typhlitis or pneumatosis intestinalis).tw. (9964)
20 18 or 19 (10403)
21 1 and 20 (86)
22 17 or 21 (436)
23 remove duplicates from 22 (396)

Database: Embase <1974 to 2018 January 05>
Search Strategy:
---

1 colostrum.mp. or exp colostrum/ (8742)
2 exp infant/ (977082)
3 newborn disease/ or exp low birth weight/ or prematurity/ (154678)
4 (newborn* or neonat* or infan* prematur* or very low birth weight or low birth weight or VLBW or LBW).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word] (764338)
5 or/2-4 (1185494)
6 1 and 5 (3848)

7 exp animals/ or exp invertebrate/ or animal experiment/ or animal model/ or animal tissue/ or animal cell/ or nonhuman/ (25390069)
8 human/ or normal human/ or human cell/ (19166171)
9 7 and 8 (19118363)
10 7 not 9 (6271706)
11 6 not 10 (1659)
12 limit 11 to ("therapy (maximizes specificity)" or "therapy (best balance of sensitivity and specificity)") (156)
13 ((doubl* adj blind*) or (singl* adj blind*) or allocat* or assign* or cross over* or crossover* or
factorial* or placebo* or random* or volunteer*).tw. (1850487)
14 crossover-procedure/ or double-blind procedure/ or randomized controlled trial/ or single-blind
procedure/ (543562)
15 13 or 14 (1943077)
16 11 and 15 (178)
17 12 or 16 (186)
18 necrotizing enterocolitis/ (9042)
19 (necroti?ing enterocolitis or enterocolitis necroticans or nec or typhlitis or pneumatosis intestinalis).tw. (12632)
20 18 or 19 (15634)
21 1 and 20 (123)
22 21 not 10 (86)
23 17 or 22 (255)

CINAHL

#	Query	Results
S35	S30 OR S34	112
S34	S1 AND S33	21
S33	S31 OR S32	1,858
S32	TX (necroti?ing enterocolitis or enterocolitis necroticans or nec or typhlitis or pneumatosis intestinalis)	1,549
S31	(MH "Enterocolitis, Necrotizing")	1,025
S30	S6 OR S29	95
S29	S5 AND S28	64
S28	S26 NOT S27	550,733
S27	(MH "Animals+")	36,860
S26	S13 OR S18 OR S25	555,026
S25	S19 OR S20 OR S21 OR S22 OR S23 OR S24	482,785
S24	(MH "Prospective Studies+")	208,456
S23	(MH "Evaluation Research+")	39,985
S22	(MH "Comparative Studies")	99,037
S21	"latin square"	137
S20	(MH "Study Design") OR (MH "Crossover Design") OR (MH "Experimental Studies+")	200,581
S19	(MH "Random Sample+")	65,875
S18	S14 OR S15 OR S16 OR S17	206,657
S17	"random*"	198,077
S16	"placebo*"	33,294

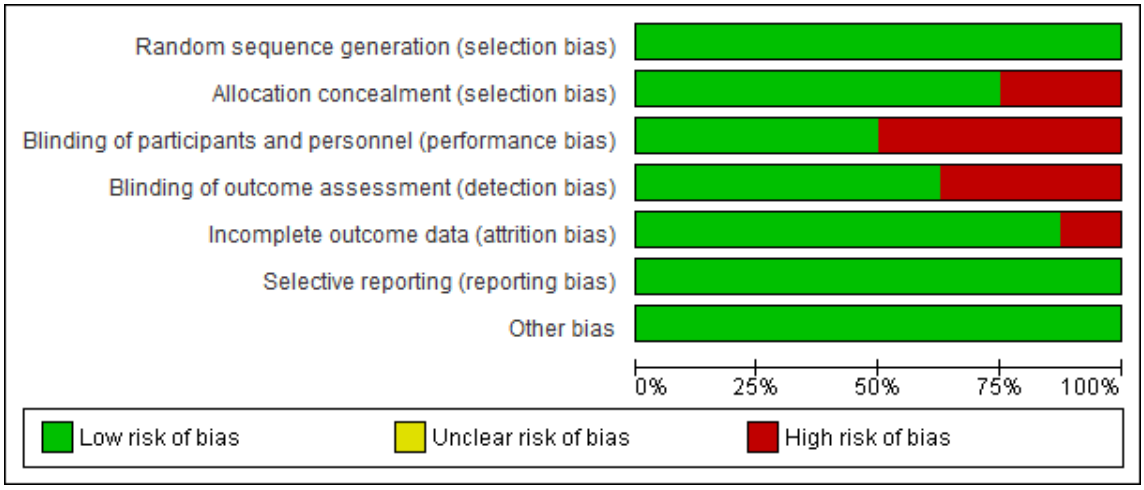
S15	(MH "Placebos")	8,191
S14	(MH "Placebo Effect")	1,160
S13	S7 OR S8 OR S9 OR S10 OR S11 OR S12	191,282
S12	"triple-blind"	128
S11	"single-blind"	8,355
S10	"double-blind"	28,990
S9	clinical W3 trial	122,214
S8	"randomi?ed controlled trial*"	65,850
S7	(MH "Clinical Trials+")	151,174
S6	S5	91
S5	S1 AND S4	222
S4	S2 OR S3	271,911
S3	infan* or newborn* or neonat* or premature or preterm or very low birth weight or low birth weight or VLBW or LBW)	271,911
S2	(MH "Infant") OR (MH "Infant, Newborn") OR (MH "Infant, Premature") OR (MH "Infant, High Risk") OR (MH "Infant, Hospitalized")	138,368
S1	(MH "Colostrum") OR "colostrum"	354

Search Name:
Date Run: 08/01/18 19:21:46.579
Description: Cochrane Central (library)

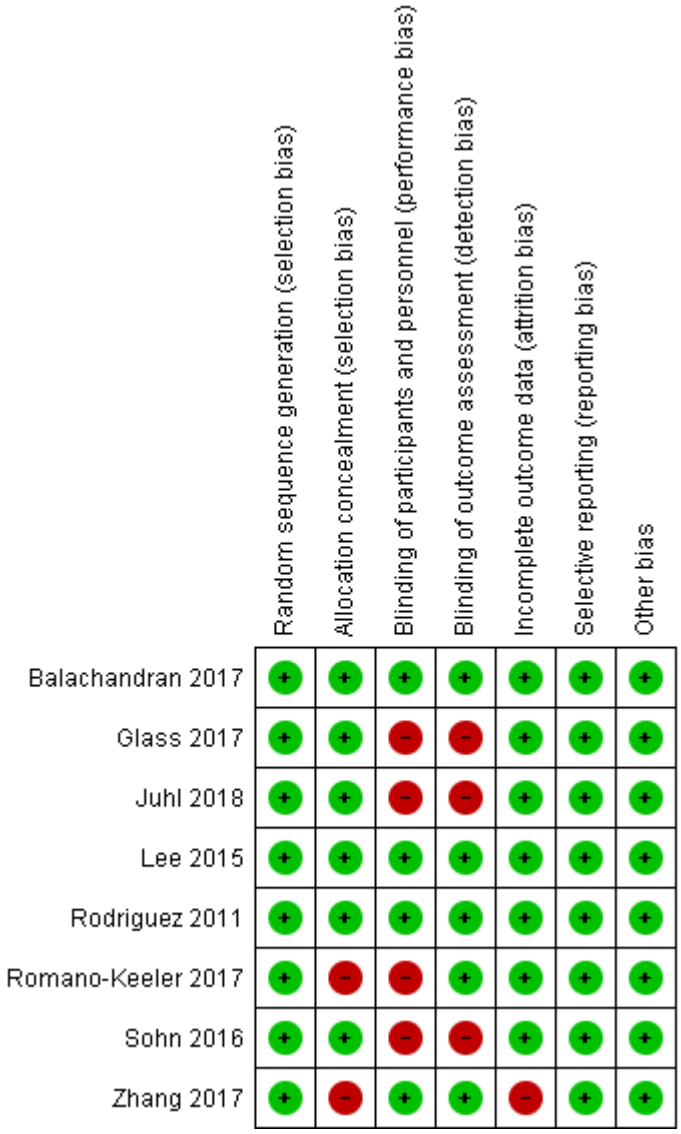
ID	Search	Hits
#1	colostrum	280
#2	#1 in Trials, with Neonatal Group in Review Groups	30
#3	infant or newborn or neonate or neonatal or premature or preterm or very low birth weight or low birth weight or VLBW or LBW	63346
#4	#1 and #3 in Trials	118
#5	#2 or #4	118
#6	(necroti?ing enterocolitis or enterocolitis necroticans or nec or typhlitis or pneumatosis intestinalis)	1434
#7	#1 and #2 in Trials	30
#8	#5 or #7	118

PubMed

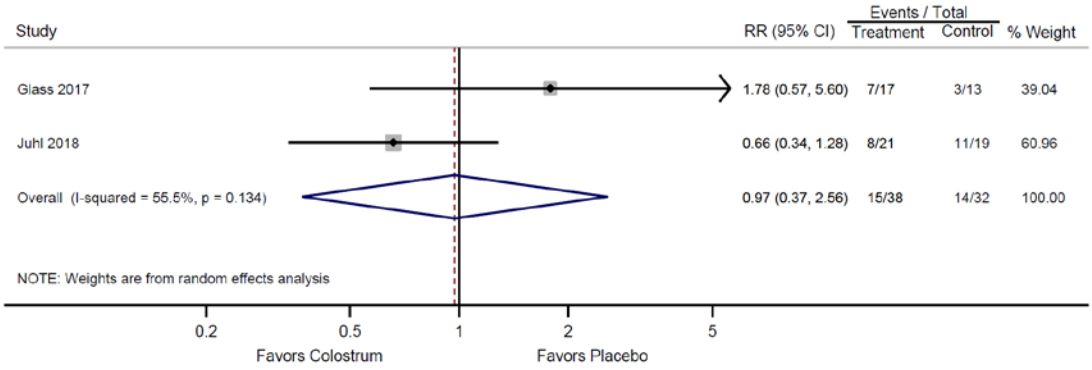
Search (((((((((("Enterocolitis, Necrotizing"[Mesh]) OR (((necrotizing enterocolitis or enterocolitis necroticans or nec or typhlitis or pneumatosis intestinalis))))) OR necrotising enterocolitis)) AND colostrum)) OR (((((infant, newborn[MeSH] OR newborn OR neonate OR neonatal OR premature OR low birth weight OR VLBW OR LBW or infan* or neonat*)))) AND colostrum))) AND (((publisher[sb] OR inprocess[sb] OR pubmednotmedline[sb] OR pubstatusaheadofprint)))
Sort by: PublicationDate.



Appendix 4. 2. Summary of risk of bias assessments among the included RCTs.
Reviewers' judgements about each risk of bias item presented as percentages across all included studies

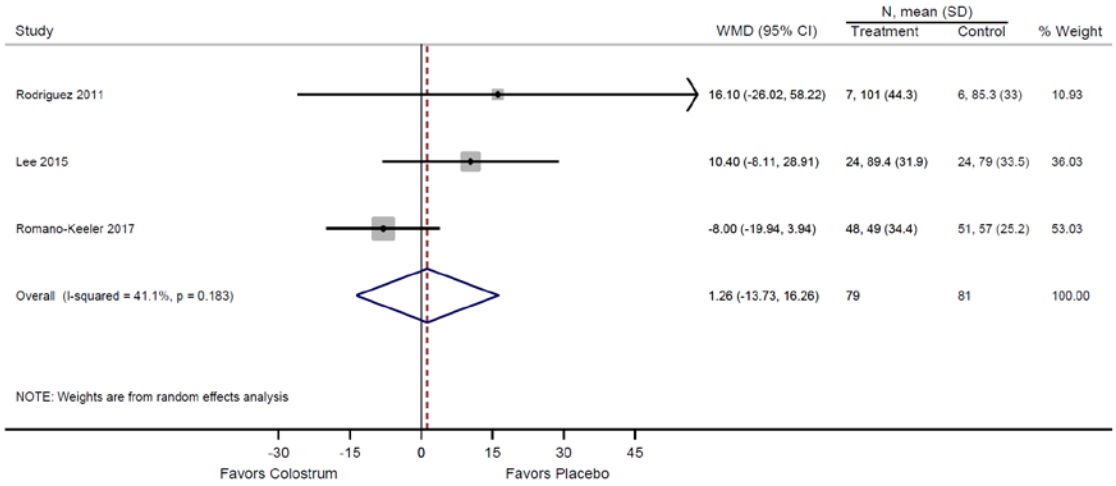


Appendix 4. 3. Summary of risk of bias assessments among the included RCTs.
Reviewers' judgements about each risk of bias item for each included study



Appendix 4. 4. Forest plot showing relative risk (RR) for feeding intolerance for colostrum vs. placebo groups.

Horizontal bars denote 95% CIs. Studies are represented as squares centred on the point estimate of the result of each study. The area of the square represents the weight given to the study in the meta-analysis. The pooled RR was calculated by DerSimonian–Laird random-effects model. The diamond represents the overall estimated effect and its 95% CIs in total (centre line of diamond, dashed line). The solid vertical line is the line of no effect.



Appendix 4. 5. Forest plot showing the weighted mean difference (WMD) in hospital stay (days) for colostrum vs. placebo groups.

Horizontal bars denote 95% CIs. Studies are represented as squares centred on the point estimate of the result of each study. The area of the square represents the weight given to the study in the meta-analysis. The pooled mean difference was calculated by DerSimonian–Laird random-effects model. The diamond represents the overall estimated effect and its 95% CIs in total (centre line of diamond, dashed line). The solid vertical line is the line of no effect.

Appendix 4. 6. Results of the meta-regression analysis for the effects of colostrum

Outcome	Covariate	Coefficient	95% CI	No. of studies	P value
NEC (stage ≥ II)	Publication year	1.17	0.58, 2.36	7	0.584
	Duration of therapy (day)	0.86	0.52, 1.43	7	0.481
	Mean birth weight (g)	1.00	0.99, 1.01	7	0.568
	Mean gestational age (weeks)	0.93	0.49, 1.77	7	0.776
	APGAR (1 minute)	1.00	0.50, 2.01	6	0.996
	APGAR (5minutes)	0.58	0.08, 4.06	6	0.484
	% C-section deliveries	1.01	0.97, 1.06	7	0.529
Mortality	Publication year	1.03	0.50, 2.15	7	0.913
	Duration of therapy (day)	1.13	0.65, 1.96	7	0.593
	Mean birth weight (g)	1.00	0.99, 1.01	7	0.415
	Mean gestational age (weeks)	0.94	0.43, 2.07	7	0.845
	APGAR (1 minute)	1.19	0.45, 3.14	5	0.603
	APGAR (5minutes)	2.76	0.15, 49.98	6	0.385
	% C-section deliveries	1.00	0.95, 1.05	7	0.963
Culture proven sepsis	Publication year	0.81	0.42, 1.59	8	0.479
	Duration of therapy (day)	1.01	0.65, 1.59	8	0.948
	Mean birth weight (g)	1.00	0.99, 1.00	8	0.636
	Mean gestational age (weeks)	1.01	0.56, 1.83	8	0.977
	APGAR (1 minute)	0.93	0.50, 1.71	6	0.746
	APGAR (5minutes)	1.08	0.18, 6.48	7	0.915
	% C-section deliveries	1.01	0.97, 1.05	8	0.692
Time to reach full feed	Publication year	0.68	-1.80, 3.16	6	0.489
	Duration of therapy (day)	0.74	-1.95, 3.43	6	0.485
	Mean birth weight (g)	-0.01	-0.03, 0.02	6	0.637
	Mean gestational age (weeks)	-1.00	-3.57, 1.57	6	0.340
	APGAR (1 minute)	-2.02	-7.71, 3.66	4	0.265
	APGAR (5minutes)	-3.31	-8.31, 1.68	6	0.139
	% C-section deliveries	0.09	-0.21, 0.38	6	0.466

Appendix 4. 7. Results of indirect meta-analysis for the effects of colostrum

Outcome*	Comparison	Effect estimate** (95% CI)
NEC (stage ≥ II)	Human colostrum vs. Placebo	0.82 (0.39, 1.71)
	Bovine colostrum vs. Placebo	4.00 (0.47, 34.34)
	Human vs. Bovine colostrum	4.87 (0.50, 47.21)
Mortality	Human colostrum vs. Placebo	0.74 (0.28, 2.00)
	Bovine colostrum vs. Placebo	0.99 (0.33, 3.03)
	Human vs. Bovine colostrum	1.34 (0.30, 5.94)
Culture proven sepsis	Human colostrum vs. Placebo	0.79 (0.51, 1.23)
	Bovine colostrum vs. Placebo	0.73 (0.33, 1.61)
	Human vs. Bovine colostrum	0.93 (0.38, 2.28)
Days to reach full feed (weighted mean difference)	Human colostrum vs. Placebo	-9.60 (-19.63, 0.43)
	Bovine colostrum vs. Placebo	-2.46 (-7.19, 2.26)
	Human vs. Bovine colostrum	-7.14 (-18.22, 3.95)

* The number of included RCTs for the remaining outcomes was not enough to allow indirect estimate of the effects.

** Effect estimates are relative risk unless otherwise reported.

Chapter 5: Comparative Effectiveness of Prophylactic Therapies for Necrotizing Enterocolitis in Preterm Infants: Protocol for a Network Meta-analysis of Randomized Trials

Behnam Sadeghirad, Ivan D. Florez, Rebecca L. Morgan, Farid Foroutan, Yaping Chang, Dena Zeraatkar, Malgorzata M. Bala, Shaneela Shahid, Taoying Lu, Long Ge, Joseph Beyene, Martin Offringa, Philip M. Sherman, Enas el Gouhary, Gordon H. Guyatt, Bradley C. Johnston

This Chapter was accepted for publication on 17 August, 2017 and was available online on 17 September, 2018. In this dissertation, we present the revised submitted version. the final published paper is available as: Sadeghirad B, Florez ID, Chang Y, et al. Comparative Effectiveness of Prophylactic Therapies for Necrotizing Enterocolitis in Preterm Infants: Protocol for a Network Meta-analysis of Randomized Trials. International Journal of Preventive Medicine, 2018; 9: 83. [PMID: 30283615] doi: 10.4103/ijpvm.IJPVM_328_17.

Author Contributions:

BS and BCJ conceived the study idea. BS wrote the first draft of the manuscript.

Comparative effectiveness of prophylactic therapies for necrotizing enterocolitis in preterm infants: Protocol for a network meta-analysis of randomized trials

Authors

Behnam Sadeghirad [1,2]; Ivan D. Florez[1,3]; Yaping Chang[1]; Farid Forutan[1,4]; Dena Zeraatkar[1]; Rebecca L. Morgan[1]; Shaneela Shahid[1,9]; Malgorzata M. Bala[5]; Joseph Beyene[1]; Martin Offringa[6]; Thomasin Adams-Webber[7]; Philip M. Sherman[8]; Enas El-Gouhary[9]; Gordon H. Guyatt[1]; Bradley C. Johnston[1,10,11]

Authors affiliations:

1. Department of Health Research Methods, Evidence, and Impact (HEI), McMaster University, Hamilton, ON, Canada
2. HIV/STI Surveillance Research Center, and WHO Collaborating Center for HIV Surveillance, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran
3. Department of Pediatrics, University of Antioquia, Medellin, Colombia
4. Heart Failure/Transplant Program, Toronto General Hospital, University Health Network, Toronto, ON, Canada
5. Department of Hygiene and Dietetics, Jagiellonian University Medical College, Krakow, Poland
6. Child Health Evaluative Sciences, The Hospital for Sick Children Research Institute, Toronto, ON, Canada
7. Hospital Library, The Hospital for Sick Children, Toronto, ON, Canada
8. Division of Gastroenterology, Hepatology and Nutrition, Hospital for Sick Children, Department of Pediatrics, University of Toronto, Toronto, ON, Canada
9. Department of Pediatrics, McMaster University, Hamilton, ON, Canada
10. Department of Community Health and Epidemiology, Dalhousie University, Halifax, NS, Canada
11. Systematic Overviews through advancing Research Technology (SORT), Child Health Evaluative Sciences, The Hospital for Sick Children Research Institute, Toronto, ON, Canada

ABSTRACT

Introduction: Necrotizing enterocolitis (NEC) is a common and devastating disease with high morbidity and mortality in premature infants. Current literature on prevention of NEC has limitations including lack of direct and indirect comparisons of available therapies.

Methods and analysis: We will search MEDLINE, EMBASE, Science Citation Index Expanded, Social Sciences Citation Index, CINAHL, Scopus, ProQuest Dissertations and Theses database, and grey literature sources to identify eligible trials evaluating NEC preventive therapies. Eligible studies will (1) enroll preterm (gestational age < 37 weeks) and/or low birth weight (birth weight < 2500 g) infants, (2) randomize infants to any preventive intervention or a placebo, or alternative active or non-active intervention. Our outcomes of interest are severe NEC (stage II or more, based on Bell's criteria), all-cause mortality, NEC related mortality, late onset sepsis, duration of hospitalization, weight gain, time to establish full enteral feeds, and treatment related adverse events. Two reviewers will independently screen trials for eligibility, assess risk of bias, and extract data. All discrepancies will be resolved by discussion. We will specify a priori explanations for heterogeneity between studies. For available comparisons between treatment and no treatment, and direct comparisons of treatments, we will conduct conventional meta-analysis using a random effects model. We will conduct a network meta-analysis using a random effects model within the Bayesian framework using Markov chain Monte Carlo methods to assess relative effects of eligible interventions. We will assess the certainty in direct, indirect, and network estimates using the GRADE approach.

Ethics and dissemination: This is a systematic review of literature not requiring primary data collection or formal ethics approval. We will disseminate our findings through a peer-reviewed publication and conference presentations.

Keywords: necrotizing enterocolitis; preterm infants; preventive therapies; multiple treatment comparison; systematic review.

Corresponding author: Bradley C. Johnston
Department of Health Research Methods, Evidence, and Impact (HEI), McMaster University
1280 Main Street West, L8S 4K1
Hamilton, Ontario, Canada, M5G 0A4
Email: bradley.johnston@sickkids.ca

Introduction

Necrotizing enterocolitis (NEC) is among the most important diseases of the gastrointestinal tract and the most frequent surgical emergency in neonates. The mechanism of NEC is poorly understood, but occurs as a result of death of intestinal tissue, which may occur as a result of bacteria in the intestinal tract, or reduced blood delivery. It is characterized by damage to the intestinal tract, which ranges from mucosal damage to full-thickness necrosis and perforation.[1, 2] The staging system originally described by Bell et al [3] categorizes NEC into 3 stages: 1) suggestive, 2) definite, and 3) severe. Stage 1 NEC presents as feeding intolerance or symptoms of advanced prematurity; infants with stage 2 NEC require medical management, and stage 3 requires surgical intervention.[4]

The incidence of NEC, which varies across countries and neonatal centers,[2] is estimated to be approximately 3 per 1,000 live births; and it occurs in 1% to 5% of neonatal intensive care unit admissions.[5, 6] NEC is mainly associated with prematurity and low birth weight.[7] The incidence of NEC in neonates of very low birth weight (<1500 g) remained unchanged from 1997 to 2007, ranging from 3% to 15%.[1, 8] According to data from 2009, the incidence of NEC increased and it is now the 11th leading cause of death in infants.[9]

Despite advances in neonatal intensive care, morbidity and mortality related to NEC has remained unchanged. The NEC associated rate of death is reported to be 20% to 30% and the rate is higher among infants in need of surgery - up to 50%.[1, 10, 11] NEC is associated with substantial economic burden, with an estimated annual hospitalization cost of more than $500 million in the United States.[1] Infants recovering from NEC are at increased risk

for prolonged parenteral nutrition and its related complications, short-bowel syndrome, serious neurodevelopmental delays, and functional disabilities.[6, 7]

Current recommendations for management include prompt, early diagnosis, medical management, and surgery if warranted.[5] The preferred approach to combat NEC is prophylactic therapy**.** More than 10 systematic reviews have evaluated various preventive strategies, including maternal and donor breast milk feeding,[12, 13] prophylactic probiotics, [14-17]oral lactoferrin,[18] supplementation of formula milk with prebiotic,[19] arginine,[20] or glutamine,[21] and immunoglobulin administration [22].

Nevertheless, considerable gaps in the current literature exist. In particular, most randomized controlled trials (RCTs) compare active treatment to non-active comparators (e.g. placebo) and there are few direct comparisons among preventive strategies. No systematic review has evaluated all RCT evidence of the leading prophylactic therapies for NEC and no review performed or compared the efficacy and safety of all preventive therapies with one another.[11-15, 18-25] The only available network meta-analysis assessing the efficacy and safety of food additives (probiotics and prebiotics) for preventing NEC included 25 RCTs,[26] while the most recent meta-analyses on probiotics included 49 RCTs. No review has used the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach for evaluating the certainty of prophylactic therapies for NEC. Further, none of the published meta-analyses used emerging methods for handling missing participant outcome data and for assessing risk of bias associated with missing data on risk of bias for NEC prevention.[27, 28] These limitations highlight the need for a comprehensive review using a systematic approach that compares all available

evidence for the prevention of NEC. We aim to conduct a systematic review and network meta-analysis of randomized trials on the effects of preventive therapies in preterm (gestational age < 37 weeks) infants on severe NEC, all-cause mortality, NEC related mortality and culture-positive sepsis.

Methods

Standardized Reporting

Our protocol conforms to the preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 guideline.[29]

Search strategy

We will systematically search MEDLINE, EMBASE, Science Citation Index Expanded and Social Sciences Citation Index, CINAHL, Scopus, ProQuest Dissertations and Theses database, and Cochrane Central Register of Controlled Trials (CENTRAL). Our grey literature search will include trial registries (including ISRCTN registry; clinicaltrials.gov; and WHO international RCT registry), BIOSIS Previews, and Google Scholar to find relevant trials. We will not apply language or publication status restrictions. We will work with an experienced medical librarian to develop a search strategy for each database. Reference lists from eligible trials and relevant literature reviews will be scanned for additional trials that may meet our inclusion criteria.

Eligibility criteria and study selection

Trials will be eligible if they enroll preterm (gestational age < 37 weeks) and/or low birth weight (birth weight < 2500 grams) infants randomized to any of preventive interventions

listed below compared to an alternative intervention, placebo or no intervention. Eligible prophylactic interventions will include: maternal or donor breast milk feeding with or without human milk fortifiers, immunoglobulin (IgG or a combination of IgG/IgA) administration, prebiotics (lactoferrin, inulin, galacto- or fructo-oligosaccharides), colostrum, arginine, glutamine, probiotics and combination of probiotics and prebiotics (synbiotics). Studies published in duplicate or studies that used data from a similar study population in different publications in part or full will be identified and we will extract data from the publication with the most complete data set (e.g. publications with largest sample size and/or longest duration of follow-up).

Pairs of reviewers, working independently will screen titles and abstracts of identified articles and acquire and assess the full-text publication for eligibility when one or both reviewers consider a study as potentially eligible. Reviewers will resolve disagreements by consensus and, if disagreements are unresolved, discuss discrepancies with a more experienced team member with relevant expertise. We will pilot this step on 10 randomly selected articles (with a ratio of 1:1 eligible and non-eligible) and repeat the process until we reach 80% agreement.

Data abstraction

To help ensure the reliability of independent data extraction, we will begin by piloting our data extraction forms on three randomly selected eligible articles, repeating the process if we find substantial challenges. After our forms have been piloted and standardized, we will conduct calibration exercises between reviewers. To calibrate, we will randomly select four articles that have met our eligibility criteria and each team member will abstract data.

Subsequently, team members will meet to resolve the disagreements. We will repeat this process until we reach agreement on 90% of data abstraction items. With accompanying data extraction instructions generated from our piloting and calibration exercises, reviewers, working in pairs, will independently extract all data and resolve discrepancies through discussion. From the included RCTs, the following data will be extracted into a standardized spreadsheet: study characteristics (the first author, publication year, country of origin, and funding source), participant and trial characteristics (sample size, mean gestational age, birth weight, and corresponding measure of variance (e.g. standard deviation)), characteristics of interventions and comparators (time of initiation, doses, species and strains if prebiotics or probiotics used, treatment durations), outcomes of interest (Severe NEC - stage II or more based on Bell's criteria,[3, 30] all-cause mortality, NEC related mortality, and late onset sepsis, duration of hospitalization, weight gain, time to establish full enteral feeds, and treatment related adverse events).

Risk of bias assessment

Among eligible studies, we will independently assess the following risk of bias issues based on the modified version of the Cochrane risk of bias tool for RCTs: random sequence generation, allocation concealment, blinding study participants (in the case of our study, infants' parents), personnel, and outcome assessors, incomplete outcome data, and selective outcome reporting.[31] The modified instrument rather than the standard response options (high, low, or unclear risk of bias) will use the following responses: 'definitely yes' or 'probably yes' (considered as low risk of bias), or 'definitely no' or 'probably no' (considered as high risk of bias).[32] These response options have published evidence of

validity for assessing blinding, and will allow our risk of bias assessments to avoid 'unclear' as a response option.[32] Any discrepancy in assessment of risk of bias will be resolved by discussion, or third party adjudication if needed. We will attempt to contact the authors of eligible studies for missing information regarding risk of bias assessments and primary/secondary outcomes.

Data synthesis

For each direct paired comparison, we will calculate relative risk and absolute risk, and the associated 95% confidence intervals (CIs) for dichotomous outcomes. For continuous outcomes, we will analyze the results using weighted mean differences with corresponding 95% CIs. We will employ methods described in Cochrane Handbook both to estimate the mean and SD where median, range, and sample size were reported, and to impute the SD if the SE or SD for the differences are not reported.[33] We will use the Q statistic and I^2 to determine statistical heterogeneity for conventional pair-wise meta-analysis and will look for clinical and/or methodological sources of heterogeneity across included RCTs.[34] We will perform subgroup analyses regardless of heterogeneity estimates. We will use the DerSimonian–Laird random-effects model for the meta-analysis of all outcomes.

Network meta-analysis methods

We will use a random effects model within the Bayesian framework using Markov chain Monte Carlo methods to assess the relative effects of eligible preventive interventions.[35, 36] However, if we observe any random-effects network estimate inconsistent with its direct estimate, we will report fixed-effects model outputs. We will simulate 100,000 iterations and test the model convergence using the Gelman-Rubin statistic.[37] For estimating the

precision of the effects, we will use 95% credible intervals, via the 2.5 and 97.5 percentiles obtained from the simulations.[38]

Although the assumptions for network meta-analysis are similar to conventional meta-analysis, additional key assumptions are transitivity (there are no effect modifiers influencing the indirect comparisons) and coherence (direct and indirect effect estimates being similar).[39] We will identify issues of incoherence by comparing direct evidence (i.e. estimates from pairwise comparisons) with indirect evidence (i.e. estimates from network meta-analysis) using node splitting method.[34, 40] We will use a Wald test to test any statistical difference between the direct and the indirect estimates.[41]

We expect results to differ between studies and we have developed three hypotheses to explain variability: (1) infants with lower birth weight will show smaller treatment effect; (2) infants receiving intervention added to their mother's milk versus donor's milk or formula will show larger treatment effects; (3) RCTs with higher risk of bias will show larger treatment effects than trials with lower risk of bias.

We will report our findings with probability statements of intervention effects. Probability rankings allow us to report a chance percentage of which interventions rank higher;[42] however, simplifying the results of a network down to probabilities can lead to misinterpretations, specifically, when particular comparisons (i.e. nodes) are not well-connected and/or when certainty in evidence varies between comparisons. Following display of the rank probabilities using rankogram, we will use the surface under the cumulative ranking (SUCRA) to aid in interpretation of relative effect of the interventions; an intervention with a SUCRA value of 100 is certain to be the best, whereas an intervention

with 0 is certain to be the worst.[42] We will use STATA (StataCorp, Release 14.2, Texas, USA) and WinBUGS (Version 1.4, Cambridge, UK) for statistical analyses.

Assessing certainty in (quality of) the evidence

To assess the certainty in (quality of) estimates of effect across each outcome of interest, we will use the GRADE (Grading of Recommendations, Assessment, Development, and Evaluation) approach that classifies evidence as high, moderate, low, or very low quality. The starting point for certainty in estimates for randomized trials is high but may be rated down based on limitations in risk of bias, imprecision, inconsistency, and indirectness and publication bias.[43] The GRADE evidence assessment will be presented in a summary of findings table. We will also use the GRADE approach to assess the certainty in indirect and network (mixed) effect estimates.[44] Indirect effect estimates are calculated from available 'loops' of evidence, which includes first order loops (based on a single common comparator treatment, the difference between the treatment A and B is based on comparisons of A and C as well as B and C) or higher order loops (more than one intervening treatment connecting the two interventions). We will visually examine the network map and where first order loops are available for indirect comparisons, the certainty of evidence will be the lower of the ratings of certainty for the two direct estimates contributing to the first order loop (for instance, for the indirect estimate of the effect between A and C through comparisons of A versus C – high quality evidence and B versus C – moderate quality evidence, the certainty will be 'moderate' – the lowest of the two direct estimates). In the absence of a first order loop, a higher order loop will be used to rate certainty in evidence, and it will be the lower of the ratings of certainty for the direct estimates contributing to the loop. However, we

may rate down the certainty further for intransitivity.[44] The transitivity assumption implies similarity of trials in terms of population, intervention, outcomes, settings, and trial methodology.[38]

Discussion

NEC is a devastating gastrointestinal condition among low birth weight neonates and has been one of the most challenging diseases to prevent and eradicate.[1, 2] Given its relatively high incidence, the high socioeconomic burden, and scarcity of evidence on the comparative effectiveness of preventive interventions which has likely contributed to variable practice patterns among clinicians, there is a need for a high-quality systematic review and network meta-analysis of the common prophylactic therapies to inform evidence-based prevention of NEC.

There may be limitations to our proposed review methods including the ability to assess risk of publication bias and assess subgroup analysis across diverse interventions using network meta-analysis methods. Our protocol has attempted to document the proposed methods a priori, including plans to address the anticipated challenges of such an NMA (e.g. handling missing participant data, assessing subgroup effects and network meta-regression, calculating absolute risk within network of preventive treatments) and assess the certainty in estimates using the GRADE approach. To ensure that our findings are translated to the neonatology community, we will publish our results in an accessible peer-reviewed journal and present our findings at national and international scientific conferences and on The Hospital for Sick Children and McMaster Children's Hospital websites.

References

1. Neu J, Walker WA. Necrotizing enterocolitis. NEJM 2011; 364:255-64.
2. Yajamanyam PK, Rasiah SV, Ewer AK. Necrotizing enterocolitis: current perspectives. Res Rep Neonatol 2014; 4:31-42.
3. Bell MJ, Ternberg JL, Feigin RD, Keating JP, Marshall R, Barton L, et al. Neonatal necrotizing enterocolitis. Therapeutic decisions based upon clinical staging. Ann Surg 1978; 187:1-7.
4. Neu J. Neonatal necrotizing enterocolitis: An update. Acta Pædiatrica 2005; 94:100-5.
5. Berman L, Moss RL. Necrotizing enterocolitis: An update. Semin Fetal Neonatal Med 2011; 16:145-50.
6. Christensen RD, Gordon PV, Besner GE. Can we cut the incidence of necrotizing enterocolitis in half-today? Fetal Pediatr Pathol 2010; 29:185-98.
7. Lin PW, Stoll BJ. Necrotising enterocolitis. The Lancet 2006; 368:1271-83.
8. Yee WH, Soraisham AS, Shah VS, Aziz K, Yoon W, Lee SK, et al. Incidence and timing of presentation of necrotizing enterocolitis in preterm infants. Pediatrics 2012; 129:e298-e304.
9. Centers for Disease Control and Prevention. Infant, neonatal, and postneonatal deaths, percent of total deaths, and mortality rates for the 15 leading causes of infant death by race and sex. United States: 2009: Published 2012 [Accessed March 12, 2015].
10. Thyoka M, de Coppi P, Eaton S, Khoo K, Hall NJ, Curry J, et al. Advanced necrotizing enterocolitis part 1: mortality. Eur J Pediatr Surg 2012; 22:8-12.
11. Raval MV, Hall NJ, Pierro A, Moss RL. Evidence-based prevention and surgical treatment of necrotizing enterocolitis—A review of randomized controlled trials. Semin Pediatr Surg 2013; 22:117-21.
12. Henderson G, Anthony MY, McGuire W. Formula milk versus maternal breast milk for feeding preterm or low birth weight infants. Cochrane Database Syst Rev 2007:CD002972.
13. Quigley M, McGuire W. Formula versus donor breast milk for feeding preterm or low birth weight infants. Cochrane Database Syst Rev 2014; 4:CD002971.
14. AlFaleh K, Anabrees J. Probiotics for prevention of necrotizing enterocolitis in preterm infants. Evidence-based child health : a Cochrane review journal 2014; 9:584-671.
15. Deshpande G, Rao S, Patole S, Bulsara M. Updated meta-analysis of probiotics for preventing necrotizing enterocolitis in preterm neonates. Pediatrics 2010; 125:921-30.
16. Stenger MR, Reber KM, Giannone PJ, Nankervis CA. Probiotics and prebiotics for the prevention of necrotizing enterocolitis. Curr Infect Dis Rep 2011; 13:13-20.

17. Wang Q, Dong J, Zhu Y. Probiotic supplement reduces risk of necrotizing enterocolitis and mortality in preterm very low-birth-weight infants: an updated meta-analysis of 20 randomized, controlled trials. J Pediatr Surg 2012; 47:241-8.

18. Pammi M, Abrams SA. Oral lactoferrin for the prevention of sepsis and necrotizing enterocolitis in preterm infants. Cochrane Database Syst Rev 2011:CD007137.

19. Srinivasjois R, Rao S, Patole S. Prebiotic supplementation of formula in preterm neonates: a systematic review and meta-analysis of randomised controlled trials. Clin Nutr 2009; 28:237-42.

20. Mitchell K, Lyttle A, Amin H, Shaireen H, Robertson H, Lodha A. Arginine supplementation in prevention of necrotizing enterocolitis in the premature infant: an updated systematic review. BMC Pediatrics 2014; 14:226.

21. Moe-Byrne T, Wagner JV, McGuire W. Glutamine supplementation to prevent morbidity and mortality in preterm infants. Cochrane Database Syst Rev 2012; 3:CD001457.

22. Foster J, Cole M. Oral immunoglobulin for preventing necrotizing enterocolitis in preterm and low birth-weight neonates. Cochrane Database Syst Rev 2004:CD001816.

23. Bury RG, Tudehope D. Enteral antibiotics for preventing necrotizing enterocolitis in low birthweight or preterm infants. Cochrane Database Syst Rev 2001:CD000405.

24. Morgan J, Young L, McGuire W. Slow advancement of enteral feed volumes to prevent necrotising enterocolitis in very low birth weight infants. Cochrane Database Syst Rev 2013; 3:CD001241.

25. Morgan J, Young L, McGuire W. Delayed introduction of progressive enteral feeds to prevent necrotising enterocolitis in very low birth weight infants. Cochrane Database Syst Rev 2014; 12:CD001970.

26. Yu W, Sui W, Mu L, Yi W, Li H, Wei L, et al. Preventing necrotizing enterocolitis by food additives in neonates: A network meta-analysis revealing the efficacy and safety. Medicine (Baltimore) 2017; 96:e6652.

27. Ebrahim S, Akl EA, Mustafa RA, Sun X, Walter SD, Heels-Ansdell D, et al. Addressing continuous data for participants excluded from trial analysis: a guide for systematic reviewers. Journal of Clinical Epidemiology 2013; 66:1014-21.e1.

28. Akl EA, Johnston BC, Alonso-Coello P, Neumann I, Ebrahim S, Briel M, et al. Addressing dichotomous data for participants excluded from trial analysis: a guide for systematic reviewers. PLoS One 2013; 8:e57132.

29. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Systematic reviews 2015; 4:1.

30. Walsh M, Kliegman R. Necrotizing enterocolitis: treatment based on staging criteria. Pediatr Clin North Am 1986; 33:179-201.

31. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011; 343:d5928.

32. Akl EA, Sun X, Busse JW, Johnston BC, Briel M, Mulla S, et al. Specific instructions for estimating unclearly reported blinding status in randomized trials were reliable and valid. J Clin Epidemiol 2012; 65:262-7.

33. Higgins JP, Green S. Cochrane Handbook for Systematic Reviews of Interventions, version [5.1.0] (updated March 2011). The Cochrane Collaboration2011.

34. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003; 327:557-60.

35. Ades AE, Sculpher M, Sutton A, Abrams K, Cooper N, Welton N, et al. Bayesian methods for evidence synthesis in cost-effectiveness analysis. Pharmacoeconomics 2006; 24:1-19.

36. Lumley T. Network meta-analysis for indirect treatment comparisons. Stat Med 2002; 21:2313-24.

37. Brooks SP, Gelman A. Alternative methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistic 1998; 7:434-55.

38. Salanti G, Higgins JP, Ades A, Ioannidis JP. Evaluation of networks of randomized trials. Stat Methods Med Res 2008; 17:279-301.

39. Donegan S, Williamson P, D'Alessandro U, Tudur Smith C. Assessing key assumptions of network meta-analysis: a review of methods. Res Synth Methods 2013; 4:291-323.

40. Lu G, Ades AE. Assessing evidence inconsistency in mixed treatment comparisons. J Am Stat Assoc 2006; 101:447-59.

41. White IR. Network meta-analysis. The Stata Journal 2015; 15:951–85.

42. Salanti G, Ades AE, Ioannidis JPA. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. J Clin Epidemiol 2011; 64:163-71.

43. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008; 336:924-6.

44. Puhan MA, Schünemann HJ, Murad MH, Li T, Brignardello-Petersen R, Singh JA, et al. A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. BMJ 2014; 349:5630.

Chapter 6: Probiotics, prebiotics, and synbiotics for prevention of mortality and morbidity in preterm infants: a systematic review and network meta-analysis of randomized trials

Behnam Sadeghirad, Ivan D. Florez, Rebecca L. Morgan, Farid Foroutan, Yaping Chang, Dena Zeraatkar, Malgorzata M. Bala, Shaneela Shahid, Taoying Lu, Long Ge, Joseph Beyene, Martin Offringa, Philip M. Sherman, Enas el Gouhary, Gordon H. Guyatt, Bradley C. Johnston

This Chapter was submitted to The Lancet on 16 March 2019 and, as of 21 May 2019, is currently under review.

Author Contributions:

BCJ and GHG conceptualized and designed the study; BCJ and BS coordinated and supervised the systematic review.BS, IDF, RLM, FF, YC, DZ, MMB, SS, and TL selected the articles, extracted the data, assessed the risk of bias. BS performed the data analysis; BS, IDF and RLM assessed the quality of the evidence and interpret the results. BS and BCJ drafted the manuscript.

Probiotics, prebiotics, and synbiotics for prevention of mortality and morbidity in preterm infants: a systematic review and network meta-analysis of randomized trials

Authors

Behnam Sadeghirad, PharmD, MPH [1,2]

Ivan D. Florez, MD, MSc [1,3]

Rebecca L. Morgan, MPH, PhD [1]

Farid Foroutan, HBSc [1,4]

Yaping Chang, PhD [1]

Dena Zeraatkar, MSc [1]

Malgorzata M. Bala, MD, PhD [5]

Shaneela Shahid, MD, MSc [1,6]

Taoying Lu, MSc [7,8]

Long Ge, PhD [1,9]

Joseph Beyene, PhD [1]

Martin Offringa, MD [10]

Philip M. Sherman, MD, FRCPC [11]

Enas el Gouhary, MD [6]

Gordon H. Guyatt, MD [1]

Bradley C. Johnston, PhD [1,12,13]

Authors affiliations:

12. Department of Health Research Methods, Evidence, and Impact (HEI), McMaster University, Hamilton, ON, Canada
13. Michael G. DeGroote Institute for Pain Research and Care, McMaster University, Hamilton, Ontario, Canada
14. Department of Pediatrics, University of Antioquia, Medellin, Colombia
15. Heart Failure/Transplant Program, Toronto General Hospital, University Health Network, Toronto, ON, Canada

16. Chair of Epidemiology and Preventive Medicine, Department of Hygiene and Dietetics, Jagiellonian University Medical College, Krakow, Poland
17. Department of Pediatrics, McMaster University, Hamilton, ON, Canada
18. Department of Clinical Epidemiology, Guangzhou University of Chinese Medicine, Guangzhou, China
19. Hospital of Guangzhou, University of Chinese Medicine, Guangzhou, China
20. Evidence Based Medicine Center, Lanzhou University, Lanzhou, China
21. Child Health Evaluative Sciences, The Hospital for Sick Children Research Institute, Toronto, ON, Canada
22. Division of Gastroenterology, Hepatology and Nutrition, Hospital for Sick Children, Department of Paediatrics, University of Toronto, Toronto, ON, Canada
23. Department of Community Health and Epidemiology, Dalhousie University, Halifax, NS, Canada
24. Systematic Overviews through advancing Research Technology (SORT), Child Health Evaluative Sciences, The Hospital for Sick Children Research Institute, Toronto, ON, Canada

Corresponding author: Bradley C. Johnston
Department of Community Health and Epidemiology, Faculty of Medicine, Dalhousie University, 5790 University Avenue, B3H 1V7
Halifax, Nova Scotia, Canada
Email: bjohnston@dal.ca

Abstract

Background: Bacteriotherapy and modulation of intestinal microbiome by administering probiotics, prebiotics, or both have long been suggested to prevent morbidity and mortality in premature infants. Our objective was to assess the comparative effectiveness of different prophylactic strategies in a network meta-analysis (NMA) of randomized trials.

Methods: We searched MEDLINE, EMBASE, Science Citation Index Expanded, CINAHL, Scopus, Cochrane Central Register of Controlled Trials, BIOSIS Previews, and Google Scholar from inception up until December 1, 2018. All-cause mortality, severe necrotizing enterocolitis (NEC - Bell stage II or more), and culture proven sepsis were our *a priori* primary outcomes. We used a frequentist approach for NMA and assessed the certainty of evidence using the GRADE approach.

Findings: We included 79 trials involving 17,655 preterm infants. Multiple-strain probiotics alone proved the only intervention with moderate- or high-quality evidence of reduced all-cause mortality relative to placebo (OR = 0.67; 95% CI: 0.52, 0.87; high certainty; absolute risk reduction = 1.5%). Compared to placebo, multiple-strain probiotics (OR = 0.35; 95% CI: 0.26, 0.47; moderate certainty; absolute risk reduction = 4.1%), and single-strain probiotics alone (OR = 0.63; 95% CI: 0.46, 0.86; moderate certainty; absolute risk reduction = 2.3%) or in combination with lactoferrin (OR = 0.04; 95% CI: 0.01, 0.62; moderate certainty; absolute risk reduction = 6.2%) were among the interventions with moderate- or high-quality evidence that significantly reduced severe NEC. Among the intervention with moderate- or high-quality evidence relative to

placebo, single-strain probiotics combined with lactoferrin (OR = 0.27; 95% CI: 0.10, 0.72; moderate certainty; absolute risk reduction = 11.00%) or alone (OR = 0.80; 95% CI: 0.65, 0.99; moderate certainty; absolute risk reduction = 2.6%) and lactoferrin alone (OR = 0.44; 95% CI: 0.27, 0.74; moderate certainty; absolute risk reduction = 5.1%) demonstrated statistically significant reduction in culture proven late-onset sepsis.

Interpretation: Moderate-to-high certainty evidence demonstrates the superiority of single and multi-strain probiotics over alternative preventive treatments. Synbiotics (multiple-strain probiotics in combination with prebiotics) provide the largest reduction in morbidity and mortality in preterm infants but this is supported by only low-to-very low certainty evidence; thus, prioritizing synbiotics in future trials may provide important insights.

Research in context

Evidence before this study

Modulating gastrointestinal microbiome and bacteriotherapy for preventing morbidity and mortality of preterm infants has generated increasing scientific interest. Necrotizing enterocolitis and neonatal sepsis are the most important causes of morbidity and mortality in preterm infants. In spite of the recent improvements in neonatal intensive care, morbidity and mortality of preterm infants have remained unchanged.

Added value of this study

Our review is the most comprehensive systematic review of bacteriotherapy for preventing morbidity and mortality of preterm infants, including all available literature from English and non-English studies. We used state-of-art methods for comparative assessments of the effects of prebiotics, probiotics and synbiotics and used the most recent advancements in GRADE methodology to present the summary of results from the network meta-analysis, providing an innovative and transparent presentation of our findings.

Implications of all the available evidence

Our results demonstrate the superiority of single and multi-strain probiotics over alternative preventive treatments based on moderate to high certainty evidence. We also point to the fact that synbiotics provide the largest reduction in morbidity and mortality in preterm infants but only supported by low-to-very low certainty evidence; thus, pointing to the evidence gap that suggests prioritizing synbiotics in future trials.

Introduction

The human gastrointestinal tract is sterile at birth, but this complex ecosystem becomes rapidly colonized by microorganisms that facilitate digestion and modulate the immune system.[1] Bacteriotherapy that involves probiotics, prebiotics, or synbiotics (products that contain both probiotics and prebiotics) results in alteration of the gut microbial flora that may prevent or treat a number of diseases.[1,2]

NEC is a devastating inflammatory disorder of the intestine and among the leading causes of mortality and morbidity in neonatal intensive care units.[3,4] With an incidence ranging from 2% to 10% in infants born before 32 weeks gestation and 5% to 22% among those born with < 1000 g of weight, NEC has no effective treatment and surgical management is associated with high mortality.[4-6] Late-onset sepsis (LOS) also has a significant burden worldwide and despite preventive strategies such as antimicrobial stewardship, limited corticosteroid use, early enteral feeding, and hand hygiene, the incidence has, in recent years, remained stable in preterm infants.[7,8]

Bacteriotherapy and modulation of the intestinal microbiome have long been suggested as a potentially effective preventive strategy for both NEC and LOS.[9-12] Numerous systematic reviews and meta-analyses of randomized trials (RCTs) [7,8,13-17] and observational studies [2,18,19] have addressed the use of probiotics and prebiotics for NEC and LOS prevention in preterm infants. No study has, however, addressed the comparative effectiveness of probiotics, prebiotics, and synbiotics for preventing morbidity and mortality in preterm infants. Hence, we conducted a systematic review and

network meta-analysis of randomized trials addressing the effects of bacteriotherapy in preterm infants (gestational age < 37 weeks) on mortality and morbidity.

Methods

We registered out protocol with PROSPERO (CRD42018085566) and previously published a detailed protocol.[20]

Search strategy and selection criteria

Using the strategies reported in our published protocol and without language restrictions,[20] we searched MEDLINE, EMBASE, Science Citation Index Expanded and Social Sciences Citation Index, CINAHL, Scopus, ProQuest Dissertations and Theses database, Cochrane Central Register of Controlled Trials (CENTRAL), BIOSIS Previews, and Google Scholar from inception up until December 1, 2018 for relevant published RCTs (**Appendix 6.1**). We reviewed reference lists from eligible new trials and related reviews for additional eligible RCTs.

Two reviewers independently screened the titles and abstracts of all identified studies and, subsequently, independently assessed eligibility of the full-texts of potentially eligible studies. Reviewers resolved discrepancies through discussion, or, if needed, by adjudication from a third reviewer. Eligible trials used bacteriotherapy (probiotics – defined in this review as living bacteria, prebiotics – defined as non-digestible compounds including lactoferrin, inulin, galacto- or fructo-oligosaccharides, or synbiotics) for prevention of morbidity or mortality in preterm (gestational age < 37

weeks) and/or low birth weight (birth weight < 2500 grams) infants. **Appendix 6.2** provides further details on eligibility criteria.

Data abstraction and risk of bias assessment

Teams of reviewers extracted data and assessed risk of bias independently and in duplicate using a modified Cochrane risk of bias instrument[21,22] (see **Appendix 6.2** for details). Primary outcomes were Severe NEC - stage II or more based on Bell's criteria,[23,24] all-cause mortality, and culture-proven sepsis. Our secondary outcomes included NEC-related mortality, duration of hospitalization, weight 37 weeks' postnatal age or at discharge, time to establish full enteral feeds (days), and feed intolerance.

Data synthesis and statistical methods

For each direct paired comparison, we calculated odds risk (OR) and associated 95% confidence intervals (CIs) for dichotomous outcomes. For continuous outcomes, we calculated weighted mean differences with corresponding 95% CIs.

Initially, we performed conventional pairwise meta-analysis using a DerSimonian–Laird random-effects in STATA (StataCorp., Release 15.1. College Station, TX). We then performed network meta-analysis (NMA) to synthesize the available evidence from the entire network of trials using the methodology of multivariate meta-analysis assuming a common heterogeneity parameter.[25,26]

We evaluated the presence of incoherence (also called inconsistency) by comparing direct evidence with indirect evidence using the node splitting method.[27,28] We also confirmed the coherence assumption in the entire network using 'design-by-treatment' model (global

test) as described by Higgins et al.[29] We performed network meta-regression adjusting effect estimates for a priori defined covariates (gestational age, birth weight, percent infants exclusively fed by mothers' or donors' milk, percent delivered by C-section and risk of bias) at the study level assuming a common fixed coefficient across comparisons. We estimated ranking probabilities using the surface under the cumulative ranking curve (SUCRA), mean ranks, and rankograms. We used hierarchical cluster analysis to group treatments according to their ranking for the primary outcomes.[30,31]

Assessing certainty (quality) of the evidence

We rated the certainty of evidence for each network estimate using the GRADE (Grading of Recommendations, Assessment, Development, and Evaluation) approach.[32-34] For each direct comparison, according to established GRADE guidance, the starting point for certainty across the body of RCTs is high, but may be rated down based on limitations in risk of bias, imprecision, inconsistency, indirectness and publication bias.[32] We rated the certainty of the indirect evidence, with a focus on the dominant lowest order loop.[33] We rated the certainty of indirect evidence as the lowest certainty of the contributing direct comparisons.

Network estimate certainty started as the higher of the direct and indirect evidence; however, we considered the relative contribution of direct and indirect evidence to the network estimate when rating the certainty. We considered rating down the certainty in the network estimate if there was incoherence between the indirect and direct estimates or if there was imprecision (wide CIs) around the treatment effect.[33,34] **Appendix 6.2** provides further details on certainty assessments.

Summary of results

To optimize NMA results for interpretation, we applied a novel approach in which we categorized the interventions - from the most effective to the least effective - based on the effect estimates obtained from the NMA and their associated certainty of evidence. For each outcome, we created groups of interventions as follows: 1) The reference intervention (placebo) and interventions no different from placebo (i.e. 95% CI includes null value) which we refer to as "among the least effective"; 2) Interventions superior to placebo, but not superior to any other of the intervention(s) superior to placebo (which we call Category 2 and describe as "inferior to the most effective, but superior to the least effective"); and 3) Interventions that proved superior to at least one category 2 intervention (which we call "among the most effective"). We then divided all three categories into two groups: those with moderate or high certainty evidence relative to placebo, and those with low or very low certainty evidence relative to placebo.[35]

Results

Description of the evidence

We identified 7,562 records through our literature search, of which we included 96 publications from 87 studies. Eight studies, failed to report any of our target outcomes,[36-43] leaving 79 eligible RCTs involving 17,655 infants (**Appendix 6.3**). **Figure 6.1** presents the details of study selection process and reasons for exclusion.

Across the included trials, the median of the average weight was 1236.3 grams (interquartile range (IQR): 1095.5, 1472.5) and the median of the average gestational age

was 30.0 weeks (IQR: 28.8,31.3). **Appendix 6.4** summarizes the characteristics of the participants.

Of the 79 studies, 73 include 2-arms, three included 3-arms, and four include 4-arms. Single-strain and multiple-strain probiotics were the most common interventions (33 and 32 studies, respectively), followed by synbiotics (8 studies), lactoferrin (6 studies), and fructo- and/or galacto-oligosaccharides (5 studies). The majority of studies assessing multi-strain probiotics and synbiotics included both *Bifidobacterium* and *Lactobacillus* species (28 out of 32 and 6 out of 8, respectively). **Appendix 6.5** presents the characteristics of the treatments used and **Figure 6.2** and **Appendix 6.7** presents the networks of eligible comparisons for primary and secondary outcomes.

Of the 79 studies, 53 proved to be at low risk of bias in terms of allocation concealment and missing participant outcome data; 26 studies proved to be at high risk of bias for blinding of infants' parents/care givers and study personnel, while 33 studies proved at high risk of bias for masking of outcome assessments. **Appendix 6.6** provides details of risk of bias assessments.

All-cause mortality

All-cause mortality was reported in 63 studies involving 16,229 infants (**Figure 6.2A**). The design-by-treatment interaction model showed no evidence of incoherence in the network. We observed incoherence in 6 paired comparisons (**Appendix 6.10**). Of the 13 available direct comparisons, in 5 comparisons, 2 or more studies were available for

conventional pairwise meta-analysis in which the I^2 was zero in 3 comparisons and < 50% for the remaining comparisons (**Appendix 6.11, e-Table 5**).

Among the studies with high or moderate certainty evidence relative to placebo, only multiple strain probiotics reduced mortality (OR = 0.67; 95% CI: 0.52, 0.87; high certainty; risk difference (RD) = -1.54; 95% CI: -2.3, -0.58) (**Table 6.1, e-Table 1** and **Table 2**). Among the studies with low or very low certainty, multiple-strain probiotics in combination with fructo-oligosaccharides (Fos) and galacto-oligosaccharides (Gos) (OR = 0.05; 95% CI: 0.01, 0.41; very low certainty; RD = -4.95; 95% CI: -5.25, -2.73) and Fos alone (OR = 0.20; 95% CI: 0.04, 0.89; very low certainty; RD = -4.18; 95% CI: -5.03, -0.56) significantly decreased all-cause mortality when compared to placebo – the former significantly decreased mortality relative to the latter. No other statistically significant differences were identified between the remainder of the treatments and placebo comparisons (Table 1, **e-Table 1** and **Table 6.2**).

NEC stage II or higher

We included 70 RCTs with 15,271 infants involving 9 preventive therapies (**Figure 6.2B**). The design-by-treatment interaction model showed no evidence of incoherence. We observed incoherence in one closed loop of evidence involving Fos and single-strain probiotics (**Appendix 6.10**). Of the 14 direct comparisons, 7 involved 2 studies or more; of these, none showed evidence of statistical heterogeneity and the results of all direct comparisons proved similar to the NMA estimates (**Appendix 6.11, e-Table 6**).

NMA results provided evidence that probiotics and synbiotics significantly reduced severe NEC in preterm infants when compared to placebo. Among the studies with high or moderate certainty evidence relative to placebo, single-strain probiotics combined with lactoferrin (OR = 0.04; 95% CI: (0.01, 0.62; moderate certainty; RD = -6.15; 95% CI: -6.44, -1.32), multiple strain probiotics (OR = 0.35; 95% CI: 0.26, 0.47; moderate certainty; RD = -4.06; 95% CI: -4.66, -3.25), and single-strain probiotics (OR = 0.63; 95% CI: (0.46, 0.86; moderate certainty; RD = -2.30; 95% CI: -3.34, -0.93) significantly reduced severe NEC (**Table 6.1** and **e-Table 1**).

Among the studies with low or very low certainty, multiple-strain probiotics in combination with Fos (OR = 0.14; 95% CI: 0.04, 0.49; low certainty; RD = -5.48; 95% CI: -6.18, -3.05) or Fos and Gos (OR = 0.09; 95% CI: 0.02, 0.52; very low certainty; RD = -5.76; 95% CI: -6.33, -2.73) and single-strain probiotics combined with Fos (OR = 0.26; 95% CI: 0.09, 0.74; low certainty; RD = -4.62; 95% CI: -5.79, -1.45) significantly decreased the likelihood of all-cause mortality when compared to placebo (**e-Table 1**). **Tables 6.1 and 6.3** show the comparative effectiveness and the certainty for all pairwise comparisons.

Culture proven late-onset sepsis

Culture proven sepsis was reported in 60 RCTs involving 14,520 infants comparing 9 preventive therapies (**Figure 6.2C**) with 14 direct comparisons. Our analysis did not show statistical evidence of incoherence either in design-by-treatment interaction model (global test) or loop-specific models (**appendix 6.10**). Heterogeneity in two of the 6 direct

comparisons was substantial (I^2 = 53.8% and 76.8% for the comparisons of multiple-strain probiotics alone or in combination with Fos and Gos versus placebo, respectively). The results of NMA were, however, similar to the direct comparisons (**Appendix 6.11, e-table 7**).

Among the studies with high or moderate certainty evidence relative to placebo, single-strain probiotics combined with lactoferrin (OR = 0.27; 95% CI: 0.10, 0.72; moderate certainty; RD = -11.00; 95% CI: -14.24, -3.07) or alone (OR = 0.80; 95% CI: 0.65, 0.99; moderate certainty; RD = -2.57; 95% CI: -4.83, 0.00) and lactoferrin alone (OR = 0.44; 95% CI: 0.27, 0.74; moderate certainty; RD = -5.10; 95% CI: -8.42, -0.40) demonstrated statistically significant reduction in likelihood of culture proven late-onset sepsis (Table 1).

Among the studies with low or very low certainty, only multiple strain probiotics reduced mortality (OR = 0.77; 95% CI: 0.62, 0.95; low certainty; RD = -2.65; 95% CI: -4.69, -0.27) (**Table 6.1, e-Table 1,** and **Table 6.3**). No other statistically significant differences were identified between the remainder of the treatments and placebo comparisons (**Table 6.1** and **Table 6.3**).

Secondary outcomes

Appendix 6.8 and **6.9** provide the detailed results of NMA of secondary outcomes, and **Table 6.1** summarizes these results. NEC-related mortality was reported in 29 studies involving 6 interventions (**Appendix 6.7, e-Figure 1**). No intervention showed significant benefit when compared to placebo. **Table 6.2** provides the results of the NMA and **e-**

Table 8 (**Appendix 6.11**) provides the results of the direct estimates of effect and their associated certainty.

Of 79 included RCTs, 16 studies reported feed intolerance involving 8 interventions in 3 direct comparisons with 2 or more studies, all with low statistical heterogeneity (I^2 < 25%) (**Appendix 6.7, e-Figure 2**). We found no statistical evidence of incoherence either in design-by-treatment interaction model (global test) or loop-specific models (**Appendix 6.10**). Among the studies with high or moderate certainty evidence relative to placebo, only single-strain probiotics (OR = 0.47; 95% CI: 0.36, 0.61; moderate certainty; RD = -11.24; 95% CI: -14.18, -7.71) showed statistically significant reduction of feed intolerance (**Table 6.1**).

Among the studies with low or very low certainty multiple-strain probiotics alone (OR = 0.48; 95% CI: 0.30, 0.76; low certainty; RD = -10.56; 95% CI: -15.46, -3.82), or in combination with Fos and Gos (OR = 0.15; 95% CI: 0.06, 0.39; very low certainty; RD = -15.92; 95% CI: -20.31, -9.19) and Lactoferrin (OR = 0.21; 95% CI: 0.05, 0.79; low certainty; RD = -21.54; 95% CI: -26.49, -4.63) demonstrated statistically significant reduction in likelihood of feed intolerance (**Table 6.1, e-Tables 1 and 3**).

The 47 studies (9,586 infants) that reported time to reach full enteral feed involved 9 interventions in 6 direct comparisons with 2 or more studies (**e-Figure 3**). The design-by-treatment interaction model showed no evidence of incoherence in the network as a whole; however, we observed incoherence in four paired comparisons (**Appendix 6.10**).

Of 6 direct comparisons with 2 or more studies, 3 had substantial heterogeneity (I^2 > 50%) (**e-Table 10**).

Among the studies with high or moderate certainty evidence relative to placebo, only single-strain probiotics (MD = -1.60 days; 95% CI: -2.58, -0.62; moderate certainty) reduced mean number of days to reach full feed (Table 1). Among the studies with low or very low certainty multiple-strain probiotics alone (MD = -2.04 days; 95% CI: -3.10, -0.99; very low certainty) or in combination with Fos and Gos (MD = -3.31 days; 95% CI: -6.50, -0.12; very low certainty), and Fos (MD = -4.35 days; 95% CI: -8.23, -0.48; very low certainty) significantly decreased the number of days to reach full enteral feeding compared to placebo (**Table 6.1, e-Tables 2 and 3**).

The 40 studies (9,483 infants) that reported duration of hospital stay involved 9 interventions in 5 direct comparisons with 2 or more studies (**e-Figure 4**). We found no global incoherence but observed incoherence in four comparisons in the loop-specific model (**appendix 6.10**). Of the 5 direct comparisons involving 2 or more RCTs, 2 had substantial heterogeneity (I^2 > 50%) (**e-Table 11**). **e-Tables 2 and 4** provide the results of all pairwise comparisons. Only single-strain and multi-strain probiotics were statistically more effective than placebo in reducing the duration of hospitalization (MD = -3.58 days; 95% CI: -5.88, -1.27; high certainty and MD = -2.68 days; 95% CI: -5.15, -0.21; low certainty, respectively) (**Table 6.1**).

The 13 studies that reported weight at 37 weeks' postnatal age or at discharge involved 7 interventions in 3 direct comparisons with 2 or more studies (**e-Figure 5**). Our analysis

showed no intervention had any statistically significant benefit compared to placebo. **e-Table 6.4** shows the NMA results (**Appendix 6.9**) and **e-Table 12** (**Appendix 6.11**) shows the direct estimates of effect with their associated certainty. We found no statistical evidence of incoherence either in design-by-treatment interaction model (global test) or loop-specific model (**Appendix 6.10**). The magnitude and direction of the effect for all NMA estimates of the effects were comparable to those of direct estimates in all secondary outcomes.

Additional analysis

Appendix 6.12 provides details of rankings and SUCRA values for all outcomes. **e-Figures 22-24** (**Appendix 6.13**) demonstrate the results of hierarchical cluster analysis to group preventive treatments based on their ranking for the primary outcomes. According to all possible cluster rankings, multiple strain probiotics and synbiotics are almost always more effective than single-strain-probiotics or prebiotics alone; this is consistent with results reported in **Table 6.1**, although moderate-to-high certainty evidence supports the benefits of multiple-strain probiotics, and only low-to-very low certainty evidence supports synbiotics benefits.

We performed network meta-regression to explore the impact of a priori defined effect modifiers (risk of bias, birth weight, gestational age, percent infants fed by breast milk, and percent delivered by C-section). Due to small number studies that reported secondary outcomes, we limited this only to the primary outcomes. In all models, none of the coefficients proved statistically significant. The analysis of comparison-adjusted funnel

plots in the network showed no evidence of small study effect for any of the outcomes (**Appendix 6.14**).

Discussion

In this systematic review and NMA comparing the effectiveness of different bacteriotherapy regimens for prevention of mortality and morbidity in preterm infants we found, across a number of outcomes, several interventions more effective than placebo. Moderate-to-high certainty evidence indicates that multi-strain probiotics are best for the prevention of all-cause mortality and stage II NEC, while moderate certainty evidence indicates that single-strain probiotics prevent all-cause and NEC-related mortality, stage II NEC, and culture proven sepsis (**Table 1**). Prebiotics alone likely have little to no benefit. Although synbiotics (single- or multiple-strain probiotics combined with lactoferrin or Fos and Gos) showed the largest relative and absolute risk reductions across most outcomes; however, only low-to-very low certainty evidence supports the benefits of these interventions (**Table 1**). We did not observe any effect modification for birth weight, gestational age, feeding with breast milk, or delivery type.

Our review has a number of strengths. It is the most comprehensive systematic review on this topic to date, including all available literature from English and non-English RCTs for comparative assessments of the effects of prebiotics, probiotics and synbiotics. The review is based on analyses using sophisticated statistical models that considered both NMA effect estimates and probability rankings. The review uses the GRADE approach for assessing the certainty in the NMA effect estimates and provides an innovative, transparent and simple presentation of our findings. This presentation captures, in a single

table the relative performance of each treatment on each outcome, categorized by the certainty of the evidence (**Table 6.1**).

Our study also has limitations, the most important of which is the small number of studies directly comparing prebiotics and synbiotics. In addition, for all comparisons, few trials compared active treatments to one another (rather than to placebo) (**Figure 6.2**). Variability in probiotic composition (i.e. variability in strain, species, and doses), which we were not able to explore further due to small number direct comparisons, makes it difficult to identify the most effective probiotic or synbiotic combination for clinical use. In addition to small numbers of patients and events in many comparisons, certainty of evidence was sometimes compromised by differences in results between direct and indirect comparisons (incoherence) and intransitivity of indirect comparisons. Our review is limited in that it did not address the relative impact of different strains of probiotics on the outcomes of interest.

Recently, two strain-specific meta-analyses and a network meta-analysis addressed probiotic effectiveness and their results are consistent with our findings for probiotics. The reviews also addressed possible variability in effectiveness of probiotic strains/species. Athalye-Jape et al looked at the effects of *L. reuteri* (DSM 17938)[15] and *B. breve* (M-16V)[44] in RCTs and non-RCTs involving preterm infants and found no significant benefits for *B. breve* on severe NEC, late-onset sepsis, all-cause mortality, and time to reach full enteral feedings. By contrast, the investigators reported significant reductions in LOS, time to reach full feedings, and duration of hospitalization as well as non-significant reductions in the incidence of severe NEC and all-cause mortality with *L.*

reuteri. An NMA addressing strain-specific effects of probiotics in 51 RCTs provided evidence that a combination of strains (multiple-strain probiotics) are usually better than any single-strain probiotics, but the paucity of studies addressing particular strains or combinations of strains limited inferences regarding comparative effectiveness.[45]

Probiotics preventive effects are believed to be strain- and/or species-specific and multifactorial.[46-48] Bifidobacterial species, specifically *B. infantis* and *B. bifidum*, which are known to produce bacteriocins and secrete molecules with anti-inflammatory properties that are different than *B. breve* in the in vitro setting.[44,48] While *Lactobacillus* species also produce bacteriocins, modulate the immune system, and can impact intestinal motility and epithelial barrier function; these effects do not apply to all species of *Lactobacillus*.[15,48] These differences could be the reason why multi-strain probiotics show better effects than single-strain probiotics. Enhanced preventive effects of multiple-strain probiotics compared to single-stain formulations might be also due to increased diversity of the intestinal microbiota or simply the consequence of a higher probiotic dose.[13]

Our results based on 79 RCTs indicate that multi-strain probiotics (specifically those containing *Lactobacillus* together with *Bifidobacterium* species) and single-strain probiotics (specific strains from *Lactobacillus* or *Bifidobacterium* species) provide the largest benefit in reducing morbidity and mortality in preterm infants. In this study, we did not address which probiotic strain(s)/species are more effective, but we are currently working on the analysis of that network.

The underlying mechanism(s) of action for the beneficial effects of synbiotics remains uncertain. Our findings suggest a potential additive or synergistic interaction when combining prebiotics and probiotics.[46] One likely explanation is that prebiotics can act as food source for probiotic colonization. In healthy term infants, the capacity of bifidobacteria strains to consume human milk oligosaccharides (HMOs) present in human breast milk as substrate is considered a primary mechanism promoting bacterial colonization.[48,49] Seven of eight synbiotic interventions tested in RCTs consisted of *Bifidobacterium* species combined with prebiotics, including Fos or Gos.

In addition, the majority of the studies (6 RCTs) used multiple-strain probiotics in the synbiotic combinations. Considering that our results showed that multi-strain probiotics are more effective than single-strain probiotics, this approach also might provide an explanation for the larger preventive effects of synbiotics. Although the magnitude of effect in synbiotics was greater than either probiotics or prebiotics alone, the certainty of evidence for synbiotics was low to very low, strongly suggesting current use of interventions with higher levels of certainty.

The most severe harm associated with the use of probiotics for prevention of morbidity and mortality in preterm infants is sepsis caused by administered probiotic(s) due to translocation across the intestinal epithelial barrier.[48,50] This is reported to be extremely rare [13,50] with none of the studies included in our review reporting cases of probiotic-associated sepsis, which is consistent with what other recent systematic reviews have reported.[8,13,15,17,18,44,45] There are two other important safety-related issues that has been discussed abundantly: regulatory issues of commercially available probiotic products and

cross-contamination.[48,50] Probiotics are being considered as dietary supplements in United States and many other countries and hence are not being regulated with the same scrutiny as drugs for their intended use before marketing.[51-53] Cross-contamination of administered probiotics to infants (i.e. receipt of the incorrect microorganism) was noted in two trials. Costeloe et al reported that half the infants in placebo arms were colonized with the study probiotic species[54] and Underwood et al described identifying *B. infantis* in the stools of infants receiving *B. lactis*.[55,56] While cross-contamination may have led to misclassification between the two trial arms, this would typically lead to a more conservative effect estimate.

Conclusion

Moderate-to-high certainty evidence supports the superiority of single and multi-strain probiotics over alternative preventive treatments. Synbiotics (multiple-strain probiotics in combination with prebiotics) provide the largest reduction in morbidity and mortality in preterm infants but this is supported by only low-to-very low certainty evidence. There are few commercially available synbiotic and probiotic products and, unfortunately, there are very few large multi-center RCTs on the effect of probiotics and no large RCT investigating the effects of synbiotics in preterm babies. While certain probiotics with evidence of efficacy and safety are now frequently used in clinical practice, synbiotics compared to multiple-strain probiotics should be prioritised in future randomized clinical trials.

Funding: Mitacs in partnership with Nestlé Canada (Mitacs is a national, not-for-profit organization that has designed and delivered research and training programs in Canada working with universities, companies, and both federal and provincial governments).

Acknowledgements

We would like to thank Andrea Papamandjaris from Nestlé Canada Inc. for her support for Mitacs award application and proofreading the manuscript. We also thank Lyubov Lytvyn for helping with the title and abstract screening. PMS is the recipient of a Canada Research Chair in Gastrointestinal Disease.

References

1. Patel R, DuPont HL. New approaches for bacteriotherapy: prebiotics, new-generation probiotics, and synbiotics. Clinical infectious diseases 2015; 60 Suppl 2: S108-21.
2. Dermyshi E, Wang Y, Yan C, et al. The "Golden Age" of Probiotics: A Systematic Review and Meta-Analysis of Randomized and Observational Studies in Preterm Infants. Neonatology 2017; 112(1): 9-23.
3. Eaton S, Rees CM, Hall NJ. Current Research on the Epidemiology, Pathogenesis, and Management of Necrotizing Enterocolitis. Neonatology 2017; 111(4): 423-30.
4. Battersby C, Santhalingam T, Costeloe K, Modi N. Incidence of neonatal necrotising enterocolitis in high-income countries: a systematic review. Archives of disease in childhood Fetal and neonatal edition 2018; 103(2): F182-f9.
5. Yee WH, Soraisham AS, Shah VS, et al. Incidence and timing of presentation of necrotizing enterocolitis in preterm infants. Pediatrics 2012; 129(2): e298-304.
6. Lin PW, Stoll BJ. Necrotising enterocolitis. Lancet (London, England) 2006; 368(9543): 1271-83.
7. Sharma D, Shastri S. Lactoferrin and neonatology - role in neonatal sepsis and necrotizing enterocolitis: present, past and future. The journal of maternal-fetal & neonatal medicine 2016; 29(5): 763-70.
8. Rao SC, Athalye-Jape GK, Deshpande GC, Simmer KN, Patole SK. Probiotic Supplementation and Late-Onset Sepsis in Preterm Infants: A Meta-analysis. Pediatrics 2016; 137(3): e20153684.
9. Caplan MS, Jilling T. New concepts in necrotizing enterocolitis. Current opinion in pediatrics 2001; 13(2): 111-5.
10. Deshpande G, Rao S, Patole S. Probiotics for prevention of necrotising enterocolitis in preterm neonates with very low birthweight: a systematic review of randomised controlled trials. Lancet (London, England) 2007; 369(9573): 1614-20.
11. Warner BB, Tarr PI. Necrotizing enterocolitis and preterm infant gut bacteria. Seminars in fetal & neonatal medicine 2016; 21(6): 394-9.
12. Neu J, Pammi M. Pathogenesis of NEC: Impact of an altered intestinal microbiome. Seminars in perinatology 2017; 41(1): 29-35.
13. Chang HY, Chen JH, Chang JH, Lin HC, Lin CY, Peng CC. Multiple strains probiotics appear to be the most effective probiotics in the prevention of necrotizing enterocolitis and mortality: An updated meta-analysis. PloS one 2017; 12(2): e0171579.
14. Sawh SC, Deshpande S, Jansen S, Reynaert CJ, Jones PM. Prevention of necrotizing enterocolitis with probiotics: a systematic review and meta-analysis. PeerJ 2016; 4: e2429.

15. Athalye-Jape G, Rao S, Patole S. Lactobacillus reuteri DSM 17938 as a Probiotic for Preterm Neonates: A Strain-Specific Systematic Review. JPEN Journal of parenteral and enteral nutrition 2016; 40(6): 783-94.
16. Pammi M, Abrams SA. Oral lactoferrin for the prevention of sepsis and necrotizing enterocolitis in preterm infants. The Cochrane database of systematic reviews 2015; (2): Cd007137.
17. AlFaleh K, Anabrees J. Probiotics for prevention of necrotizing enterocolitis in preterm infants. Evid Based Child Health: A Cochrane Review Journal 2014; 9(3): 584-671.
18. Olsen R, Greisen G, Schroder M, Brok J. Prophylactic Probiotics for Preterm Infants: A Systematic Review and Meta-Analysis of Observational Studies. Neonatology 2016; 109(2): 105-12.
19. Schulzke SM, Deshpande GC, Patole SK. Neurodevelopmental outcomes of very low-birth-weight infants with necrotizing enterocolitis: a systematic review of observational studies. Archives of pediatrics & adolescent medicine 2007; 161(6): 583-90.
20. Sadeghirad B, Florez ID, Chang Y, et al. Comparative effectiveness of prophylactic therapies for necrotizing enterocolitis in preterm infants: Protocol for a network meta-analysis of randomized trials. Int J Prev Med 2018; 9: 83.
21. Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011; 343: d5928.
22. Akl EA, Sun X, Busse JW, et al. Specific instructions for estimating unclearly reported blinding status in randomized trials were reliable and valid. Journal of clinical epidemiology 2012; 65(3): 262-7.
23. Bell MJ, Ternberg JL, Feigin RD, et al. Neonatal necrotizing enterocolitis. Therapeutic decisions based upon clinical staging. Ann Surg 1978; 187(1): 1-7.
24. Walsh M, Kliegman R. Necrotizing enterocolitis: treatment based on staging criteria. Pediatr Clin North Am 1986; 33(1): 179-201.
25. White IR. Network meta-analysis. The Stata Journal 2015; 15(4): 951–85.
26. White IR, Barrett JK, Jackson D, Higgins JP. Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. Research synthesis methods 2012; 3(2): 111-25.
27. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003; 327(7414): 557-60.
28. Lu G, Ades AE. Assessing evidence inconsistency in mixed treatment comparisons. J Am Stat Assoc 2006; 101(474): 447-59.
29. Higgins JP, Jackson D, Barrett JK, Lu G, Ades AE, White IR. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. Research synthesis methods 2012; 3(2): 98-110.

30. Chaimani A, Higgins JP, Mavridis D, Spyridonos P, Salanti G. Graphical tools for network meta-analysis in STATA. PloS one 2013; 8(10): e76654.
31. Chaimani A, Salanti G. Visualizing assumptions and results in network meta-analysis: The network graphs package. Stata Journal 2015; 15(4): 905-50.
32. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008; 336(7650): 924-6.
33. Puhan MA, Schunemann HJ, Murad MH, et al. A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. BMJ 2014; 349: g5630.
34. Brignardello-Petersen R, Bonner A, Alexander PE, et al. Advances in the GRADE approach to rate the certainty in estimates from a network meta-analysis. Journal of clinical epidemiology 2018; 93: 36-44.
35. Florez ID, Veroniki AA, Al Khalifah R, et al. Comparative effectiveness and safety of interventions for acute diarrhea and gastroenteritis in children: A systematic review and network meta-analysis. PloS one 2018; 13(12): e0207701.
36. Li Y, Shimizu T, Hosaka A, Kaneko N, Ohtsuka Y, Yamashiro Y. Effects of bifidobacterium breve supplementation on intestinal flora of low birth weight infants. Pediatrics international 2004; 46(5): 509-15.
37. Coleta E, Gheonea M, Sarbu M. Oral supplementation with probiotics in premature infants-a randomised clinical trial. Intensive Care Medicine 2013; 39 (24th Annual Meeting of the European Society of Paediatric and Neonatal Intensive Care): S113.
38. Koksal N, Varal I, Ozkan H, Bagci O, Dotan P. Effect of probiotic support on feeding intolerance and mortality at preterm infants. Journal of Perinatal Medicine Conference: 12th World Congress of Perinatal Medicine 2015; 43(no pagination).
39. Kapiki AC, C.;Oikonomidou, C.;Triantafylldou, A.;Loukatou, E.;Pertrohilou, V. The effect of a fructo-oligosaccharide supplemented formula on gut flora of preterm infants. Early Human Development 2007; 83(5): 335-9.
40. Wang C, Shoji H, Sato H, Nagata S, Ohtsuka Y, Shimizu T, Yamashiro Y. Effects of oral administration of bifidobacterium breve on fecal lactic acid and short-chain fatty acids in low birth weight infants. J Pediatr Gastroenterol Nutr 2007; 44(2): 252-7.
41. Zeber-Lubecka N, Kulecka M, Ambrozkiewicz F, et al. Effect of Saccharomyces boulardii and Mode of Delivery on the Early Development of the Gut Microbial Community in Preterm Infants. PLoS ONE 2016; 11(2).
42. Indrio F, Riezzo G, Raimondi F, Bisceglia M, Cavallo L, Francavilla R. The effects of probiotics on feeding tolerance, bowel habits, and gastrointestinal motility in preterm newborns. The Journal of pediatrics 2008; 152(6): 801-6.

43. Partty A, Luoto R, Kalliomaki M, Salminen S, Isolauri E. Effects of early prebiotic and probiotic supplementation on development of gut microbiota and fussing and crying in preterm infants: a randomized, double-blind, placebo-controlled trial. The Journal of pediatrics 2013; 163(5): 1272-7.e1-2.
44. Athalye-Jape G, Rao S, Simmer K, Patole S. Bifidobacterium breve M-16V as a Probiotic for Preterm Infants: A Strain-Specific Systematic Review. JPEN Journal of parenteral and enteral nutrition 2018; 42(4): 677-88.
45. van den Akker CHP, van Goudoever JB, Szajewska H, et al. Probiotics for Preterm Infants: A Strain-Specific Systematic Review and Network Meta-analysis. Journal of pediatric gastroenterology and nutrition 2018; 67(1): 103-22.
46. Johnson-Henry KC, Abrahamsson TR, Wu RY, Sherman PM. Probiotics, Prebiotics, and Synbiotics for the Prevention of Necrotizing Enterocolitis. Advances in nutrition (Bethesda, Md) 2016; 7(5): 928-37.
47. Sherman PM, Ossa JC, Johnson-Henry K. Unraveling mechanisms of action of probiotics. Nutrition in clinical practice 2009; 24(1): 10-4.
48. Underwood MA. Impact of probiotics on necrotizing enterocolitis. Seminars in perinatology 2017; 41(1): 41-51.
49. Yu ZT, Chen C, Newburg DS. Utilization of major fucosylated and sialylated human milk oligosaccharides by isolated human gut microbes. Glycobiology 2013; 23(11): 1281-92.
50. Doron S, Snydman DR. Risk and safety of probiotics. Clinical infectious diseases 2015; 60 Suppl 2: S129-34.
51. de Simone C. The Unregulated Probiotic Market. Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association 2018.
52. Venugopalan V, Shriner KA, Wong-Beringer A. Regulatory oversight and safety of probiotic use. Emerging infectious diseases 2010; 16(11): 1661-5.
53. Hoffmann DE, Fraser CM, Palumbo F, Ravel J, Rowthorn V, Schwartz J. Probiotics: achieving a better regulatory fit. Food and drug law journal 2014; 69(2): 237-72, ii.
54. Costeloe K, Hardy P, Juszczak E, Wilks M, Millar MR. Bifidobacterium breve BBG-001 in very preterm infants: A randomised controlled phase 3 trial. The Lancet 2016; 387(10019): 649-60.
55. Underwood MA, Salzman NH, Bennett SH, et al. A Randomized Placebo-controlled Comparison of 2 Prebiotic/Probiotic Combinations in Preterm Infants: Impact on Weight Gain, Intestinal Microbiota, and Fecal Short-chain Fatty Acids. Journal of pediatric gastroenterology and nutrition 2009; 48(2): 216-25.
56. Underwood MA, Kalanetra KM, Bokulich NA, et al. A comparison of two probiotic strains of bifidobacteria in premature infants. J Pediatr 2013; 163(6): 1585-91.e9.

Table 6. 1: Network meta-analysis results sorted based on GRADE certainty of evidence and treatments effectiveness for the comparisons of active treatments vs. placebo for primary and secondary outcomes.

	All-cause Mortality OR (95% CI)	NEC (stage ≥ II) OR (95% CI)	Culture proven sepsis OR (95% CI)	NEC-related mortality OR (95% CI)	Feed intolerance OR (95% CI)	Reduction in days to reach full feed MD (95% CI)	Reduction in days of hospitalization MD (95% CI)	Weight (gr) MD (95% CI)
Multiple-strain Probiotic	**0.67 (0.52, 0.87)**	**0.35 (0.26, 0.47)**	**0.77 (0.62, 0.95)**	0.52 (0.26, 1.01)	**0.48 (0.30, 0.76)**	**-2.04 (-3.10, -0.99)**	**-2.68 (-5.15, -0.21)**	51.88 (-36.11, 139.86)
Single-strain Probiotic	0.83 (0.67, 1.04)	**0.63 (0.46, 0.86)**	**0.80 (0.65, 0.99)**	0.71 (0.39, 1.27)	**0.47 (0.36, 0.61)**	**-1.60 (-2.58, -0.62)**	**-3.58 (-5.88, -1.27)**	17.86 (-152.87, 188.59)
Multiple-strain Probiotic + Fos & Gos	**0.05 (0.01, 0.41)**	**0.09 (0.02, 0.52)**	0.73 (0.32, 1.64)	0.17 (0.02, 1.56)	**0.15 (0.06, 0.39)**	**-3.31 (-6.50, -0.12)**	-9.00 (-22.11, 4.11)	-
Single-strain Probiotic + Lactoferrin	0.63 (0.28, 1.41)	**0.04 (0.01, 0.62)**	**0.27 (0.10, 0.72)**	-	-	-0.97 (-4.33, 2.39)	1.98 (-5.43, 9.39)	-
Lactoferrin	0.64 (0.33, 1.22)	0.43 (0.18, 1.00)	**0.44 (0.27, 0.74)**	1.13 (0.02, 59.5)	**0.21 (0.05, 0.79)**	-1.44 (-3.95, 1.07)	1.75 (-4.02, 7.53)	33.77 (-182.31, 249.86)
Fos	**0.20 (0.04, 0.89)**	0.81 (0.36, 1.83)	0.85 (0.33, 2.19)	0.34 (0.01, 7.77)	0.37 (0.10, 1.39)	**-4.35 (-8.23, -0.48)**	-6.19 (-14.01, 1.62)	27.32 (-184.43, 239.07)
Single-strain Probiotic + Fos	0.30 (0.08, 1.07)	**0.26 (0.09, 0.74)**	0.67 (0.25, 1.80)	1.71 (0.25, 11.8)	0.48 (0.17, 1.35)	-1.35 (-5.52, 2.82)	-2.19 (-10.08, 5.69)	36.32 (-170.76, 243.40)
Multiple-strain Probiotic + Fos	0.94 (0.38, 2.32)	**0.14 (0.04, 0.49)**	0.62 (0.30, 1.28)	-	0.13 (0.01, 2.55)	0.46 (-2.63, 3.55)	0.45 (-4.14, 5.05)	-
Fos & Gos	0.95 (0.21, 4.42)	1.73 (0.67, 4.45)	0.58 (0.29, 1.14)	-	1.00 (0.36, 2.78)	-1.14 (-3.57, 1.28)	-2.86 (-10.30, 4.58)	59.34 (-119.34, 238.02)

Footnote: OR = odds ratio; MD = mean difference; Fos = Fructo-oligosaccharides, Gos = galacto-oligosaccharides. Results are the mean difference, or odds ratio, and associated 95% confidence intervals (95% CIs) between the intervention and placebo from the network meta-analysis. Mean difference values < 0 indicates the treatment is more effective than placebo. An OR < 1 indicates the treatment is superior to placebo; Underlined numbers in bold represent statistically significant results.

Table legends and description of color gradients:

	Statistically significant difference with placebo and at least one other tx	Statistically significant difference with placebo	Statistically no difference with placebo
High or moderate certainty evidence	**Among the most effective**	**Inferior to the most effective, but superior to placebo**	**No more effective than placebo**
Low or very low certainty evidence	**May be among the most effective**	**May be inferior to the most effective, but superior to placebo**	**May be no more effective than placebo**

Table 6. 2: Network meta-analysis results for all-cause mortality (top half) and NEC-related mortality (bottom half).

Placebo	0.63 (0.28,1.41)	0.30 (0.08,1.07)	0.83 (0.67,1.04)	**0.67 (0.52,0.87)**	**0.05 (0.01,0.41)**	0.94 (0.38,2.32)	0.64 (0.33,1.22)	0.95 (0.21,4.42)	**0.20 (0.04,0.89)**
-	**SinglePrb & Lactoferrin**	0.47 (0.11,2.13)	1.32 (0.57,3.03)	1.06 (0.46,2.45)	**0.08 (0.01,0.75)**	1.49 (0.45,4.97)	1.01 (0.39,2.56)	1.51 (0.27,8.50)	0.31 (0.06,1.71)
1.71 (0.25,11.78)	-	**SinglePrb & Fos**	2.78 (0.77,9.97)	2.23 (0.61,8.13)	0.17 (0.01,1.95)	3.14 (0.66,14.87)	2.12 (0.51,8.81)	3.17 (0.43,23.2)	0.66 (0.11,4.04)
0.71 (0.39,1.27)	-	0.41 (0.06,2.70)	**SinglePrb**	0.80 (0.57,1.13)	**0.06 (0.01,0.50)**	1.13 (0.45,2.86)	0.76 (0.38,1.52)	1.14 (0.24,5.39)	0.24 (0.05,1.07)
0.52 (0.26,1.01)	-	0.30 (0.04,2.32)	0.73 (0.30,1.80)	**MultiPrb**	**0.07 (0.01,0.62)**	1.41 (0.55,3.59)	0.95 (0.47,1.91)	1.42 (0.30,6.74)	0.30 (0.06,1.35)
0.17 (0.02,1.56)	-	0.10 (0.01,1.88)	0.25 (0.03,2.39)	0.34 (0.03,3.34)	**MultiPrb & Fos & Gos**	**18.92 (1.91,187.3)**	**12.75 (1.40,115.9)**	**19.11 (1.41,259.2)**	3.97 (0.30,52.8)
-	-	-	-	-	-	**MultiPrb & Fos**	0.67 (0.22,2.04)	1.01 (0.17,5.97)	0.21 (0.04,1.21)
1.13 (0.02,59.49)	-	0.66 (0.01,54.12)	1.61 (0.03,88.15)	2.19 (0.04,122.0)	6.51 (0.07,602.5)	-	**Lactoferrin**	1.50 (0.28,7.93)	0.31 (0.06,1.60)
-	-	-	-	-	-	-	-	**Fos & Gos**	0.21 (0.02,1.77)
0.34 (0.01,7.77)	-	0.20 (0.01,4.14)	0.48 (0.02,10.71)	0.65 (0.03,16.16)	1.93 (0.04,88.96)	-	0.30 (0.01,46.47)	-	**Fos**

Footnote: Results are Odds Ratio (95% CIs) from the network meta-analysis. Odds ratios < 1 means the treatment in bottom right is better. Numbers in bold represent statistically significant results. Colors represent the certainty in evidence for each pairwise comparison. Fos = Fructo-oligosaccharides, Gos = galacto-oligosaccharides; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.

High certainty evidence	**Moderate certainty evidence**	**Low certainty evidence**	**Very low certainty evidence**

Table 6. 3: Network meta-analysis results for NEC stage II or more (top half) and culture proven sepsis (bottom half).

Placebo	**0.04 (0.00,0.62)**	**0.26 (0.09,0.74)**	**0.63 (0.46,0.86)**	**0.35 (0.26,0.47)**	**0.09 (0.02,0.52)**	**0.14 (0.04,0.49)**	0.43 (0.18,1.00)	1.73 (0.67,4.45)	0.81 (0.36,1.83)
0.27 (0.10,0.72)	**SinglePrb & Lactoferrin**	7.13 (0.35,146.0)	**17.32 (1.00,299.8)**	9.60 (0.56,165.9)	2.55 (0.09,70.3)	3.72 (0.17,83.2)	11.65 (0.65,208.5)	**47.36 (2.39,939.5)**	**22.16 (1.16,422.8)**
0.67 (0.25,1.80)	2.50 (0.61,10.17)	**SinglePrb & Fos**	2.43 (0.81,7.26)	1.35 (0.45,3.99)	0.36 (0.05,2.68)	0.52 (0.11,2.48)	1.63 (0.42,6.29)	**6.64 (1.62,27.21)**	3.11 (0.99,9.75)
0.80 (0.65,0.99)	**3.00 (1.09,8.31)**	1.20 (0.44,3.26)	**SinglePrb**	**0.55 (0.36,0.85)**	**0.15 (0.03,0.85)**	**0.21 (0.06,0.80)**	0.67 (0.27,1.67)	**2.73 (1.01,7.37)**	1.28 (0.53,3.08)
0.77 (0.62,0.95)	**2.88 (1.04,7.95)**	1.15 (0.42,3.18)	0.96 (0.71,1.29)	**MultiPrb**	0.27 (0.05,1.53)	0.39 (0.10,1.43)	1.21 (0.49,2.99)	**4.93 (1.83,13.29)**	2.31 (0.97,5.50)
0.73 (0.32,1.64)	2.73 (0.76,9.82)	1.09 (0.30,3.94)	0.91 (0.39,2.11)	0.95 (0.41,2.19)	**MultiPrb & Fos & Gos**	1.46 (0.17,12.41)	4.56 (0.67,31.16)	**18.55 (2.60,132.29)**	**8.68 (1.29,58.35)**
0.62 (0.30,1.28)	2.33 (0.68,7.95)	0.93 (0.27,3.19)	0.77 (0.36,1.66)	0.81 (0.38,1.72)	0.85 (0.29,2.53)	**MultiPrb & Fos**	3.13 (0.68,14.49)	**12.72 (2.60,62.20)**	**5.95 (1.34,26.47)**
0.44 (0.27,0.74)	1.66 (0.59,4.67)	0.67 (0.22,2.04)	**0.55 (0.32,0.97)**	0.58 (0.33,1.01)	0.61 (0.23,1.59)	0.72 (0.29,1.74)	**Lactoferrin**	**4.07 (1.14,14.54)**	1.90 (0.59,6.19)
0.58 (0.29,1.14)	2.16 (0.65,7.20)	0.86 (0.26,2.88)	0.72 (0.35,1.47)	0.75 (0.37,1.53)	0.79 (0.27,2.28)	0.93 (0.34,2.51)	1.30 (0.55,3.04)	**Fos & Gos**	0.47 (0.13,1.63)
0.85 (0.33,2.19)	3.19 (0.81,12.56)	1.28 (0.41,3.96)	1.06 (0.41,2.74)	1.11 (0.42,2.91)	1.17 (0.34,4.06)	1.37 (0.42,4.51)	1.92 (0.66,5.61)	1.48 (0.46,4.74)	**Fos**

Footnote: Results are Odds Ratio (95% CIs) from the network meta-analysis. Odds ratios < 1 means the treatment in bottom right is better. Numbers in bold represent statistically significant results. Colors represent the certainty in evidence for each pairwise comparison. Fos = Fructo-oligosaccharides, Gos = galacto-oligosaccharides; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.

High certainty evidence	Moderate certainty evidence	Low certainty evidence	Very low certainty evidence

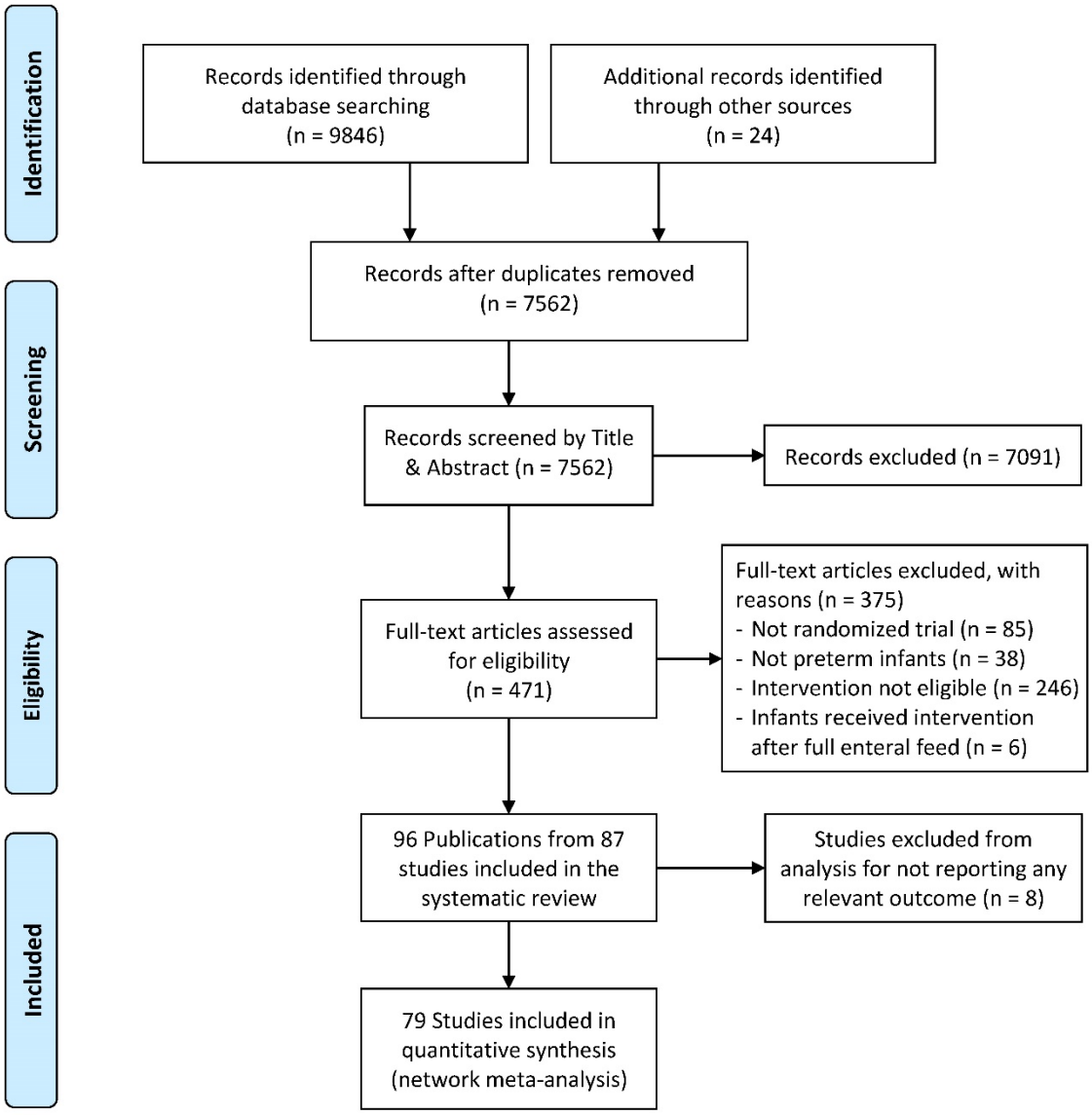


Figure 6. 1: Flow diagram for study selection

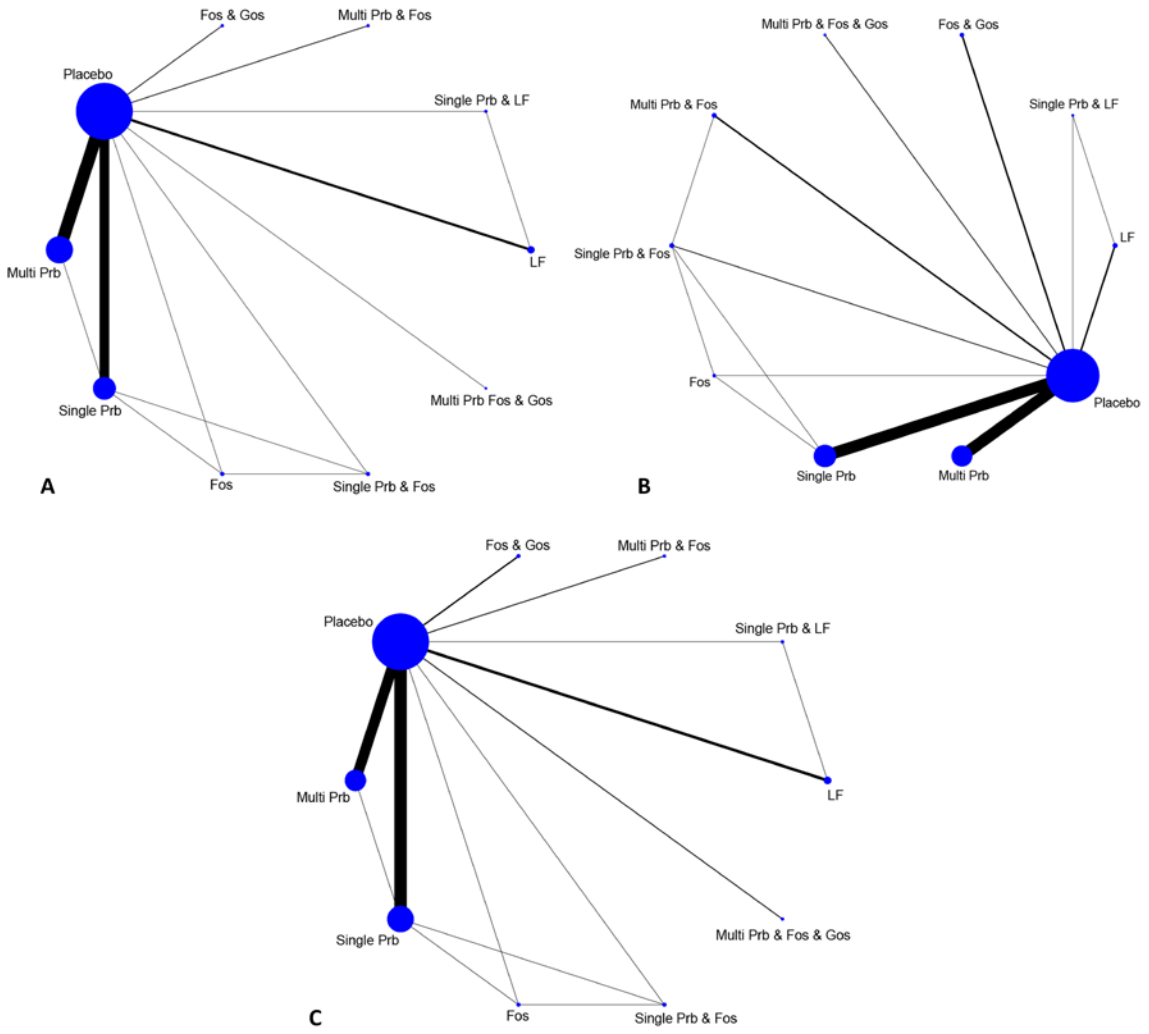


Figure 6. 2: Network of eligible comparisons for primary outcomes. (A) All-cause mortality, (B) NEC stage II or more, and (C) Culture proven sepsis.

The size of the node corresponds to the number of infants randomized to that intervention. The interventions directly compared are linked with a line; the thickness of the line corresponds to the number of studies that assessed the comparison. Fos = Fructo-oligosaccharides, Gos = galacto-oligosaccharides; LF = Lactoferrin; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.

Appendices

Appendix 6. 1. Search strategies

MEDLINE – from 1946

# ▲	Searches
1	enterocolitis, necrotizing/
2	(necroti?ing enterocolitis or enterocolitis necroticans or nec or typhlitis or pneumatosis intestinalis).tw.
3	1 or 2
4	randomized controlled trial.pt.
5	controlled clinical trial.pt.
6	randomized.ab.
7	placebo.ab.
8	drug therapy.fs.
9	randomly.ab.
10	trial.ab.
11	groups.ab.
12	or/4-11
13	exp animals/ not humans.sh.
14	12 not 13
15	3 and 14

Embase Classic and Embase – from 1947

# ▲	Searches
1	enterocolitis, necrotizing/
2	(necroti?ing enterocolitis or enterocolitis necroticans or nec or typhlitis or pneumatosis intestinalis).tw.
3	1 or 2
4	crossover-procedure/
5	double-blind procedure/
6	randomized controlled trial/
7	single-blind procedure/
8	((doubl* adj blind*) or (singl* adj blind*) or allocat* or assign* or cross over* or crossover* or factorial* or placebo* or random* or volunteer*).tw.
9	or/4-8

10	exp Animals/ not humans/
11	9 not 10
12	3 and 11

Cochrane Central Register of Controlled Trials

# ▲	Searches
1	enterocolitis, necrotizing/
2	(necroti?ing enterocolitis or enterocolitis necroticans or nec or typhlitis or pneumatosis intestinalis).tw.
3	**1 or 2**

Science Citation Index Expanded (SCI-EXPANDED) – from 1900
Social Sciences Citation Index (SSCI) – from 1956

Set	
# 5	#3 NOT #4 *Indexes=SCI-EXPANDED, SSCI Timespan=All years*
# 4	TS=(rat or rats or mouse or mice or pig or pigs or piglet or piglets or porcine or animal) *Indexes=SCI-EXPANDED, SSCI Timespan=All years*
# 3	#2 AND #1 *Indexes=SCI-EXPANDED, SSCI Timespan=All years*
# 2	TS=((doubl* NEAR blind*) or (singl* NEAR blind*) or allocat* or assign* or "cross over*" or crossover* or factorial* or placebo* or random* or volunteer*) *Indexes=SCI-EXPANDED, SSCI Timespan=All years*
# 1	TS=("necrotising enterocolitis" or "necrotizing enterocolitis" or "enterocolitis necroticans" or "nec" or "typhlitis" or "pneumatosis intestinalis") *Indexes=SCI-EXPANDED, SSCI Timespan=All years*

BIOSIS Previews 1980 to present (Oct 22, 2014)

Set	
# 5	#3 NOT #4 *Indexes=BIOSIS Previews Timespan=All years*
# 4	TS=(rat or rats or mouse or mice or pig or pigs or piglet or piglets or porcine) *Indexes=BIOSIS Previews Timespan=All years*
# 3	#2 AND #1 *Indexes=BIOSIS Previews Timespan=All years*
# 2	TS=((doubl* NEAR blind*) or (singl* NEAR blind*) or allocat* or assign* or "cross over*" or crossover* or factorial* or placebo* or random* or volunteer*) *Indexes=BIOSIS Previews Timespan=All years*
# 1	TS=("necrotising enterocolitis" or "necrotizing enterocolitis" or "enterocolitis necroticans" or "nec" or "typhlitis" or "pneumatosis intestinalis") *Indexes=BIOSIS Previews Timespan=All years*

CINAHL (EBSCO Host)

#	Searches
S17	S3 AND S16
S16	S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14 OR S15
S15	TX allocat* random*
S14	(MH "Quantitative Studies")
S13	(MH "Placebos")
S12	(MH "Placebos")
S11	TX placebo*
S10	TX random* allocat*
S9	(MH "Random Assignment")
S8	TX randomi* control* trial*
S7	TX ((singl* n1 blind*) or (singl* n1 mask*)) or TX ((doubl* n1 blind*) or (doubl* n1 mask*)) or TX ((tripl* n1 blind*) or (tripl* n1 mask*)) or TX ((trebl* n1 blind*) or (trebl* n1 mask*))
S6	TX clinic* n1 trial*
S5	PT Clinical trial
S4	(MH "Clinical Trials+")
S3	S1 OR S2
S2	TX ("necrotising enterocolitis" or "necrotizing enterocolitis" or "enterocolitis necroticans" or "nec" or "typhlitis" or "pneumatosis intestinalis")
S1	(MH "Enterocolitis, Necrotizing")

Scopus

((TITLE-ABS-KEY ("necrotising enterocolitis" OR "necrotizing enterocolitis" OR "enterocolitis necroticans" OR "nec" OR "typhlitis" OR "pneumatosis intestinalis")) AND (TITLE-ABS-KEY ((doubl* W/1 blind*) OR (singl* W/1 blind*) OR allocat* OR assign* OR "cross over*" OR crossover* OR factorial* OR placebo* OR random* OR volunteer*))) AND NOT (TITLE-ABS-KEY (rat OR rats OR mouse OR mice OR pig OR pigs OR piglet OR piglets OR porcine OR animal))

ProQuest Dissertations & Theses Full Text

(TI,AB,FT("necrotising enterocolitis" OR "necrotizing enterocolitis" OR "enterocolitis necroticans" OR "typhlitis" OR "pneumatosis intestinalis") AND TI,AB,FT((doubl* NEAR/1 blind*) OR (singl* NEAR/1 blind*) OR allocat* OR assign* OR "cross over*"

OR crossover* OR factorial* OR placebo* OR random* OR volunteer*)) NOT TI,AB,FT(rat OR rats OR mouse OR mice OR pig OR pigs OR piglet OR piglets OR porcine OR animal)

Google scholar

("necrotizing enterocolitis" OR "necrotising enterocolitis" OR "enterocolitis necroticans" OR "typhlitis" OR "pneumatosis intestinalis") AND (random AND trial)

WHO International Clinical Trials Registry Platform

("necrotizing enterocolitis" OR "necrotising enterocolitis" OR "enterocolitis necroticans" OR "typhlitis" OR "pneumatosis intestinalis")

The ISRCTN registry

("necrotizing enterocolitis" OR "necrotising enterocolitis" OR "enterocolitis necroticans" OR "typhlitis" OR "pneumatosis intestinalis")

clinicaltrials.gov

("necrotizing enterocolitis" OR "necrotising enterocolitis" OR "pneumatosis intestinalis") | Interventional Studies

Appendix 6. 2. Additional methods
Eligibility criteria

Type of studies

We included all randomized trials (RCTs) if they investigated the efficacy or safety of any probiotics, prebiotics, or synbiotics given alone or in combination with other preventive therapies compared to control or no treatment, placebo, or each other. Non-control studies such as pre-post studies, non-RCTs such as quasi-randomized trials, observational or cross-sectional studies were excluded. We included conference abstracts if enough outcome data was reported for the analysis.

Participants

We included RCTs enrolling preterm (gestational age <37 weeks) and/or low birth weight (birth weight <2500 g) infants. We excluded studies that enrolled term infants or included both term and infants, unless data for preterm infants were reported separately or > 80% of infants were preterm. We also excluded studies that enrolled infants once they achieved full enteral feed, or enrolled infants with early onset sepsis, feed intolerance, or necrotising enterocolitis. RCTs enrolling infants with abnormal antenatal Doppler, intra-uterine growth restriction, or small for gestational age were eligible.

Intervention

Eligible prophylactic interventions include any probiotics – defined in this review as living bacteria, prebiotics – defined as non-digestible compounds that provides a beneficial effect on the host by stimulating the growth of selected indigenous bacteria including lactoferrin, inulin, galacto- or fructo-oligosaccharides, or synbiotics, products

that contain both probiotics and prebiotics. We excluded RCTs that used fermented probiotics (dead organisms) as intervention.

Data abstraction and risk of bias assessment

Reviewers extracted the following data, independently and in duplicate: (i) general study information (author's name, publication year, country of origin, and funding source), (ii) study population details (sample size, mean gestational age, birth weight, percent caesarean deliveries, percent infants fed exclusively with mother's, donor's, or formula milk), (iii) details of the intervention and comparison (e.g. probiotics species and strains, dosage, time of initiation, and duration of therapy), and (iv) outcomes (Severe NEC - stage II or more based on Bell's criteria,[1,2] all-cause mortality, NEC-related mortality, and culture-proven sepsis, duration of hospitalization, weight 37 weeks postnatal age or at discharge, time to establish full enteral feeds (days), and feed intolerance).

Two reviewers independently assessed risk of bias using the modified Cochrane risk of bias instrument[3,4] that addresses the following issues: random sequence generation, allocation concealment, blinding of study participants (in the case of our study, infants' parents), blinding of healthcare providers, blinding of data collectors, and outcome assessors/adjudicators, incomplete outcome data (loss to follow-up > 5% of randomized population were considered at high risk of bias), and other potential sources of bias. For assessing the eligibility criteria, data extraction, and risk of bias assessments, we performed two rounds of calibration exercise. Assessment of non-English articles was performed by data extractors with the same language of publication.

We contacted authors of eligible studies (including conference abstracts) to get necessary outcome data for the analysis which was not reported in the original publication. When studies were published in duplicate or studies that used data from a similar study population in different publications in part or full, we extracted data from the publication with the most complete data set (e.g. publications with largest sample size and/or longest duration of follow-up). We used methods described by Cochrane Handbook[5] and Hozo et al[6] to estimate the mean and standard deviation where median, range, and sample size were reported, and to impute the standard deviation if the standard error or deviation for the differences are not reported. We considered no difference between infants receiving formula, parental nutrition, control or no treatment and infants receiving placebo and merged these comparisons into a single node called placebo (PLC).

Data synthesis

We used *mvmeta* command and network suite in STATA to carry out network meta-analysis (NMA), check the model assumptions, and present the results.[7,8] We assessed heterogeneity between RCTs for each direct comparison with visual inspection of the forest plots and the I^2 statistic. We planned to assess small study effect using the comparison-adjusted funnel plot in the network[8] and Harbord's test for all direct comparisons with at least 10 RCTs.[9]

We assessed the presence of incoherence by comparing direct evidence (i.e. estimates from pairwise comparisons) with indirect evidence (i.e. estimates from network meta-analysis) using the node splitting method.[10,11] In this approach, incoherence is assessed locally by evaluating the coherence assumption in each closed loop of the network

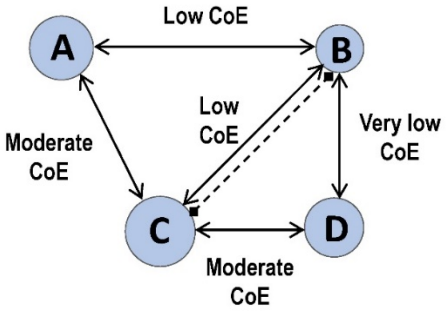
separately as the difference between direct and indirect estimates for a specific comparison in the loop. We assumed a common heterogeneity estimate within each loop. We also confirmed the coherence assumption in the entire network using 'design-by-treatment' model (global test) as described by Higgins et al.[12] When there was considerable local and global incoherence, which was not resolved by network meta-regression or subgroup analysis, we investigated the source of incoherence in loops of evidence. For one outcome (NEC - stage II or more), we found out that the global and local incoherence was dominated by a small 4-arm study[13] with factorial design (arm 1: B. lactis [N 50], arm 2: B. longum [N 48], arm 3: B. lactis and B. longum [N 43], arm 4: placebo [N 52]) with number of infants with severe NEC (stage II or more) in arm 3 being considerably higher that the other trial arms (arm 1: 2 infants, arm 2: 1 infant, arm 3: 5 infants and 3 infants in placebo arm). Findings for this outcome for arm 3 was not consistent with other outcomes of the study (including mortality, late-onset sepsis, feed intolerance, or weight). Hence, we decide not to include data for arm 3 of the study in the network meta-analysis of NEC. This resulted in significant improvement in local and global incoherence.

For direct comparisons, we considered $I^2 \leq 25\%$ as low heterogeneity, $25\% < I^2 \leq 50\%$ as moderate heterogeneity, and $I^2 > 50\%$ as substantial heterogeneity. The following potential sources of heterogeneity were identified a priori: gestational age, birth weight, percent infants exclusively fed by mothers' or donors' milk, percent delivered by C-section and risk of bias. The overall risk of bias for each study was assessed by taking the average of the 3 most important risk of bias items, allocation concealment, blinding, and

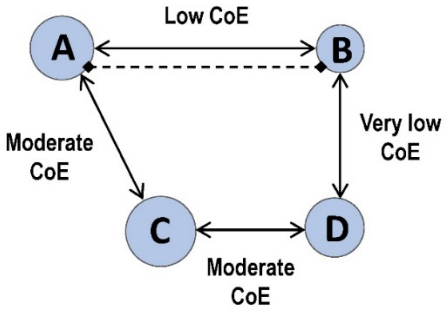
missing participant data. We performed network meta-regression adjusting effect estimates for covariates at study level assuming a common fixed coefficient across comparisons. Network meta-regression was performed only for primary outcomes (all-cause mortality, severe NEC, and culture proven sepsis).

Assessing certainty (quality) of the evidence using GRADE approach

We rated the certainty of evidence for each outcome in duplicate using the GRADE (Grading of Recommendations, Assessment, Development, and Evaluation) framework.[14-16] To rate certainty for network meta-analysis, both direct and indirect comparisons are considered. Direct comparisons are those for which a head-to-head comparison (the intervention of A versus B, the comparison of interest) is available. Indirect comparisons are those made via a third intervention – a common comparator (inferring the effect of A versus B through trials of A versus C and B versus C, described in derail below).



Example of a first-order loop



Example of a second-order loop

Initially, we rated the certainty in direct estimates according to the traditional GRADE guidance.[14] Then, rated the certainty in indirect estimate, with a focus on the dominant

lowest order loop.[15] Indirect effect estimates are calculated from available 'loops' of evidence, which includes first order loops (based on a single common comparator treatment, the difference between the treatment B and C is based on comparisons of A and C as well as A and B) or higher order loops (more than one intervening treatment connecting the two interventions, e.g. the loop of A vs B, B vs D, C vs D, and A vs C as depicted above). In the final step, we rated the certainty in of the network estimates. The overall certainty was rated based on four levels: high, moderate, low and very low.

For all certainty assessments of the indirect comparisons, information obtained from the first and second order loops in the network was used as shown in the example figure above. In the depicted hypothetical networks, each circle (node) indicates a treatment, each arrow indicates a direct comparison, and the dashed line indicates an indirect comparison. For the indirect comparison of B vs C (dash line), **in the left-side network**, the pathway of A – B – C is considered a first-order loop and the pathway of A – B – C – D is considered a second-order loop. certainty for indirect comparisons was decided based on the certainty of the first-order loop, which its certainty will be the lowest certainty of the direct comparisons within the first-order loop. In the depicted example in the left side network, the certainty of the indirect comparison of B vs C is the lower certainty among the two direct comparisons of A vs C (moderate) and A vs B (low), which is low. In case, an indirect comparison had more than one first-order loops, the loop with more number patients and studies was used as the certainty for the indirect comparison.

In cases where no first-order loop was available, the certainty for an indirect comparison was derived from a second-order loop from the lowest certainty of the direct comparisons within the second-order loop. For example, for the indirect comparison of A vs B (dash line), **in the right-side network**, the certainty is the lowest of the three direct comparisons within the second-order loop of A vs. C (moderate), C vs. D (moderate) and D vs. B (very low), which is very low. In addition, we considered further rating down each indirect comparison for intransitivity. We assumed transitivity when an indirect comparison could be considered as a valid method to compare two treatments, because the studies were sufficiently similar in important clinical and methodological characteristics, i.e. similar enough considering the distribution of effect modifiers.[17,18]

Network estimate certainty started as the higher of the direct and indirect evidence; however, we considered the relative contribution of direct and indirect evidence to the network estimate. When the certainty of the direct evidence was high and direct evidence contributes to the network estimate at least as much as the indirect evidence, we ignored the certainty rating of indirect evidence. We considered rating down the certainty if there was incoherence between the indirect and direct estimates (also called local inconsistency).[16,19] We evaluated the presence of incoherence by comparing direct evidence (i.e. estimates from pairwise comparisons) with indirect evidence (i.e. estimates from network meta-analysis) using the node splitting method.[10,11] A p value less than 0.05 was considered as significant incoherence between the direct and indirect comparisons. We did not rate down the certainty rating of the network estimate twice when both intransitivity and incoherence were present as incoherence can be considered the

statistical manifestation of intransitivity. We evaluated imprecision using the network estimate; if the 95% confidence interval excluded an odds ratio (OR) of 1, we did not rate down for imprecision. When the results did not exclude an OR of 1, we rated down for imprecision if the lower boundary of the 95% CI was below 0.8 or the upper boundary was above 1.25.

Appendix 6. 3. References to trials included in the network meta-analysis

1. Akin IM, Atasay B, Dogu F, et al. Oral Lactoferrin to Prevent Nosocomial Sepsis and Necrotizing Enterocolitis of Premature Neonates and Effect on T-Regulatory Cells. *Am J Perinatol* 2014; **31**(12): 1111-9.
2. Al-Hosni M, Duenas M, Hawk M, et al. Probiotics-supplemented feeding in extremely low-birth-weight infants. *J Perinatol* 2012; **32**(4): 253-9.
3. Amini E, Dalili H, Niknafs N, Shariat M, Nakhostin M, Jedari-Attari S. The Effect of Probiotics in Prevention of Necrotising Enterocolitis in Preterm Neonates in Comparison with Control Group. *Iran J Pediatr* 2017; **27**(6): e7663.
4. Armanian AM, Sadeghnia A, Hoseinzadeh M, et al. The effect of neutral oligosaccharides on reducing the incidence of necrotizing enterocolitis in preterm infants: A randomized clinical trial. *International Journal of Preventive Medicine* 2014; **5**(11): 1387-95.
5. Arora S, Khurana MS, Saini R. To study the role of probiotics in the prevention of necrotizing enterocolitis in preterm neonates. *International Journal of Contemporary Pediatrics* 2017; **4**(5): 1792-7.
6. Barrington KJ, Assaad MA, Janvier A. The Lacuna Trial: a double-blind randomized controlled pilot trial of lactoferrin supplementation in the very preterm infant. *J Perinatol* 2016; **36**(8): 666-9.
7. Bin-Nun A, Bromiker R, Wilschanski M, et al. Oral probiotics prevent necrotizing enterocolitis in very low birth weight neonates. *J Pediatr* 2005; **147**(2): 192-6.
8. Braga TD, Da Silva GAP, De Lira PIC, De Carvalho Lima M. Efficacy of Bifidobacterium breve and Lactobacillus casei oral supplementation on necrotizing enterocolitis in very-low-birth-weight preterm infants: A double-blind, randomized, controlled trial. *Am J Clin Nutr* 2011; **93**(1): 81-6.
9. Chowdhury T, Ali MM, Hossain MM, et al. Efficacy of Probiotics Versus Placebo in the Prevention of Necrotizing Enterocolitis in Preterm Very Low Birth Weight Infants: A Double-Blind Randomized Controlled Trial. *Journal of the College of Physicians and Surgeons--Pakistan : JCPSP* 2016; **26**(9): 770-4.
10. Chrzanowska-Liszewska D, Seliga-Siwecka J, Kornacka MK. The effect of Lactobacillus rhamnosus GG supplemented enteral feeding on the microbiotic flora of preterm infants-double blinded randomized control trial. *Early Hum Dev* 2012; **88**(1): 57-60.
11. Costalos C, Skouteri V, Gounaris A, et al. Enteral feeding of premature infants with Saccharomyces boulardii. *Early Hum Dev* 2003; **74**(2): 89-96.
12. Costeloe K, Hardy P, Juszczak E, Wilks M, Millar MR. Bifidobacterium breve BBG-001 in very preterm infants: A randomised controlled phase 3 trial. *The Lancet* 2016; **387**(10019): 649-60.

13. Dani C, Biadaioli R, Bertini G, Martelli E, Rubaltelli FF. Probiotics feeding in prevention of urinary tract infection, bacterial sepsis and necrotizing enterocolitis in preterm infants. A prospective double-blind study. *Biology of the neonate* 2002; **82**(2): 103-8.

14. Dashti AS, Afjeh SA, Basiry A, Shirvani F, Seifi K, Taheri ZM. Prophylactic probiotics for prevention of necrotizing enterocolitis (NEC) in low birth weight neonates. *Arch Pediatr Infect Dis* 2014; **2**(1): 174-9.

15. Dasopoulou M, Briana DD, Boutsikou T, et al. Motilin and gastrin secretion and lipid profile in preterm neonates following prebiotics supplementation: A double-blind randomized controlled study. *J Parenter Enter Nutr* 2015; **39**(3): 359-68.

16. Demirel G, Erdeve O, Celik IH, Dilmen U. Saccharomyces boulardii for prevention of necrotizing enterocolitis in preterm infants: A randomized, controlled study. *Acta Paediatr Int J Paediatr* 2013; **102**(12): e560-e5.

17. Deng J, Chen K. Early minimal feeding combined with probiotics to prevent necrotizing enterocolitis in preterm infant. *Chinese Journal of Modern Drug Application* 2010; **4**: 13-4.

18. Di M, Li X. Effects of Bifidobacterium supplementation for prevention of necrotizing enterocolitis in preterm infants: a randomized, controlled trial. *Zhong Guo She Qu Yi Shi* 2010; **231**(69).

19. Dilli D, Aydin B, Fettah ND, et al. The propre-save study: Effects of probiotics and prebiotics alone or combined on necrotizing enterocolitis in very low birth weight infants. *J Pediatr* 2015; **166**(3): 545-51.

20. Dongol Singh S, Klobassa D, Resch B, Urlesberger B, Shrestha R. Placebo Controlled Introduction of Prophylactic Supplementation of Probiotics to Decrease the Incidence of Necrotizing Enterocolitis at Dhulikhel Hospital in Nepal. *Kathmandu Univ Med J* 2017; **60**(4): 319-23.

21. Dutta S, Ray P, Narang A. Comparison of Stool Colonization in Premature Infants by Three Dose Regimes of a Probiotic Combination: A Randomized Controlled Trial. *Am J Perinatol* 2015; **32**(8): 733-40.

22. Fernández-Carrocera LA, Solis-Herrera A, Cabanillas-Ayón M, et al. Double-blind, randomised clinical assay to evaluate the efficacy of probiotics in preterm newborns weighing less than 1500 g in the prevention of necrotising enterocolitis. *Arch Dis Child Fetal Neonatal Ed* 2013; **98**(1): F5-F9.

23. Fujii T, Ohtsuka Y, Lee T, et al. Bifidobacterium breve enhances transforming growth factor beta1 signaling by regulating Smad7 expression in preterm infants. *J Pediatr Gastroenterol Nutr* 2006; **43**(1): 83-8.

24. Guney-Varal I, Koksal N, Ozkan H, Bagci O, Dogan P. The effect of early administration of combined multi-strain and multi-species probiotics on

gastrointestinal morbidities and mortality in preterm infants: A randomized controlled trial in tertiary care unit. *Turkish Journal of Pediatrics* 2017; **59**(1): 13-9.

25. Hariharan D, Balasubramanian L, Kannappan V, Veluswami G. Probiotic supplementation in VLBW preterm infants improves feeding tolerance and reduces risk of gram negative sepsis. *J Pediatr Gastroenterol Nutr* 2016; **62**: 655.
26. Hays S, Jacquot A, Gauthier H, et al. Probiotics and growth in preterm infants: A randomized controlled trial, PREMAPRO study. *Clin Nutr* 2016; **35**(4): 802-11.
27. Hernández-Enríquez NP, Rosas-Sumano AB, Monzoy-Ventre MA, Galicia-Flores L. Lactobacillus reuteri DSM 17938 en la prevención de enterocolitis necrosante en recién nacidos prematuros. Estudio piloto de eficacia y seguridad. *Rev Mex Pediatr* 2016; **83**(2): 37-43.
28. Hikaru H, Koichi K, Yayoi Y, et al. Bifidobacteria prevents preterm infants from developing infection and sepsis. *Int J Probiotics Prebiotics* 2010; **5**(1): 33-6.
29. Hua XT, Tang J, Mu DZ. Effect of oral administration of probiotics on intestinal colonization with drug-resistant bacteria in preterm infants. [Chinese]. *Chinese Journal of Contemporary Pediatrics* 2014; **16**(6): 606-9.
30. Huang B, Yang H, Huang X. Probiotics supplementation for prevention of necrotizing enterocolitis in very low-birth-weight neonates: a randomized, controlled trial. *Journal of Guangdong Medical College* 2009; **27**: 37-9.
31. Indrio F, Riezzo G, Tafuri S, et al. Probiotic Supplementation in Preterm: Feeding Intolerance and Hospital Cost. *Nutrients* 2017; **9**(9).
32. Jacobs SE, Tobin JM, Opie GF, et al. Probiotic effects on late-onset sepsis in very preterm infants: A randomized controlled trial. *Pediatrics* 2013; **132**(6): 1055-62.
33. Kanic Z, Micetic Turk D, Burja S, Kanic V, Dinevski D. Influence of a combination of probiotics on bacterial infections in very low birthweight newborns. *Wien Klin Wochenschr* 2015; **127**: 210-5.
34. Kaur G, Gathwala G. Efficacy of Bovine Lactoferrin Supplementation in Preventing Late-onset Sepsis in low Birth Weight Neonates: A Randomized Placebo-Controlled Clinical Trial. *J Trop Pediatr* 2015; **61**(5): 370-6.
35. Ke D, Su Z, Li L. Effects of Bifido supplement for prevention of necrotizing enterocolitis in preterm infants: a randomized controlled trial. *Chin Pediatr Emerg Med* 2008; **12**: 69-71.
36. Kitajima H, Sumida Y, Tanaka R, Yuki N, Takayama H, Fujimura M. Early administration of Bifidobacterium breve to preterm infants: randomised controlled trial. *Arch Dis Child Fetal Neonatal Ed* 1997; **76**(2): F101-7.
37. Lin HC, Hsu CH, Chen HL, et al. Oral probiotics prevent necrotizing enterocolitis in very low birth weight preterm infants: a multicenter, randomized, controlled trial. *Pediatrics* 2008; **122**(4): 693-700.

38. Lin HC, Su BH, Chen AC, et al. Oral probiotics reduce the incidence and severity of necrotizing enterocolitis in very low birth weight infants. *Pediatrics* 2005; **115**(1): 1-4.

39. Manzoni P, Meyer M, Stolfi I, et al. Bovine lactoferrin supplementation for prevention of necrotizing enterocolitis in very-low-birth-weight neonates: A randomized clinical trial. *Early Hum Dev* 2014; **90**(SUPPL.1): S60-S5.

40. Manzoni P, Mostert M, Leonessa ML, et al. Oral supplementation with Lactobacillus casei subspecies rhamnosus prevents enteric colonization by Candida species in preterm neonates: a randomized study. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2006; **42**(12): 1735-42.

41. Mihatsch WA, Vossbeck S, Eikmanns B, Hoegel J, Pohlandt F. Effect of bifidobacterium lactis on the incidence of nosocomial infections in very-low-birth-weight infants: A randomized controlled trial. *Neonatology* 2010; **98**(2): 156-63.

42. Millar MR, Bacon C, Smith SL, Walker V, Hall MA. Enteral Feeding of Premature-Infants with Lactobacillus Gg. *Archives of Disease in Childhood* 1993; **69**(5): 483-7.

43. Modi N, Uthaya S, Fell J, Kulinskaya E. A randomized, double-blind, controlled trial of the effect of prebiotic oligosaccharides on enteral tolerance in preterm infants (ISRCTN77444690). *Pediatr Res* 2010; **68**(5): 440-5.

44. Mohan R, Koebnick C, Schildt J, Mueller M, Radke M, Blaut M. Effects of Bifidobacterium lactis Bb12 supplementation on body weight, fecal pH, acetate, lactate, calprotectin, and IgA in preterm infants. *Pediatr Res* 2008; **64**(4): 418-22.

45. Nandhini LP, Biswal N, Adhisivam B, Mandal J, Bhat BV, Mathai B. Synbiotics for decreasing incidence of necrotizing enterocolitis among preterm neonates - a randomized controlled trial. *Journal of Maternal-Fetal & Neonatal Medicine* 2016; **29**(5): 821-5.

46. Ochoa TJ, Zegarra J, Cam L, et al. Randomized controlled trial of lactoferrin for prevention of sepsis in peruvian neonates less than 2500 g. *The Pediatric infectious disease journal* 2015; **34**(6): 571-6.

47. Oncel MY, Sari FN, Arayici S, et al. Lactobacillus Reuteri for the prevention of necrotising enterocolitis in very low birthweight infants: A randomised controlled trial. *Arch Dis Child Fetal Neonatal Ed* 2014; **99**(2): F110-F5.

48. Patole S, Keil AD, Chang A, et al. Effect of Bifidobacterium breve M-16V supplementation on fecal bifidobacteria in preterm neonates - A randomised double blind placebo controlled trial. *PLoS ONE* 2014; **9**(3): e89511.

49. Punnahitananda S, Thaithumyanon P, Soongsawang K. Nosocomial infection and necrotizing enterocolitis in preterm neonates treated with Lactobacillus acidophilus and Bifidobacterium infantis in a neonatal intensive care unit: A randomized

controlled study. *The 14th Congress of the Federation of Asia Oceania Perinatal Societies* 2006.

50. Qiao LX, Zhu WY, Zhang HY, Wang H. Effect of early administration of probiotics on gut microflora and feeding in pre-term infants: a randomized controlled trial. *The journal of maternal-fetal & neonatal medicine : the official journal of the European Association of Perinatal Medicine, the Federation of Asia and Oceania Perinatal Societies, the International Society of Perinatal Obstet* 2017; **30**(1): 13-6.
51. Ren B. Preventive effect of Bifidobacterium tetravaccine tablets in premature infants with necrotizing enterocolitis. *Journal of Pediatric Pharmacy* 2010; **16**(2): 24-5.
52. Reuman PD, Duckworth DH, Smith KL, Kagan R, Bucciarelli RL, Ayoub EM. Lack of effect of Lactobacillus on gastrointestinal bacterial colonization in premature infants. *Pediatric infectious disease* 1986; **5**(6): 663-8.
53. Rojas MA, Lozano JM, Rojas MX, et al. Prophylactic probiotics to prevent death and nosocomial infection in preterm infants. *Pediatrics* 2012; **130**(5): e1113-e20.
54. Romeo MG, Romeo DM, Trovato L, et al. Role of probiotics in the prevention of the enteric colonization by Candida in preterm newborns: incidence of late-onset sepsis and neurological outcome. *J Perinatol* 2011; **31**(1): 63-9.
55. Rouge C, Piloquet H, Butel MJ, et al. Oral supplementation with probiotics in very-low-birth-weight preterm infants: a randomized, double-blind, placebo-controlled trial. *Am J Clin Nutr* 2009; **89**(6): 1828-35.
56. Roy A, Chaudhuri J, Sarkar D, Ghosh P, Chakraborty S. Role of Enteric Supplementation of Probiotics on Late-onset Sepsis by Candida species in Preterm Low Birth Weight Neonates: A Randomized, Double Blind, Placebo-controlled Trial. *North American journal of medical sciences* 2014; **6**(1): 50-7.
57. Sadowska-Krawczenko I, Korbal P, Polak A, Wietlicka-Piszcz M, Szajewska H. Lactobacilllus rhamnosus ATC A07FA for preventing necrotizing enterocolitis in very-low-birth-weight preterm infants: A randomized controlled trial (preliminary results). [Polish]. *Pediatria polska* 2012; **87**(2): 139-45.
58. Saengtawesin V, Tangpolkaiwalsak R, Kanjanapattankul W. Effect of oral probiotics supplementation in the prevention of necrotizing enterocolitis among very low birth weight preterm infants. *Journal of the Medical Association of Thailand = Chotmaihet thangphaet* 2014; **97**(Suppl 6): S20-S5.
59. Samanta M, Sarkar M, Ghosh P, Ghosh J, Sinha M, Chatterjee S. Prophylactic probiotics for prevention of necrotizing enterocolitis in very low birth weight newborns. *J Trop Pediatr* 2009; **55**(2): 128-31.
60. Sari FN, Dizdar EA, Oguz S, Erdeve O, Uras N, Dilmen U. Oral probiotics: Lactobacillus sporogenes for prevention of necrotizing enterocolitis in very low-

birth weight infants: A randomized, controlled trial. *Eur J Clin Nutr* 2011; **65**(4): 434-9.

61. Serce O, Benzer D, Gursoy T, Karatekin G, Ovali F. Efficacy of saccharomyces boulardii on necrotizing enterocolitis or sepsis in very low birth weight infants: A randomised controlled trial. *Early Hum Dev* 2013; **89**(12): 1033-6.
62. Serce O, Gursoy T, Karatekin G, Ovali F. Effects of prebiotic and probiotic combination on necrotizing enterocolitis and sepsis prophylaxis in very low birth weight infants. *Journal of Perinatal Medicine* 2013; **41**.
63. Shadkam MN, Jalalizadeh F, Nasiriani K. Effects of probiotic lactobacillus reuteri (DSM 17938) on the incidence of necrotizing enterocolitis in very low birth weight premature infants. *Iranian Journal of Neonatology* 2015; **6**(4): 15-20.
64. Shashidhar A, Suman Rao PN, Nesargi S, Bhat S, Chandrakala BS. Probiotics for Promoting Feed Tolerance in Very Low Birth Weight Neonates - A Randomized Controlled Trial. *Indian Pediatr* 2017; **54**(5): 363-7.
65. Sherman MP, Adamkin DH, Niklas V, et al. Randomized Controlled Trial of Talactoferrin Oral Solution in Preterm Infants. *J Pediatr* 2016; **175**: 68-73.
66. Sinha A, Gupta SS, Chellani H, et al. Role of probiotics VSL#3 in prevention of suspected sepsis in low birthweight infants in India: a randomised controlled trial. *BMJ Open* 2015; **5**(7).
67. Sreenivasa B, Sunil KP, Suresh BM, Ragavendra K. Role of synbiotics in the prevention of necrotizing enterocolitis in preterm neonates: a randomized controlled trial. *International Journal of Contemporary Pediatrics* 2017; **2**(2): 127-30.
68. Stratiki Z, Costalos C, Sevastiadou S, et al. The effect of a bifidobacter supplemented bovine milk on intestinal permeability of preterm infants. *Early Hum Dev* 2007; **83**(9): 575-9.
69. Tewari VV, Dubey SK, Gupta G. Bacillus clausii for Prevention of Late-onset Sepsis in Preterm Infants: A Randomized Controlled Trial. *J Trop Pediatr* 2015; **61**(5): 377-84.
70. Totsu S, Yamasaki C, Terahara M, Uchiyama A, Kusuda S. Bifidobacterium and enteral feeding in preterm infants: Cluster-randomized trial. *Pediatrics International* 2014; **56**(5): 714-9.
71. Underwood MA, Salzman NH, Bennett SH, et al. A Randomized Placebo-controlled Comparison of 2 Prebiotic/Probiotic Combinations in Preterm Infants: Impact on Weight Gain, Intestinal Microbiota, and Fecal Short-chain Fatty Acids. *J Pediatr Gastroenterol Nutr* 2009; **48**(2): 216-25.
72. Van Niekerk E, Kirsten GF, Nel DG, Blaauw R. Probiotics, feeding tolerance, and growth: A comparison between HIV-exposed and unexposed very low birth weight infants. *Nutrition* 2014; **30**(6): 645-53.

73. Van Niekerk E, Nel DG, Blaauw R, Kirsten GF. Probiotics Reduce Necrotizing Enterocolitis Severity in HIV-exposed Premature Infants. *J Trop Pediatr* 2015; **61**(3): 155-64.

74. Wejryd E, et al. Probiotics promoted head growth in extremely low birthweight infants in a double-blind placebo-controlled trial. Acta Paediatr, 2019. 108: 62-69.

75. Westerbeek EAM, Hensgens RL, Mihatsch WA, Boehm G, Lafeber HN, van Elburg RM. The effect of neutral and acidic oligosaccharides on stool viscosity, stool frequency and stool pH in preterm infants. *Acta Paediatr* 2011; **100**(11): 1426-31.

76. Xiao-yuan Z, Lian-qiao LI, Xuan-xuan GAO, Li-duan SU. Relative Factors of Neonatal Necrotizing Enterocolitis and Preventive Effect of Microeco-Preparation. *Journal of Applied Clinical Pediatrics* 2007; **22**(18):1392-93.

77. Xu L, Wang Y, Fu J, Sun M, Mao Z, Vandenplas Y. A double-blinded randomized trial on growth and feeding tolerance with Saccharomyces boulardii CNCM I-745 in formula-fed preterm infants. *Jornal de Pediatria* 2016; **92**(3): 296-301.

78. Yang S, Haiying Y, Bin G, Shu X, Xianglan D, Jiang W. The Clinical Application Value of Endangered Preterm Infants Given Earlier Amounts of Micro Feedings and Adding Probiotics. *Journal of Pediatric Pharmacy* 2011; **17**(3):21-24.

79. Zhou N. The observation of effect of probiotics in the prevention of neonatal necrotizing enterocolitis. *Chinese Journal of Ethnomedicine and Ethnopharmacy* 2012; **21**: 81.

Appendix 6. 4. Characteristics of participants of trials included in the network meta-analysis

Ref #	Study (year)	LN	BW (g)	GA (w)	APGAR M5	% C-Sec	% MV	% IUGR	% SGA	% MM fed	% FM fed
1	Akin (2014)	EN	1298.5	29.9	9.0	90.0	6.0	-	-	58.0	0.0
2	Al-Hosni (2012)	EN	778.5	25.7	7.5	51.5	-	-	0.0	-	0.0
3	Amini (2017)	EN	1153.3	29.6	-	-	-	-	-	-	-
4	Armanian (2014)	EN	1224.7	30.4	-	-	-	-	-	100.0	0.0
5	Arora (2017)	EN	1700.0	32.9	-	82.0	-	9.3	22.7	100.0	0.0
6	Barrington (2016)	EN	1095.5	28.2	-	55.7	-	-	-	92.4	-
7	Bin-Nun (2005)	EN	1131.4	29.5	8.0	78.0	-	-	20.0	61.4	24.1
8	Braga (2011)	EN	1173.7	29.4	8.0	51.5	-	-	23.0	-	0.0
9	Chowdhury (2016)	EN	1324.0	31.5	-	-	-	-	-	-	0.0
10	Chrzanowska-Lisiewska (2012)	EN	1257.8	29.5	6.6	72.3	-	-	100.0	0.0	100.0
11	Costalos (2003)	EN	1648.1	31.4	-	44.8	-	15.0	-	0.0	100.0
12	Costeloe (2016)	EN	1041.0	28.0	-	52.7	-	-	-	46.1	34.1
13	Dani (2002)	EN	1334.9	30.8	-	79.3	45.5	17.3	-		35.0
14	Dashti (2014)	EN	1406.4	31.2	9.2	82.4	50.0	-	-	34.6	5.9
15	Dasopoulou (2013)	EN	2003.2	34.0	9.0	78.4	-	-	-	0.0	100.0
16	Demirel (2013)	EN	1147.4	29.3	6.0	80.4	-	-	-	-	-
17	Deng (2010)	CHN	1628.7	32.8	-	-	-	-	-	-	-
18	Di (2010)	CHN	-	33.0	-	-	-	-	-	-	-
19	Dilli (2015)	EN	1204.3	28.7	7.0	34.5	-	-	-		21.8
20	Dongol Singh (2017)	EN	-	-	-	-	-	-	-	-	-
21	Dutta (2014)	EN	1323.3	30.9	8.9	32.9	29.5	-	-	-	7.4
22	Fernández-Carrocera (2013)	EN	1130.0	31.1	-	-	2.0	-	78.0	-	-
23	Fujii (2006)	EN	1427.7	31.3	-	-	-	-	-	-	-
24	Güney-Varal (2017)	EN	1546.5	29.6	7.5	80.9	-	-	-	-	-
25	Hariharan (2016)	EN	959.2	29.0	-	-	36.7	-	-	-	-
26	Hays (2016)	EN	1170.0	29.2	-	78.2	51.8	-	-	-	0.5

Ref #	Study (year)	LN	BW (g)	GA (w)	APGAR M5	% C-Sec	% MV	% IUGR	% SGA	% MM fed	% FM fed
1	Akin (2014)	EN	1298.5	29.9	9.0	90.0	6.0	-	-	58.0	0.0
2	Al-Hosni (2012)	EN	778.5	25.7	7.5	51.5	-	-	0.0	-	0.0
3	Amini (2017)	EN	1153.3	29.6	-	-	-	-	-	-	-
4	Armanian (2014)	EN	1224.7	30.4	-	-	-	-	-	100.0	0.0
5	Arora (2017)	EN	1700.0	32.9	-	82.0	-	9.3	22.7	100.0	0.0
6	Barrington (2016)	EN	1095.5	28.2	-	55.7	-	-	-	92.4	-
7	Bin-Nun (2005)	EN	1131.4	29.5	8.0	78.0	-	-	20.0	61.4	24.1
8	Braga (2011)	EN	1173.7	29.4	8.0	51.5	-	-	23.0	-	0.0
9	Chowdhury (2016)	EN	1324.0	31.5	-	-	-	-	-	-	0.0
10	Chrzanowska-Lisiewska (2012)	EN	1257.8	29.5	6.6	72.3	-	-	100.0	0.0	100.0
11	Costalos (2003)	EN	1648.1	31.4	-	44.8	-	15.0	-	0.0	100.0
12	Costeloe (2016)	EN	1041.0	28.0	-	52.7	-	-	-	46.1	34.1
13	Dani (2002)	EN	1334.9	30.8	-	79.3	45.5	17.3	-		35.0
14	Dashti (2014)	EN	1406.4	31.2	9.2	82.4	50.0	-	-	34.6	5.9
15	Dasopoulou (2013)	EN	2003.2	34.0	9.0	78.4	-	-	-	0.0	100.0
16	Demirel (2013)	EN	1147.4	29.3	6.0	80.4	-	-	-	-	-
17	Deng (2010)	CHN	1628.7	32.8	-	-	-	-	-	-	-
27	Hernandez-Enriquez (2016)	SPN	1293.3	31.4	-	72.3	13.6	-	-	-	0.0
28	Hikaru (2010)	EN	1036.4	28.3	-	-	-	-	-	-	
29	Hua (2014)	CHN	1786.6	33.1	-	60.3	-	-	23.0	24.1	75.9
30	Huang (2009)	CHN	1100.0	30.1	-	-	-	-	-	-	-
31	Indrio (2017)	EN	1439.1	-	-	85.0	-	-	0.0	0.0	100.0
32	Jacobs (2013)	EN	1055.5	27.9	8.0	67.0	-	-	-	50.8	-
33	Kanic (2015)	EN	1064.2	28.5	7.5	-	77.5	-	-	31.3	-
34	Kaur (2015)	EN	1489.3	34.1	8.2	35.4	-	-	-	-	-
35	Ke (2008)	CHN	-	33.5	-	-	-	-	-	-	-
36	Kitajima (1997)	EN	1026.0	28.2	-	-	-	-	13.2	-	0.0

Ref #	Study (year)	LN	BW (g)	GA (w)	APGAR M5	% C-Sec	% MV	% IUGR	% SGA	% MM fed	% FM fed
1	Akin (2014)	EN	1298.5	29.9	9.0	90.0	6.0	-	-	58.0	0.0
2	Al-Hosni (2012)	EN	778.5	25.7	7.5	51.5	-	-	0.0	-	0.0
3	Amini (2017)	EN	1153.3	29.6	-	-	-	-	-	-	-
4	Armanian (2014)	EN	1224.7	30.4	-	-	-	-	-	100.0	0.0
5	Arora (2017)	EN	1700.0	32.9	-	82.0	-	9.3	22.7	100.0	0.0
6	Barrington (2016)	EN	1095.5	28.2	-	55.7	-	-	-	92.4	-
7	Bin-Nun (2005)	EN	1131.4	29.5	8.0	78.0	-	-	20.0	61.4	24.1
8	Braga (2011)	EN	1173.7	29.4	8.0	51.5	-	-	23.0	-	0.0
9	Chowdhury (2016)	EN	1324.0	31.5	-	-	-	-	-	-	0.0
10	Chrzanowska-Lisiewska (2012)	EN	1257.8	29.5	6.6	72.3	-	-	100.0	0.0	100.0
11	Costalos (2003)	EN	1648.1	31.4	-	44.8	-	15.0	-	0.0	100.0
12	Costeloe (2016)	EN	1041.0	28.0	-	52.7	-	-	-	46.1	34.1
13	Dani (2002)	EN	1334.9	30.8	-	79.3	45.5	17.3	-		35.0
14	Dashti (2014)	EN	1406.4	31.2	9.2	82.4	50.0	-	-	34.6	5.9
15	Dasopoulou (2013)	EN	2003.2	34.0	9.0	78.4	-	-	-	0.0	100.0
16	Demirel (2013)	EN	1147.4	29.3	6.0	80.4	-	-	-	-	-
17	Deng (2010)	CHN	1628.7	32.8	-	-	-	-	-	-	-
37	Lin (2005)	EN	1087.2	28.3	-	55.6	-	-	22.6	-	0.0
38	Lin (2008)	EN	1053.1	-	-	65.7	-	-	21.4	65.4	0.0
39	Manzoni (2006)	EN	1192.8	29.4	6.5	32.5	-	-	-	-	0.0
40	Manzoni (2014)	EN	1134.8	29.6	7.4	78.1	-	-	-	28.0	17.9
41	Mihatsch (2010)	EN	863.4	26.7	7.9	69.4	-	-	10.0	17.2	67.2
42	Millar (1993)	EN	1472.5	30.3	-	40.0	-	-	-	-	-
43	Modi (2010)	EN	1538.7	30.5	-	-	-	-	-	7.8	14.9
44	Mohan (2008)	EN	1425.3	31.2	8.1	88.4	-	-	-	-	15.9
45	Nandhini (2015)	EN	1437.1	31.5	-	9.2	3.7	-	11.0	100.0	0.0
46	Ochoa (2015)	EN	1591.0	32.1	-	83.2	-	-	26.3	-	-

Ref #	Study (year)	LN	BW (g)	GA (w)	APGAR M5	% C-Sec	% MV	% IUGR	% SGA	% MM fed	% FM fed
1	Akin (2014)	EN	1298.5	29.9	9.0	90.0	6.0	-	-	58.0	0.0
2	Al-Hosni (2012)	EN	778.5	25.7	7.5	51.5	-	-	0.0	-	0.0
3	Amini (2017)	EN	1153.3	29.6	-	-	-	-	-	-	-
4	Armanian (2014)	EN	1224.7	30.4	-	-	-	-	-	100.0	0.0
5	Arora (2017)	EN	1700.0	32.9	-	82.0	-	9.3	22.7	100.0	0.0
6	Barrington (2016)	EN	1095.5	28.2	-	55.7	-	-	-	92.4	-
7	Bin-Nun (2005)	EN	1131.4	29.5	8.0	78.0	-	-	20.0	61.4	24.1
8	Braga (2011)	EN	1173.7	29.4	8.0	51.5	-	-	23.0	-	0.0
9	Chowdhury (2016)	EN	1324.0	31.5	-	-	-	-	-	-	0.0
10	Chrzanowska-Lisiewska (2012)	EN	1257.8	29.5	6.6	72.3	-	-	100.0	0.0	100.0
11	Costalos (2003)	EN	1648.1	31.4	-	44.8	-	15.0	-	0.0	100.0
12	Costeloe (2016)	EN	1041.0	28.0	-	52.7	-	-	-	46.1	34.1
13	Dani (2002)	EN	1334.9	30.8	-	79.3	45.5	17.3	-		35.0
14	Dashti (2014)	EN	1406.4	31.2	9.2	82.4	50.0	-	-	34.6	5.9
15	Dasopoulou (2013)	EN	2003.2	34.0	9.0	78.4	-	-	-	0.0	100.0
16	Demirel (2013)	EN	1147.4	29.3	6.0	80.4	-	-	-	-	-
17	Deng (2010)	CHN	1628.7	32.8	-	-	-	-	-	-	-
47	Oncel (2014)	EN	1059.5	28.1	8.0	78.8	-	-	8.2	15.0	13.5
48	Patole (2014)	EN	1060.2	28.5	-	69.9	-	-	25.2	-	1.3
49	Punnahitananda (2006)	EN	1135.8	29.2	-	-	-	-	-	-	
50	Qiao (2016)	EN	1623.0	32.3	-	-	28.3	-	-	0.0	100.0
51	Ren (2010)	CHN	1700.0	31.0	-	-	-	-	-	-	-
52	Reuman (1986)	EN	1371.5	30.6	7.1	-	-	-	-	-	-
53	Rojas (2012)	EN	1522.9	32.0	9.0	82.9		-	28.9	2.8	40.5
54	Romeo (2011)	EN	1961.7	33.5	-	91.0	-	-	-	-	-
55	Rouge (2009)	EN	1084.8	28.1	8.9	67.0	-	4.3	-	-	-
56	Roy (2014)	EN	1130.5	32.1	-	80.4	9.8		-	-	0.0

Ref #	Study (year)	LN	BW (g)	GA (w)	APGAR M5	% C-Sec	% MV	% IUGR	% SGA	% MM fed	% FM fed
1	Akin (2014)	EN	1298.5	29.9	9.0	90.0	6.0	-	-	58.0	0.0
2	Al-Hosni (2012)	EN	778.5	25.7	7.5	51.5	-	-	0.0	-	0.0
3	Amini (2017)	EN	1153.3	29.6	-	-	-	-	-	-	-
4	Armanian (2014)	EN	1224.7	30.4	-	-	-	-	-	100.0	0.0
5	Arora (2017)	EN	1700.0	32.9	-	82.0	-	9.3	22.7	100.0	0.0
6	Barrington (2016)	EN	1095.5	28.2	-	55.7	-	-	-	92.4	-
7	Bin-Nun (2005)	EN	1131.4	29.5	8.0	78.0	-	-	20.0	61.4	24.1
8	Braga (2011)	EN	1173.7	29.4	8.0	51.5	-	-	23.0	-	0.0
9	Chowdhury (2016)	EN	1324.0	31.5	-	-	-	-	-	-	0.0
10	Chrzanowska-Lisiewska (2012)	EN	1257.8	29.5	6.6	72.3	-	-	100.0	0.0	100.0
11	Costalos (2003)	EN	1648.1	31.4	-	44.8	-	15.0	-	0.0	100.0
12	Costeloe (2016)	EN	1041.0	28.0	-	52.7	-	-	-	46.1	34.1
13	Dani (2002)	EN	1334.9	30.8	-	79.3	45.5	17.3	-		35.0
14	Dashti (2014)	EN	1406.4	31.2	9.2	82.4	50.0	-	-	34.6	5.9
15	Dasopoulou (2013)	EN	2003.2	34.0	9.0	78.4	-	-	-	0.0	100.0
16	Demirel (2013)	EN	1147.4	29.3	6.0	80.4	-	-	-	-	-
17	Deng (2010)	CHN	1628.7	32.8	-	-	-	-	-	-	-
57	Sadowska-Krawczenko (2012)	POL	973.1	29.5	7.0	65.5	-	41.8	-	-	0.0
58	Saengtawesin (2014)	EN	1229.6	30.8	-	65.0	-	-	8.3	-	0.0
59	Samanta (2009)	EN	1191.4	30.1	-	47.8	-	-	34.4	100.0	0.0
60	Sari (2011)	EN	1254.5	29.6	-	71.5	-	-	7.2	28.1	0.0
61	Serce (2013) A	EN	1144.0	28.8	7.5	84.6	-	-	11.0	-	-
62	Serce (2013) B	EN	1236.3	29.0	-	89.0	-	-	-	-	20.0
63	Shadkam (2015)	EN	1407.5	30.9	-	-	-	-	-	100.0	0.0
64	Shashidhar (2017)	EN	1223.0	31.1	8.0	62.5	-	-	35.6	45.2	0.0
65	Sherman (2016)	EN	1147.5	28.0	8.0	77.5	-	-	16.7	75.8	24.2
66	Sinha (2015)	EN	2262.0	-	-	5.8	-	-	-	-	0.0

Ref #	Study (year)	LN	BW (g)	GA (w)	APGAR M5	% C-Sec	% MV	% IUGR	% SGA	% MM fed	% FM fed
1	Akin (2014)	EN	1298.5	29.9	9.0	90.0	6.0	-	-	58.0	0.0
2	Al-Hosni (2012)	EN	778.5	25.7	7.5	51.5	-	-	0.0	-	0.0
3	Amini (2017)	EN	1153.3	29.6	-	-	-	-	-	-	-
4	Armanian (2014)	EN	1224.7	30.4	-	-	-	-	-	100.0	0.0
5	Arora (2017)	EN	1700.0	32.9	-	82.0	-	9.3	22.7	100.0	0.0
6	Barrington (2016)	EN	1095.5	28.2	-	55.7	-	-	-	92.4	-
7	Bin-Nun (2005)	EN	1131.4	29.5	8.0	78.0	-	-	20.0	61.4	24.1
8	Braga (2011)	EN	1173.7	29.4	8.0	51.5	-	-	23.0	-	0.0
9	Chowdhury (2016)	EN	1324.0	31.5	-	-	-	-	-	-	0.0
10	Chrzanowska-Lisiewska (2012)	EN	1257.8	29.5	6.6	72.3	-	-	100.0	0.0	100.0
11	Costalos (2003)	EN	1648.1	31.4	-	44.8	-	15.0	-	0.0	100.0
12	Costeloe (2016)	EN	1041.0	28.0	-	52.7	-	-	-	46.1	34.1
13	Dani (2002)	EN	1334.9	30.8	-	79.3	45.5	17.3	-		35.0
14	Dashti (2014)	EN	1406.4	31.2	9.2	82.4	50.0	-	-	34.6	5.9
15	Dasopoulou (2013)	EN	2003.2	34.0	9.0	78.4	-	-	-	0.0	100.0
16	Demirel (2013)	EN	1147.4	29.3	6.0	80.4	-	-	-	-	-
17	Deng (2010)	CHN	1628.7	32.8	-	-	-	-	-	-	-
67	Sreenivasa (2015)	EN	1465.0	31.3	-	-	-	-	10.0	-	0.0
68	Stratiki (2007)	EN	1500.0	30.8	9.0	36.0	13.3	-	-	0.0	100.0
69	Tewari (2015)	EN	1363.0	30.0	-	60.7	-	13.2	-	-	0.0
70	Totsu (2014)	EN	1007.7	28.6	7.5	68.6	-	-	-	-	-
71	Underwood (2009)	EN	1417.8	29.7	8.0	70.0	-	-	-	-	17.8
72	van Neikerk (2014)	EN	1009.0	28.7	-	78.0	-	25.7	-	0.0	0.0
73	van Neikerk (2015)	EN	972.0	28.7	-	74.0	-	18.2	-	100.0	0.0
74	Wejryd (2018)	EN	733.0	25.5	-	-	-	-	-	-	0.0
75	Westerbeek (2011)	EN	1273.8	29.6	-	44.2	-	-	-	62.8	-
76	Xiao-yuan (2007)	CHN	1745.0	31.0	-	-	-	-	-	-	4.0

Ref #	Study (year)	LN	BW (g)	GA (w)	APGAR M5	% C-Sec	% MV	% IUGR	% SGA	% MM fed	% FM fed
1	Akin (2014)	EN	1298.5	29.9	9.0	90.0	6.0	-	-	58.0	0.0
2	Al-Hosni (2012)	EN	778.5	25.7	7.5	51.5	-	-	0.0	-	0.0
3	Amini (2017)	EN	1153.3	29.6	-	-	-	-	-	-	-
4	Armanian (2014)	EN	1224.7	30.4	-	-	-	-	-	100.0	0.0
5	Arora (2017)	EN	1700.0	32.9	-	82.0	-	9.3	22.7	100.0	0.0
6	Barrington (2016)	EN	1095.5	28.2	-	55.7	-	-	-	92.4	-
7	Bin-Nun (2005)	EN	1131.4	29.5	8.0	78.0	-	-	20.0	61.4	24.1
8	Braga (2011)	EN	1173.7	29.4	8.0	51.5	-	-	23.0	-	0.0
9	Chowdhury (2016)	EN	1324.0	31.5	-	-	-	-	-	-	0.0
10	Chrzanowska-Lisiewska (2012)	EN	1257.8	29.5	6.6	72.3	-	-	100.0	0.0	100.0
11	Costalos (2003)	EN	1648.1	31.4	-	44.8	-	15.0	-	0.0	100.0
12	Costeloe (2016)	EN	1041.0	28.0	-	52.7	-	-	-	46.1	34.1
13	Dani (2002)	EN	1334.9	30.8	-	79.3	45.5	17.3	-		35.0
14	Dashti (2014)	EN	1406.4	31.2	9.2	82.4	50.0	-	-	34.6	5.9
15	Dasopoulou (2013)	EN	2003.2	34.0	9.0	78.4	-	-	-	0.0	100.0
16	Demirel (2013)	EN	1147.4	29.3	6.0	80.4	-	-	-	-	-
17	Deng (2010)	CHN	1628.7	32.8	-	-	-	-	-	-	-
77	Xu (2016)	EN	1951.9	33.0	-	-	-	-	-	0.0	100.0
78	Yang (2011)	CHN	-	-	-	-	-	-	-	0.0	100.0
79	Zhou (2012)	CHN	1891.3	34.3	-	-	-	-	-	-	-

Footnote: LN = Language, EN = English, CHN = Chinese, POL = Polish, SPN = Spanish, BW = birth weight, GA = gestational age (weeks), APGAR M5 = APGAR score at 5-minute, % C-Sec = percent infants delivered by caesarean section, % MV = percent infants with mechanical ventilation support, % IUGR = percent infants with intrauterine growth restriction, % SGA = percent infants small for gestational age, % MM fed = percent infant exclusively fed with mother's milk, % FM fed = percent infant exclusively fed with formula milk.

Appendix 6. 5. Treatment characteristics of trials and outcomes included in the network meta-analysis

Ref #	Study (year)	Arm 1	n	Probiotic formulation	Trt duration (weeks)	Arm 2	n	Probiotic formulation	Trt duration (weeks)	Arm 3	n	Probiotic formulation	Trt duration (weeks)	Arm 4	n	Outcomes
1	Akin (2014)	Lactoferrin	22	NA	NR	Placebo	25	NA	-	-	-	-	-	-	-	ACM, NEC, NRM, LOS, TFF, HOS
2	Al-Hosni (2012)	Multi-strain probiotics	50	L. rhamnosus B. infantis	up to 34 wks PMA or discharge	Control/No treatment	51	NA	-	-	-	-	-	-	-	ACM, NEC, LOS, WT
3	Amini (2017)	Synbiotics	57	L. acidophilus, L. rhamnosus; L. bulgaricus; L. casei B. infantis S. Thermophilus	NR	Control/No treatment	58	NA	-	-	-	-	-	-	-	NEC, TFF, HOS
4	Armanian (2014)	GOS & FOS	25	NA	NR	Placebo	50	NA	-	-	-	-	-	-	-	ACM, NEC, LOS, FIT, TFF, HOS, WT
5	Arora (2017)	Multi-strain probiotics	75	L. rhamnosus; L. acidophilus; B. longum S. boulardii	2	Control/No treatment	75	NA	-	-	-	-	-	-	-	ACM, NEC, NRM, LOS, TFF, HOS
6	Barrington (2016)	Lactoferrin	40	NA	up to 36 wks PMA or discharge	Control/No treatment	39	NA	-	-	-	-	-	-	-	ACM, NEC, LOS, TFF
7	Bin-Nun (2005)	Multi-strain probiotics	72	B. infantis; B. bifidus S. Thermophilus	up to 36 wks PMA	Placebo	73	NA	-	-	-	-	-	-	-	ACM, NEC, NRM, LOS, TFF
8	Braga (2011)	Multi-strain probiotics	119	L. casei B. breve	4	Control/No treatment	112	NA	-	-	-	-	-	-	-	ACM, NEC, LOS, TFF
9	Chowdhury (2016)	Multi-strain probiotics	52	L. rhamnosus; L. acidophilus; L. casei; B. infantis; B. bifidum; B. longum;	until discharge	Control/No treatment	50	NA	-	-	-	-	-	-	-	NEC, TFF, HOS

Ref #	Study (year)	Arm 1	n	Probiotic formulation	Trt duration (weeks)	Arm 2	n	Probiotic formulation	Trt duration (weeks)	Arm 3	n	Probiotic formulation	Trt duration (weeks)	Arm 4	n	Outcomes
10	Chrzanowska-Lisiewska (2012)	Single strain probiotics	21	L. rhamnosus	6	Preterm Formula	26	NA	-	-	-	-	-	-	-	NEC, LOS, HOS
11	Costalos (2003)	Single strain probiotics	51	S. boulardii	≥ 4	Preterm Formula	36	NA	-	-	-	-	-	-	-	NEC, LOS, FIT, TFF
12	Costeloe (2016)	Single strain probiotics	650	B. breve	up to 36 weeks PMA or discharge	Placebo	660	NA	-	-	-	-	-	-	-	ACM, NEC, NRM, LOS, TFF, HOS
13	Dani (2002)	Single strain probiotics	295	L. rhamnosus	until discharge	Placebo	290	NA	-	-	-	-	-	-	-	ACM, NEC, NRM, LOS
14	Dashti (2014)	Multi-strain probiotics	69	L. acidophilus; L. rhamnosus; L. bulgaricus; L. casei B. breve; B. longum; S. thermophilus	NR	Placebo	67	NA	-	-	-	-	-	-	-	ACM, NEC, LOS, TFF, HOS
15	Dasopoulou (2013)	GOS & FOS	85	NA	2	Preterm Formula	82	NA	-	-	-	-	-	-	-	NEC, LOS, WT
16	Demirel (2013)	Single strain probiotics	135	S. boulardii	until discharge	Control/No treatment	136	NA	-	-	-	-	-	-	-	ACM, NEC, LOS, FIT, TFF, HOS, WT
17	Deng (2010)	Multi-strain probiotics	63	L. acidophilus B. longum E. faecalis	2	Control/No treatment	62	NA	-	-	-	-	-	-	-	ACM, NEC, NRM
18	Di (2010)	Multi-strain probiotics	182	B. subtilis E. faecium	until discharge	Control/No treatment	173	NA	-	-	-	-	-	-	-	ACM, NEC, NRM
19	Dilli (2015)	Single strain probiotics	100	B. lactis	8	Synbiotics	100	B. lactis	8	FOS	100	NA	8	Placebo	100	ACM, NEC, NRM, LOS, FIT, TFF, HOS, WT

Ref #	Study (year)	Arm 1	n	Probiotic formulation	Trt duration (weeks)	Arm 2	n	Probiotic formulation	Trt duration (weeks)	Arm 3	n	Probiotic formulation	Trt duration (weeks)	Arm 4	n	Outcomes
20	Dongol Singh (2017)	Single strain probiotics	37	L. casei	Until full enteral feeding	Placebo	35	NA	-	-	-	-	-	-	-	ACM, NEC
21	Dutta (2014)	Multi-strain probiotics (high dose)	38	L. rhamnosus; L. acidophilus B. longum S. boulardii	3	Multi-strain probiotics (high dose)	38	L. rhamnosus; L. acidophilus B. longum S. boulardii	2	Multi-strain probiotics (low dose)	38	L. rhamnosus; L. acidophilus B. longum S. boulardii	3	Placebo	35	ACM, NEC, LOS
22	Fernández-Carrocera (2013)	Multi-strain probiotics	75	L. rhamnosus; L. acidophilus; L. casei; L. plantarum B. infantis S. thermophilus	until discharge	Control/No treatment	75	NA	-	-	-	-	-	-	-	ACM, NEC, LOS, TFF, HOS
23	Fujii (2006)	Single strain probiotics	11	B. breve	until discharge	Placebo	8	NA	-	-	-	-	-	-	-	NEC, LOS, HOS
24	Güney-Varal (2017)	Synbiotics	70	L. casei; L. plantorum B. animalis	until discharge	Control/No treatment	40	NA	-	-	-	-	-	-	-	ACM, NEC, NRM, LOS, FIT, TFF, HOS
25	Hariharan (2016)	Multi-strain probiotics	93	L. acidophilus B. bifidum S. boulardii	6	Control/No treatment	103	NA	-	-	-	-	-	-	-	ACM, NEC, LOS, FIT, TFF
26	Hays (2016)	Multi-strain probiotics	47	B. Lactis; B. longum	4 to 6	Single strain probiotics	48	B. longum	4 to 6	Single strain probiotics	50	B. lactis	4 to 6	Placebo	52	ACM, NEC, LOS, TFF
27	Hernandez-Enriquez (2016)	Single strain probiotics	24	L. reuteri	3	Control/No treatment	20	NA	-	-	-	-	-	-	-	NEC, TFF, HOS
28	Hikaru (2010)	Single strain probiotics	108	B. breve	Until discharge	Placebo	100	NA	-	-	-	-	-	-	-	ACM, LOS, TFF, HOS, WT

Ref #	Study (year)	Arm 1	n	Probiotic formulation	Trt duration (weeks)	Arm 2	n	Probiotic formulation	Trt duration (weeks)	Arm 3	n	Probiotic formulation	Trt duration (weeks)	Arm 4	n	Outcomes
29	Hua (2014)	Multi-strain probiotics	119	L. bulgaricus B. longum S. thermophilus	2	Control/No treatment	138	NA	-	-	-	-	-	-	-	ACM, NEC, LOS, FIT
30	Huang (2009)	Single strain probiotics	95	B. adolescentis	1 to 2	Control/No treatment	88	NA	-	-	-	-	-	-	-	ACM, NEC, NRM
31	Indrio (2017)	Single strain probiotics	30	L. reuteri	4	Placebo	30	NA	-	-	-	-	-	-	-	TFF, HOS, WT
32	Jacobs (2013)	Multi-strain probiotics	548	B. infantis; B. lactis S. thermophilus	term corrected age/until discharge	Placebo	551	NA	-	-	-	-	-	-	-	ACM, NEC, NRM, LOS, TFF, HOS, WT
33	Kanic (2015)	Multi-strain probiotics	40	L. acidophilus B. infantis E. faecium	until discharge	Control/No treatment	40	NA	-	-	-	-	-	-	-	ACM, NEC, LOS, TFF, HOS
34	Kaur (2015)	Lactoferrin	63	NA	4	Placebo	67	NA	-	-	-	-	-	-	-	ACM, LOS, FIT
35	Ke (2008)	Multi-strain probiotics	438	L. acidophilus B. longum E. faecalis	until discharge	Control/No treatment	446	NA	-	-	-	-	-	-	-	ACM, NEC, NRM
36	Kitajima (1997)	Single strain probiotics	45	B. breve	4	Control/No treatment	46	NA	-	-	-	-	-	-	-	ACM, NEC, NRM, LOS
37	Lin (2005)	Multi-strain probiotics	180	L. acidophilus B. infantis	until discharge	Control/No treatment	187	NA	-	-	-	-	-	-	-	ACM, NEC, LOS, HOS
38	Lin (2008)	Multi-strain probiotics	222	L. acidophilus B. bifidum	6	Placebo	221	NA	-	-	-	-	-	-	-	ACM, NEC, NRM, LOS, TFF, HOS
39	Manzoni (2006)	Single strain probiotics	39	L. rhamnosus	6/discharge	MBM or DBM	41	NA	-	-	-	-	-	-	-	ACM, NEC, LOS, TFF, HOS
40	Manzoni (2014)	Synbiotics	238	L. rhamnosus	4 to 6	Lactoferrin	247	NA	4 to 6	Placebo	258	-	-	-	-	ACM, NEC, LOS, TFF, HOS

Ref #	Study (year)	Arm 1	n	Probiotic formulation	Trt duration (weeks)	Arm 2	n	Probiotic formulation	Trt duration (weeks)	Arm 3	n	Probiotic formulation	Trt duration (weeks)	Arm 4	n	Outcomes
41	Mihatsch (2010)	Single strain probiotics	91	B. lactis	until discharge	Placebo	89	NA	-	-	-	-	-	-	-	ACM, NEC, NRM, LOS, TFF
42	Millar (1993)	Single strain probiotics	10	L. rhamnosus	2 wks	Placebo	10	NA	-	-	-	-	-	-	-	LOS, FIT
43	Modi (2010)	GOS & FOS	73	NA	40 wks corrected age/until discharge	Control/No treatment	81	NA	-	-	-	-	-	-	-	NEC, TFF
44	Mohan (2008)	Single strain probiotics	37	B. lactis	3 wks	Placebo	32	NA	-	-	-	-	-	-	-	NEC, WT
45	Nandhini (2015)	Synbiotics	108	L. acidophilus; L. rhamnosus; L. casei; L. bulgaricus; L. plantaris B. longum; B. breve; B. infantis	1 wks	Control/No treatment	110	NA	-	-	-	-	-	-	-	ACM, NEC, LOS, HOS
46	Ochoa (2015)	Lactoferrin	95	NA	4 wks	Placebo	95	NA	-	-	-	-	-	-	-	ACM, LOS
47	Oncel (2014)	Single strain probiotics	200	L. reuteri	4 wks	Placebo	200	NA	-	-	-	-	-	-	-	ACM, NEC, NRM, LOS, FIT, TFF, HOS
48	Patole (2014)	Single strain probiotics	79	B. breve	corrected age of 37 wks	Placebo	80	NA	-	-	-	-	-	-	-	ACM, NEC, LOS, TFF, WT
49	Punnahitananda (2006)	Multi-strain probiotics	89	L. acidophilus B. infantis	4/until discharge	Placebo	85	NA	-	-	-	-	-	-	-	ACM, NRM
50	Qiao (2016)	Multi-strain probiotics	30	L. acidophilus B. longum E. faecium	2 wks	Placebo	30	NA	-	-	-	-	-	-	-	FIT

Ref #	Study (year)	Arm 1	n	Probiotic formulation	Trt duration (weeks)	Arm 2	n	Probiotic formulation	Trt duration (weeks)	Arm 3	n	Probiotic formulation	Trt duration (weeks)	Arm 4	n	Outcomes
51	Ren (2010)	Multi-strain probiotics	80	B. infantis L. acidophilus B. cereus E. Faecalis	1 to 2 wks	Control/No treatment	70	NA	-	-	-	-	-	-	-	ACM, NEC, NRM
52	Reuman (1986)	Single strain probiotics	15	L. acidophilus	4 wks	Placebo	15	NA	-	-	-	-	-	-	-	ACM, NEC, HOS
53	Rojas (2012)	Single strain probiotics	372	L. reuteri	until discharge	Placebo	378	NA	-	-	-	-	-	-	-	ACM, NEC, LOS, FIT, HOS
54	Romeo (2011)	Single strain probiotics	83	L. reuteri	6/until discharge	Single strain probiotics	83	L. rhamnosus	6/until discharge	Control/No treatment	83	-	-	-	-	LOS, FIT, TFF, HOS
55	Rouge (2009)	Multi-strain probiotics	45	L. rhamnosus B. longum	until discharge	Placebo	49	NA	-	-	-	-	-	-	-	ACM, NEC, NRM, LOS, HOS
56	Roy (2014)	Multi-strain probiotics	56	L. acidophilus B. bifidum; B. lactis; B. longum	6/until discharge	Placebo	56	NA	-	-	-	-	-	-	-	ACM, NEC, LOS, TFF, HOS
57	Sadowska-Krawczenko (2012)	Single strain probiotics	30	L. rhamnosus	Until discharge	Placebo	25	NA	-	-	-	-	-	-	-	ACM, NEC, LOS, HOS
58	Saengtawesin (2014)	Multi-strain probiotics	31	L. acidophilus B. bifidum	6/until discharge	Control/No treatment	29	NA	-	-	-	-	-	-	-	ACM, NEC, NRM, LOS, TFF, HOS
59	Samanta (2009)	Multi-strain probiotics	91	L. acidophilus B. infantis; B. lactis; B. longum	until discharge	MBM	95	NA	-	-	-	-	-	-	-	ACM, NEC, LOS, TFF, HOS
60	Sari (2011)	Single strain probiotics	121	L. sporogenes	until discharge	Placebo	121	NA	-	-	-	-	-	-	-	ACM, NEC, NRM, LOS, FIT, TFF, HOS

Ref #	Study (year)	Arm 1	n	Probiotic formulation	Trt duration (weeks)	Arm 2	n	Probiotic formulation	Trt duration (weeks)	Arm 3	n	Probiotic formulation	Trt duration (weeks)	Arm 4	n	Outcomes
61	Serce (2013) A	Single strain probiotics	104	S. boulardii	until discharge	Placebo	104	NA	-	-	-	-	-	-	-	ACM, NEC, NRM, LOS, TFF, HOS
62	Serce (2013) B	Synbiotics	43	L. rhamnosus; L. casei; L. plantaris B. lactis	until discharge	Placebo	60	NA	-	-	-	-	-	-	-	NEC, NRM, LOS, TFF
63	Shadkam (2015)	Single strain probiotics	30	L. reuteri	Until full enteral feeding	Placebo	30	NA	-	-	-	-	-	-	-	ACM, NEC, LOS, TFF, WT
64	Shashidhar (2017)	Multi-strain probiotics	48	L. acidophilus; L. rhamnosus B. longum S. boulardii	4	Placebo	48	NA	-	-	-	-	-	-	-	ACM, NEC, TFF, HOS
65	Sherman (2016)	Lactoferrin	59	NA	4	Placebo	60	NA	-	-	-	-	-	-	-	ACM, NEC, LOS, TFF, HOS, WT
66	Sinha (2015)	Multi-strain probiotics	668	L. acidophilus; L. plantarum; L. casei; B. breve; B. infantis; B. longum S. thermophilus	4	Placebo	672	NA	-	-	-	-	-	-	-	ACM, LOS
67	Sreenivasa (2015)	Synbiotics	100	L. acidophilus B. bifidum; B. longum; S. thermophilus	Until full enteral feeding	Control/No treatment	100	NA	-	-	-	-	-	-	-	ACM, NEC, LOS, TFF, HOS
68	Stratiki (2007)	Single strain probiotics	41	B. lactis	until discharge	Preterm Formula	36	NA	-	-	-	-	-	-	-	ACM, NEC, LOS, TFF
69	Tewari (2015)	Single strain probiotics	123	B. clausii	6	Placebo	121	NA	-	-	-	-	-	-	-	ACM, NEC, LOS, FIT

Ref #	Study (year)	Arm 1	n	Probiotic formulation	Trt duration (weeks)	Arm 2	n	Probiotic formulation	Trt duration (weeks)	Arm 3	n	Probiotic formulation	Trt duration (weeks)	Arm 4	n	Outcomes
70	Totsu (2014)	Single strain probiotics	153	B. bifidum	until weight reach 2 kg	Placebo	130	NA	-	-	-	-	-	-	-	ACM, NEC, NRM, LOS, TFF, HOS, WT
71	Underwood (2009)	Synbiotics	31	L. acidophilus B. bifidum; B. infantis; B. longum	4/until discharge	Synbiotics	30	L. rhamnosus	4 / until discharge	Placebo	29	-	-	-	-	NEC, FIT
72	van Neikerk (2014)	Multi-strain probiotics	37	L. rhamnosus B. infantis	4	Placebo	37	NA	-	-	-	-	-	-	-	ACM, NEC, NRM, LOS, TFF
73	van Neikerk (2015)	Multi-strain probiotics	54	L. rhamnosus B. infantis	4	Placebo	54	NA	-	-	-	-	-	-	-	ACM, NEC, NRM, LOS, TFF
74	Wejryd (2018)	Single strain probiotics	68	L. reuteri	4	Placebo	66	NA	-	-	-	-	-	-	-	ACM, NEC, LOS, TFF
75	Westerbeek (2011)	GOS & FOS	55	NA	4	Placebo	58	NA	-	-	-	-	-	-	-	ACM, NEC, LOS, TFF, HOS
76	Xiao-yuan (2007)	Multi-strain probiotics	276	L. acidophilus B. infantis E. faecalis	until discharge	Control/No treatment	248	NA	-	-	-	-	-	-	-	ACM, NEC, NRM, LOS
77	Xu (2016)	Single strain probiotics	51	S. boulardii	4/until discharge	Control/No treatment	49	NA	-	-	-	-	-	-	-	NEC, LOS, HOS
78	Yang (2011)	Multi-strain probiotics	31	L. acidophilus B. longum E. faecalis	Until full enteral feeding	Preterm Formula	31	NA	-	-	-	-	-	-	-	ACM, NEC, NRM, TFF
79	Zhou (2012)	Multi-strain probiotics	75	L. acidophilus B. infantis E. faecalis B. cereus	1	Control/No treatment	50	NA	-	-	-	-	-	-	-	ACM, NEC, NRM

Footnote: n = number analyzed (as reported in study publication); wks = weeks

Appendix 6. 6. Summary of risk of bias assessments for included trials

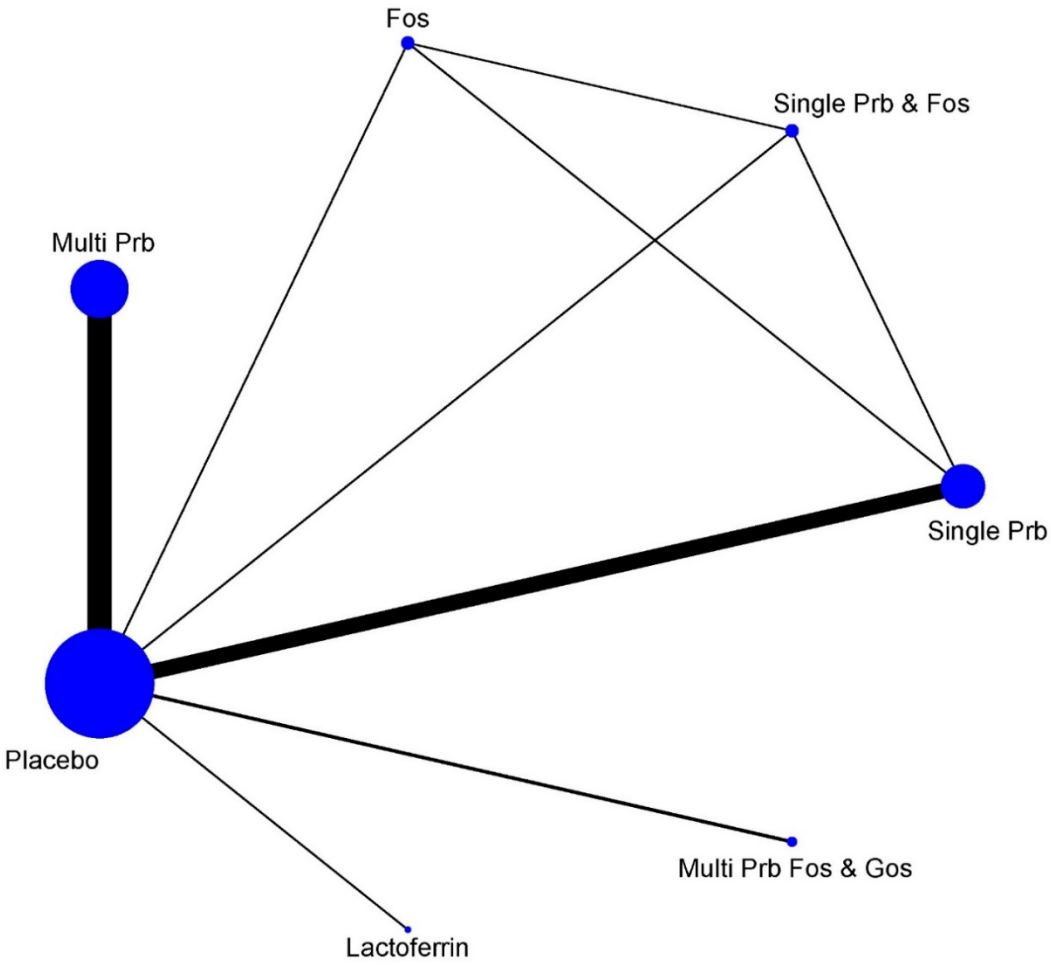
Ref #	Study (year)	Sequence generation	Allocation concealment	Parents blinded	Health care providers blinded	Data collectors/ outcome assessors blinded	Missingness (> 5%: high)	funding
1	Akin (2014)	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	non-industry
2	Al-Hosni (2012)	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	no funding statement
3	Amini (2017)	High Risk	High Risk	High Risk	High Risk	High Risk	Low Risk	non-industry
4	Armanian (2014)	Low Risk	High Risk	Low Risk	High Risk	High Risk	High Risk	non-industry
5	Arora (2017)	Low Risk	High Risk	High Risk	High Risk	High Risk	Low Risk	no funding
6	Barrington (2016)	Low Risk	Low Risk	Low Risk	Low Risk	High Risk	Low Risk	any industry
7	Bin-Nun (2005)	Low Risk	High Risk	Low Risk	Low Risk	Low Risk	Low Risk	non-industry
8	Braga (2011)	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	no funding statement
9	Chowdhury (2016)	High Risk	High Risk	High Risk	High Risk	High Risk	High Risk	no funding statement
10	Chrzanowska-Lisiewska (2012)	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	no funding statement
11	Costalos (2003)	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	no funding statement
12	Costeloe (2016)	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	non-industry
13	Dani (2002)	Low Risk	Low Risk	High Risk	High Risk	Low Risk	Low Risk	no funding statement
14	Dashti (2014)	High Risk	High Risk	Low Risk	Low Risk	High Risk	Low Risk	no funding statement
15	Dasopoulou (2013)	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	High Risk	any industry
16	Demirel (2013)	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	no funding statement
17	Deng (2010)	High Risk	High Risk	High Risk	High Risk	High Risk	Low Risk	no funding
18	Di (2010)	High Risk	High Risk	Low Risk	Low Risk	High Risk	Low Risk	no funding
19	Dilli (2015)	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	no funding statement
20	Dongol Singh (2017)	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	no funding statement
21	Dutta (2014)	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	any industry
22	Fernández-Carrocera (2013)	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	no funding statement
23	Fujii (2006)	Low Risk	Low Risk	High Risk	High Risk	High Risk	Low Risk	non-industry

Ref #	Study (year)	Sequence generation	Allocation concealment	Parents blinded	Health care providers blinded	Data collectors/ outcome assessors blinded	Missingness (> 5%: high)	funding
24	Güney-Varal (2017)	High Risk	High Risk	High Risk	High Risk	High Risk	High Risk	no funding statement
25	Hariharan (2016)	Low Risk	High Risk	High Risk	High Risk	High Risk	Low Risk	no funding statement
26	Hays (2016)	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	any industry
27	Hernandez-Enriquez (2016)	High Risk	Low Risk	High Risk	High Risk	High Risk	Low Risk	no funding statement
28	Hikaru (2010)	Low Risk	Low Risk	Low Risk	Low Risk	High Risk	Low Risk	any industry
29	Hua (2014)	Low Risk	High Risk	Low Risk	Low Risk	High Risk	Low Risk	non-industry
30	Huang (2009)	High Risk	High Risk	High Risk	High Risk	High Risk	Low Risk	no funding
31	Indrio (2017)	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	any industry
32	Jacobs (2013)	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	non-industry
33	Kanic (2015)	High Risk	High Risk	Low Risk	High Risk	High Risk	Low Risk	no funding statement
34	Kaur (2015)	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	High Risk	no funding statement
35	Ke (2008)	High Risk	High Risk	Low Risk	Low Risk	High Risk	Low Risk	no funding
36	Kitajima (1997)	Low Risk	High Risk	High Risk	High Risk	High Risk	Low Risk	no funding statement
37	Lin (2005)	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	non-industry
38	Lin (2008)	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	non-industry
39	Manzoni (2006)	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	no funding statement
40	Manzoni (2014)	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	any industry
41	Mihatsch (2010)	Low Risk	Low Risk	Low Risk	High Risk	High Risk	Low Risk	no funding statement
42	Millar (1993)	High Risk	High Risk	Low Risk	Low Risk	Low Risk	Low Risk	any industry
43	Modi (2010)	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	no funding statement
44	Mohan (2008)	Low Risk	Low Risk	Low Risk	Low Risk	High Risk	Low Risk	any industry
45	Nandhini (2015)	Low Risk	Low Risk	High Risk	High Risk	High Risk	Low Risk	no funding statement
46	Ochoa (2015)	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	non-industry
47	Oncel (2014)	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	no funding statement

Ref #	Study (year)	Sequence generation	Allocation concealment	Parents blinded	Health care providers blinded	Data collectors/ outcome assessors blinded	Missingness (> 5%: high)	funding
48	Patole (2014)	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	non-industry
49	Punnahitananda (2006)	High Risk	High Risk	High Risk	High Risk	High Risk	Low Risk	no funding statement
50	Qiao (2016)	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	no funding statement
51	Ren (2010)	Low Risk	High Risk	High Risk	High Risk	High Risk	Low Risk	no funding
52	Reuman (1986)	High Risk	High Risk	Low Risk	Low Risk	Low Risk	Low Risk	no funding statement
53	Rojas (2012)	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	non-industry
54	Romeo (2011)	Low Risk	High Risk	High Risk	High Risk	High Risk	Low Risk	no funding statement
55	Rouge (2009)	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	non-industry
56	Roy (2014)	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	no funding statement
57	Sadowska-Krawczenko (2012)	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	no funding statement
58	Saengtawesin (2014)	Low Risk	Low Risk	High Risk	High Risk	High Risk	Low Risk	non-industry
59	Samanta (2009)	Low Risk	High Risk	Low Risk	Low Risk	High Risk	Low Risk	no funding statement
60	Sari (2011)	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	High Risk	no funding statement
61	Serce (2013) A	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	any industry
62	Serce (2013) B	Low Risk	Low Risk	High Risk	High Risk	High Risk	Low Risk	no funding statement
63	Shadkam (2015)	Low Risk	Low Risk	Low Risk	Low Risk	High Risk	Low Risk	non-industry
64	Shashidhar (2017)	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	High Risk	no funding
65	Sherman (2016)	Low Risk	Low Risk	Low Risk	High Risk	High Risk	Low Risk	any industry
66	Sinha (2015)	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	non-industry
67	Sreenivasa (2015)	Low Risk	High Risk	High Risk	High Risk	High Risk	Low Risk	no funding
68	Stratiki (2007)	High Risk	Low Risk	Low Risk	Low Risk	High Risk	High Risk	any industry
69	Tewari (2015)	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	no funding statement
70	Totsu (2014)	Low Risk	Low Risk	High Risk	Low Risk	Low Risk	High Risk	no funding
71	Underwood (2009)	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	High Risk	non-industry

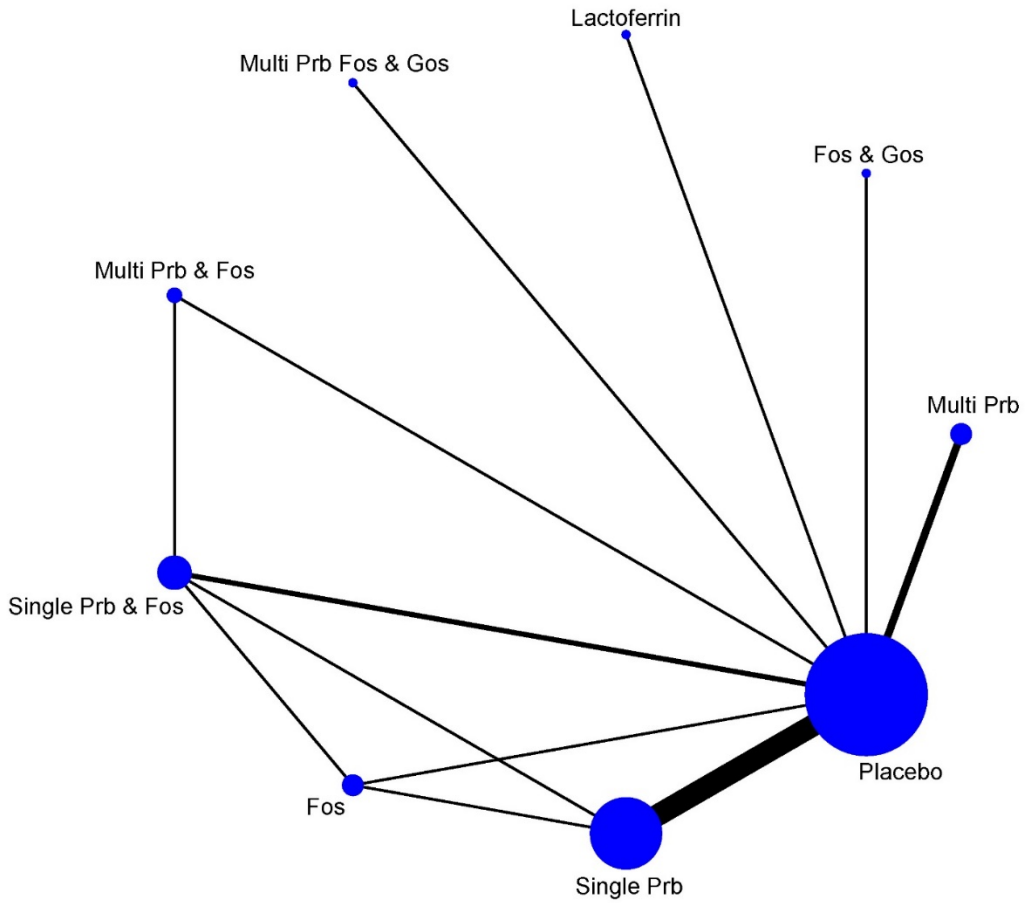
Ref #	Study (year)	Sequence generation	Allocation concealment	Parents blinded	Health care providers blinded	Data collectors/ outcome assessors blinded	Missingness (> 5%: high)	funding
72	van Neikerk (2014)	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	any industry
73	van Neikerk (2015)	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	any industry
74	Wejryd (2018)	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	non-industry
75	Westerbeek (2011)	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	High Risk	any industry
76	Xiao-yuan (2007)	High Risk	High Risk	High Risk	High Risk	High Risk	Low Risk	no funding
77	Xu (2016)	Low Risk	High Risk	Low Risk	Low Risk	High Risk	High Risk	no funding
78	Yang (2011)	High Risk	High Risk	High Risk	High Risk	High Risk	Low Risk	no funding
79	Zhou (2012)	High Risk	High Risk	High Risk	High Risk	High Risk	Low Risk	no funding

Appendix 6. 7. Networks of treatment comparisons for the secondary outcomes



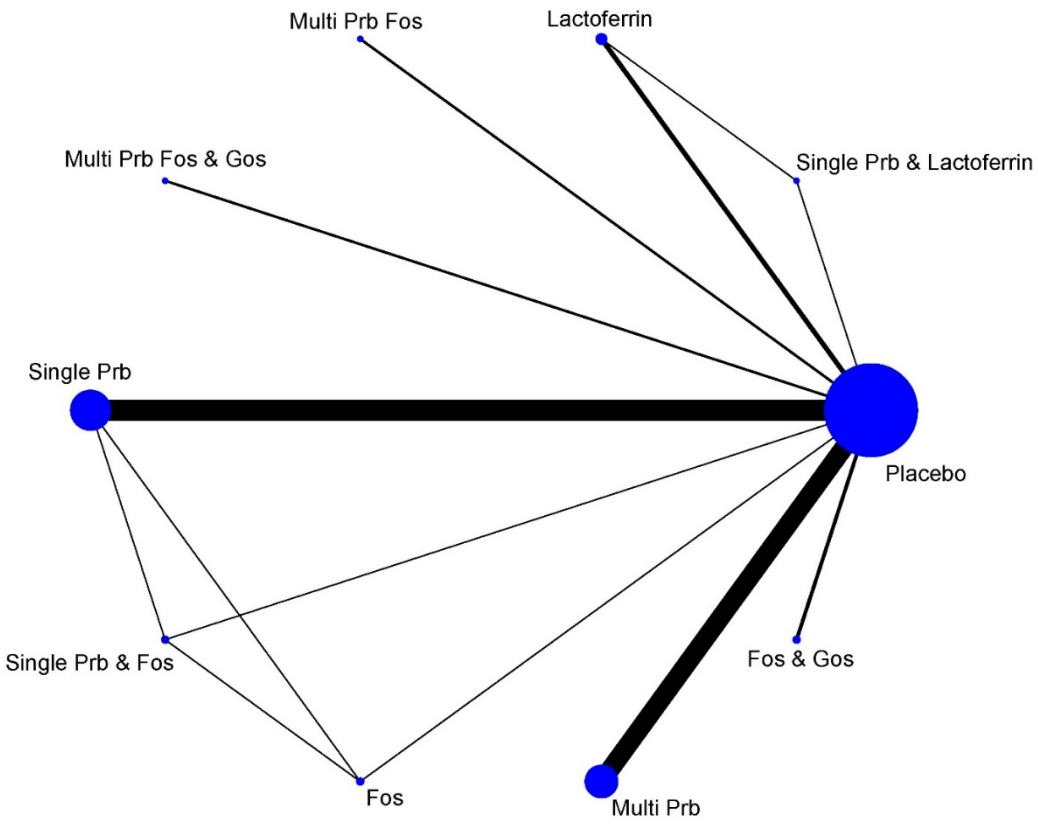
e-figure 1: Network of eligible comparisons for NEC-related mortality

The size of the node corresponds to the number of infants randomized to that intervention. The interventions directly compared are linked with a line; the thickness of the line corresponds to the number of studies that assessed the comparison. Fos = Fructo-oligosaccharides, Gos = galacto-oligosaccharides; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.



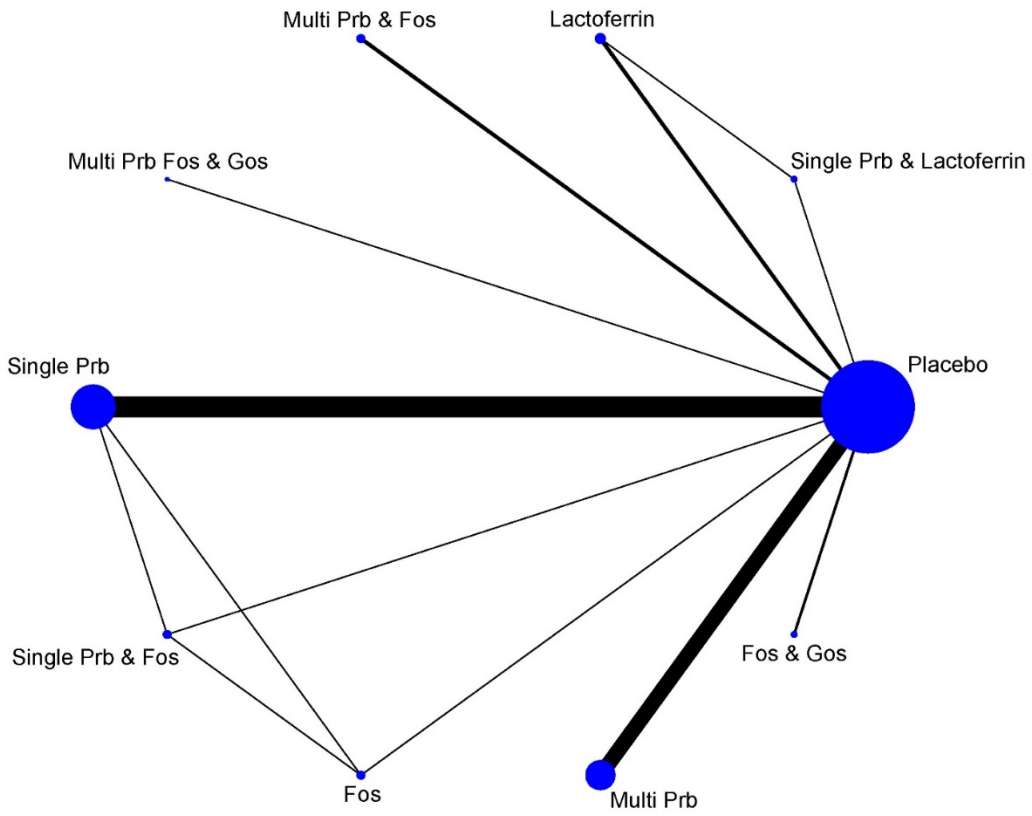
e-figure 2: Network of eligible comparisons for feed intolerance

The size of the node corresponds to the number of infants randomized to that intervention. The interventions directly compared are linked with a line; the thickness of the line corresponds to the number of studies that assessed the comparison. Fos = Fructo-oligosaccharides, Gos = galacto-oligosaccharides; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.



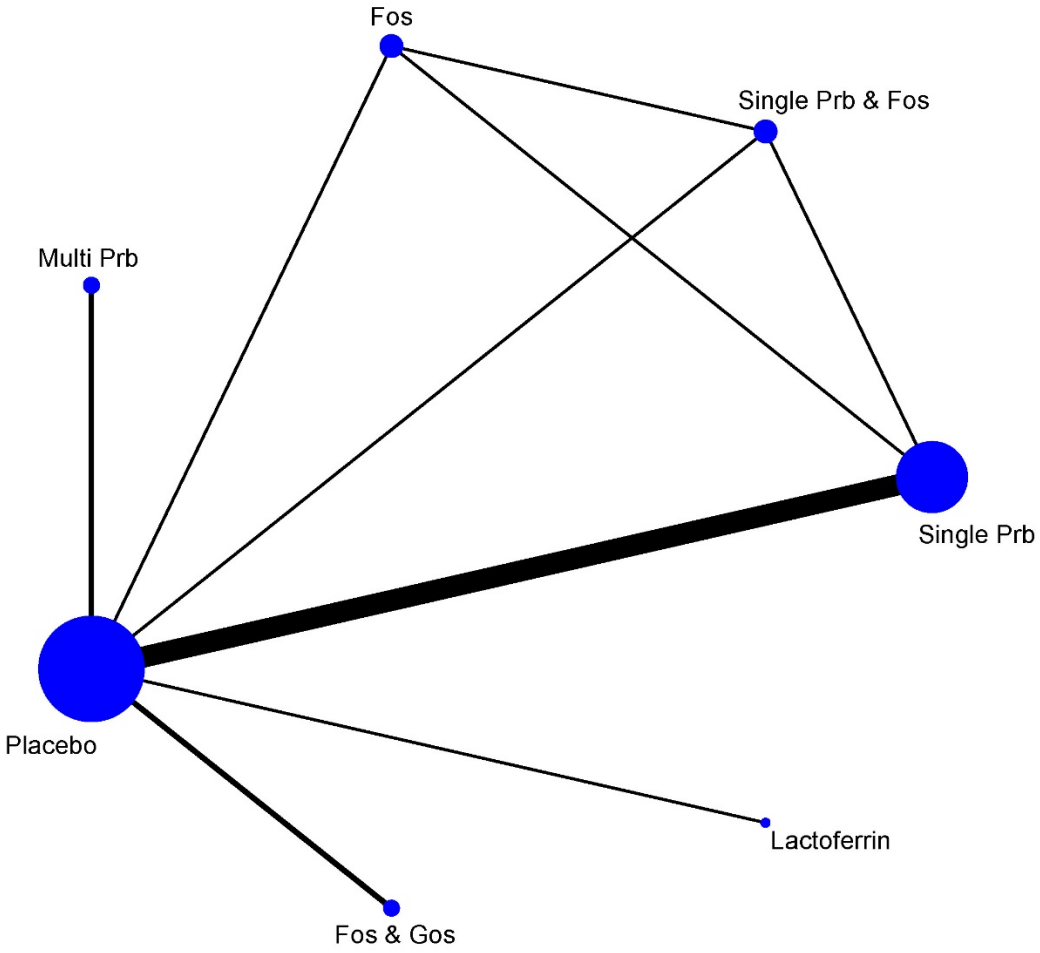
e-figure 3: Network of eligible comparisons for time to reach full enteral feeding

The size of the node corresponds to the number of infants randomized to that intervention. The interventions directly compared are linked with a line; the thickness of the line corresponds to the number of studies that assessed the comparison. Fos = Fructo-oligosaccharides, Gos = galacto-oligosaccharides; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.



e-figure 4: Network of eligible comparisons for duration of hospital stay

The size of the node corresponds to the number of infants randomized to that intervention. The interventions directly compared are linked with a line; the thickness of the line corresponds to the number of studies that assessed the comparison. Fos = Fructo-oligosaccharides, Gos = galacto-oligosaccharides; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.



e-figure 5: Network of eligible comparisons for weight gain

The size of the node corresponds to the number of infants randomized to that intervention. The interventions directly compared are linked with a line; the thickness of the line corresponds to the number of studies that assessed the comparison. Fos = Fructo-oligosaccharides, Gos = galacto-oligosaccharides; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.

Appendix 6. 8. GRADE presentation of primary and secondary outcomes

e-Table 1: Network meta-analysis results (odds ratio and risk difference and their 95% CIs) and SUCRA values sorted based on GRADE certainty of evidence for the comparisons of active treatments vs. placebo (PLC) for primary and secondary binary outcomes

Outcome	Credibility	Classification	Intervention	OR (95% CI)	SUCRA	% RD (95% CI)
All-cause Mortality	High (Moderate to High certainty)	Inferior to the most effective/ superior to the least effective	Multiple-strain Probiotic	**0.67 (0.52, 0.87)**	48.9	-1.54 (-2.30, -0.58)
		Least effective	Single-strain Probiotic	0.83 (0.67, 1.04)	49.4	-0.80 (-1.64, 0.23)
			Lactoferrin	0.64 (0.33, 1.22)	31.0	-0.28 (-1.52, 1.38)
	Low (Low to very low certainty)	May be the most effective	Multiple-strain Probiotic + Fos & Gos	**0.05 (0.01, 0.41)**	96.8	-4.95 (-5.25, -2.73)
		May be inferior to the most effective/ superior than the least effective	Fos	**0.20 (0.04, 0.89)**	82.6	-4.18 (-5.03, -0.56)
		May be among the least effective	Single-strain Probiotic + Fos	0.30 (0.08, 1.07)	74.3	-3.63 (-4.80, 0.29)
			Single-strain Probiotic + Lactoferrin	0.63 (0.28, 1.41)	48.4	-1.62 (-3.57, 2.53)
			Multiple-strain Probiotic + Fos	0.94 (0.38, 2.32)	25.2	0.20 (3.06, -6.34)
			Fos & Gos	0.95 (0.21, 4.42)	29.5	-0.25 (-4.13, 16.73)
NEC (stage II or more)	High (Moderate to High certainty)	Among the most effective	Single-strain Probiotic + Lactoferrin	**0.04 (0.01, 0.62)**	91.6	-6.15 (-6.44, -1.32)
			Multiple-strain Probiotic	**0.35 (0.26, 0.47)**	57.0	-4.06 (-4.66, -3.25)
		Inferior to the most effective/ superior to the least effective	Single-strain Probiotic	**0.63 (0.46, 0.86)**	33.2	-2.30 (-3.34, -0.93)
		Least effective	Lactoferrin	0.43 (0.18, 1.00)	48.7	-0.76 (-2.90, 2.68)
	Low (Low to very low certainty)	May be among the most effective	Multiple-strain Probiotic + Fos & Gos	**0.09 (0.02, 0.52)**	84.4	-5.76 (-6.33, -2.73)
			Multiple-strain Probiotic + Fos	**0.14 (0.04, 0.49)**	79.9	-5.48 (-6.18, -3.05)
		May be inferior to the most effective/ superior than the least effective	Single-strain Probiotic + Fos	**0.26 (0.09, 0.74)**	64.9	-4.62 (-5.79, -1.45)
		May be among the least effective	Fos	0.81 (0.36, 1.83)	23.4	-1.20 (-3.97, 4.65)
			Fos & Gos	1.73 (0.67, 4.45)	3.2	4.10 (-2.03, 18.72)
Culture proven sepsis	High (Moderate to High certainty)	Most effective	Single-strain Probiotic + Lactoferrin	**0.27 (0.10, 0.72)**	93.1	-11.00 (-14.24, -3.07)
		Inferior to the most effective/ superior than the least effective	Lactoferrin	**0.44 (0.27, 0.74)**	79.1	-5.10 (-8.42, -0.40)
			Single-strain Probiotic	**0.80 (0.65, 0.99)**	35.1	-2.57 (-4.83, 0.13)
		Among the least effective	Single-strain Probiotic + Fos	0.67 (0.25, 1.80)	49.7	-4.97 (-11.72, 11.37)
			Fos	0.85 (0.33, 2.19)	32.9	-2.09 (-10.19, 16.43)

Outcome	Credibility	Classification	Intervention	OR (95% CI)	SUCRA	% RD (95% CI)
	Low (Low to very low certainty)	**May be inferior to the most effective/ superior than the least effective**	Multiple-strain Probiotic	**0.77 (0.62, 0.95)**	39.6	-2.65 (-4.69, -0.27)
		May be among the least effective	Fos & Gos	0.58 (0.29, 1.14)	61.2	-4.86 (-9.43, 2.68)
			Multiple-strain Probiotic + Fos	0.62 (0.30, 1.28)	54.9	-4.75 (-9.82, 4.17)
			Multiple-strain Probiotic + Fos & Gos	0.73 (0.32, 1.64)	43.6	-3.87 (-9.87, 7.58)
Feed intolerance	High (Moderate to High)	**Inferior to the most effective/ superior than the least effective**	Single-strain Probiotic	**0.47 (0.36,0.61)**	45.9	-11.24 (-14.18, -7.71)
		Among the least effective	Fos	0.37 (0.10,1.39)	54.8	-17.47 (-25.43, 10.88)
	Low (Low to very low)	**May be among the most effective**	Multiple-strain Probiotic + Fos & Gos	**0.15 (0.06,0.39)**	86.1	-15.92 (-20.31, -9.19)
		May be inferior to the most effective/ superior than the least effective	Multiple-strain Probiotic	**0.48 (0.30,0.76)**	44.3	-10.56 (-15.46, -3.82)
			Lactoferrin	**0.21 (0.05,0.79)**	75.0	-21.54 (-26.49, -4.63)
		May be among the least effective	Multiple-strain Probiotic + Fos	0.13 (0.01,2.55)	76.3	-24.45 (-28.32, 45.41)
			Single-strain Probiotic + Fos	0.48 (0.17,1.35)	44.2	-14.26 (-23.17, 9.42)
			Fos & Gos	1.00 (0.36,2.78)	13.9	0.01 (-13.32, 24.98)

e-Table 2: Network meta-analysis results (mean difference – MD and their 95% CIs) and SUCRA values sorted based on GRADE certainty of evidence for the comparisons of active treatments vs. placebo (PLC) for time to reach full enteral feeding (days) and duration of hospitalization (days)

Outcome	certainty	Classification	Intervention	MD (95% CI) vs PLC	SUCRA
Time to Reach Full Enteral Feeding (days)	High (Moderate to High)	Inferior to the most effective / superior than the least effective	Single-strain Probiotic	**-1.60 (-2.58, -0.62)**	52.9
		Among the least effective	Lactoferrin	-1.44 (-3.95, 1.07)	49.8
			Fos & Gos	-1.14 (-3.57, 1.28)	43.3
			Single-strain Probiotic + Lactoferrin	-0.97 (-4.33, 2.39)	40.7
	Low (Low to very low)	May be inferior to the most effective / superior than the least effective	Multiple-strain Probiotic	**-2.04 (-3.10, -0.99)**	64.0
			Fos	**-4.35 (-8.23, -0.48)**	89.0
			Multiple-strain Probiotic + Fos & Gos	**-3.31 (-6.50, -0.12)**	79.8
		May be among the least effective	Single-strain Probiotic + Fos	-1.35 (-5.52, 2.82)	47.0
			Multiple-strain Probiotic + Fos	0.46 (-2.63, 3.55)	16.7
Duration of Hospitalization (days)	High (Moderate to High)	Inferior to the most effective / superior than the least effective	Single-strain Probiotic	**-3.58 (-5.88, -1.27)**	68.8
		Among the least effective	Single-strain Probiotic + Lactoferrin	1.98 (-5.43, 9.39)	20.9
			Lactoferrin	1.75 (-4.02, 7.53)	20.2
	Low (Low to very low)	May be inferior to the most effective / superior than the least effective	Multiple-strain Probiotic	**-2.68 (-5.15, -0.21)**	59.7
		May be among the least effective	Multiple-strain Probiotic + Fos & Gos	-9.00 (-22.11, 4.11)	83.5
			Fos	-6.19 (-14.01, 1.62)	79.9
			Fos & Gos	-2.86 (-10.30, 4.58)	57.5
			Single-strain Probiotic + Fos	-2.19 (-10.08, 5.69)	51.1
			Multiple-strain Probiotic + Fos	0.45 (-4.14, 5.05)	28.4

Appendix 6. 9. Network meta-analysis results (league tables) for secondary outcomes

e-Table 3: NMA results for feed intolerance (top half) and reduction in number of days to reach full feeding (bottom half)

Placebo	-	0.48 (0.17,1.35)	**0.47 (0.36,0.61)**	**0.48 (0.30,0.76)**	**0.15 (0.06,0.39)**	0.13 (0.01,2.55)	**0.21 (0.05,0.79)**	1.00 (0.36,2.78)	0.37 (0.10,1.39)
-0.97 (-4.33,2.39)	**SinglePrb & Lactoferrin**	-	-	-	-	-	-	-	-
-1.35 (-5.52,2.82)	-0.38 (-5.75,4.98)	**SinglePrb & Fos**	0.99 (0.34,2.88)	1.01 (0.32,3.17)	0.32 (0.08,1.29)	0.27 (0.01,5.71)	0.44 (0.08,2.38)	2.10 (0.49,9.04)	0.77 (0.17,3.40)
-1.60 (-2.58,-0.62)	-0.64 (-4.14,2.87)	-0.25 (-4.37,3.87)	**SinglePrb**	1.02 (0.61,1.73)	**0.32 (0.12,0.85)**	0.28 (0.01,5.48)	0.44 (0.11,1.73)	2.13 (0.74,6.10)	0.78 (0.20,3.03)
-2.04 (-3.10,-0.99)	-1.07 (-4.60,2.45)	-0.69 (-5.00,3.62)	-0.44 (-1.88,1.01)	**MultiPrb**	**0.31 (0.11,0.89)**	0.27 (0.01,5.49)	0.43 (0.10,1.78)	2.08 (0.68,6.38)	0.76 (0.19,3.13)
-3.31 (-6.50,-0.12)	-2.34 (-6.97,2.29)	-1.96 (-7.22,3.30)	-1.71 (-5.05,1.63)	-1.27 (-4.63,2.09)	**MultiPrb & Fos & Gos**	0.86 (0.04,19.41)	1.37 (0.27,7.03)	**6.60 (1.65,26.38)**	2.42 (0.47,12.35)
0.46 (-2.63,3.55)	1.43 (-3.14,6.00)	1.81 (-3.38,7.01)	2.07 (-1.18,5.31)	2.50 (-0.76,5.77)	3.77 (-0.67,8.21)	**MultiPrb & Fos**	1.59 (0.06,41.43)	7.66 (0.33,177.40)	2.81 (0.11,70.67)
-1.44 (-3.95,1.07)	-0.47 (-3.83,2.88)	-0.09 (-4.97,4.79)	0.16 (-2.54,2.86)	0.60 (-2.12,3.32)	1.87 (-2.18,5.92)	-1.90 (-5.88,2.07)	**Lactoferrin**	4.81 (0.89,25.94)	1.77 (0.27,11.69)
-1.14 (-3.57,1.28)	-0.18 (-4.32,3.97)	0.21 (-4.62,5.03)	0.46 (-2.16,3.07)	0.90 (-1.75,3.54)	2.17 (-1.84,6.17)	-1.61 (-5.54,2.32)	0.30 (-3.19,3.78)	**Fos & Gos**	0.37 (0.07,1.97)
-4.35 (-8.23,-0.48)	-3.38 (-8.52,1.76)	-3.00 (-7.54,1.54)	-2.75 (-6.57,1.08)	-2.31 (-6.33,1.71)	-1.04 (-6.07,3.99)	-4.81 (-9.78,0.15)	-2.91 (-7.54,1.72)	-3.21 (-7.78,1.36)	**Fos**

Footnote: Results are Odds Ratio (95% CIs) – top right half and mean change (95% CI) – bottom left half from the network meta-analysis. Odds ratios < 1 and mean difference values less than 0 indicate the treatment in bottom right is better. Numbers in bold represent statistically significant results. Colors represent the certainty in evidence for each pairwise comparison. Fos = Fructo-oligosaccharides, Gos = galacto-oligosaccharides; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.

High certainty evidence	Moderate certainty evidence	Low certainty evidence	Very low certainty evidence

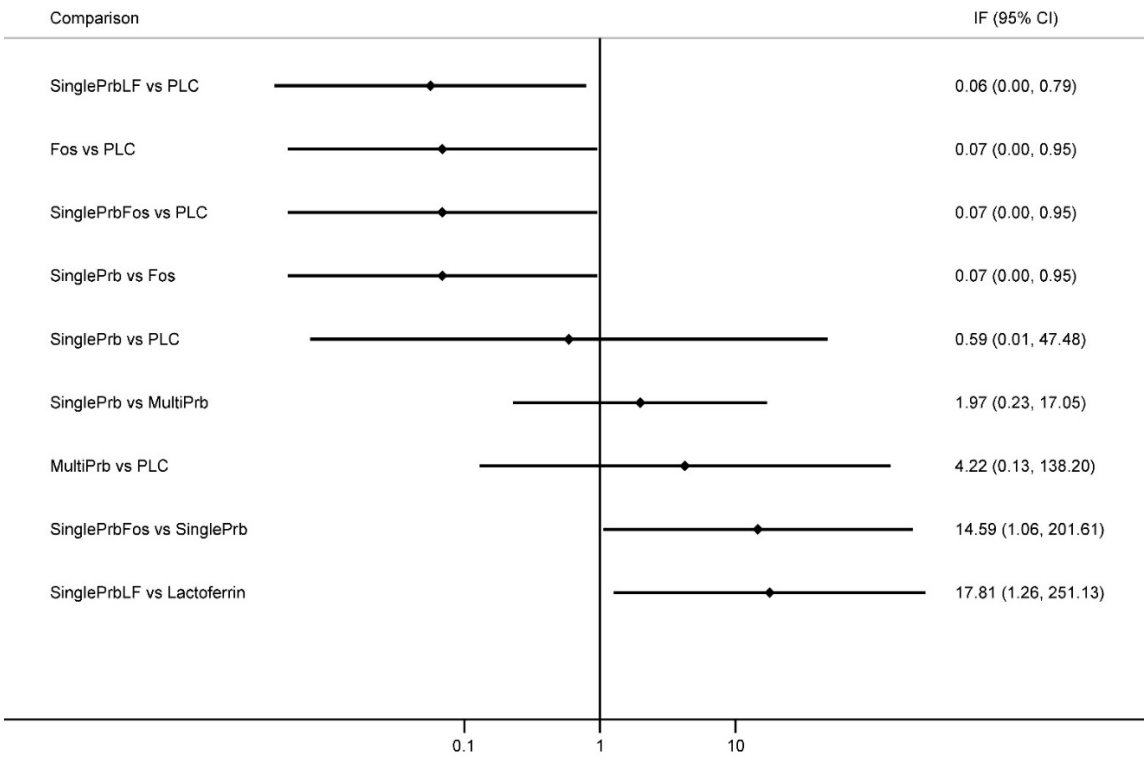
e-Table 4: NMA results for duration of hospitalization (days – top half) and weight change (grams change from birth until discharge/at 37 weeks' postnatal age – bottom half)

Placebo	1.98 (-5.43,9.39)	-2.19 (-10.08,5.69)	**-3.58 (-5.88,-1.27)**	**-2.68 (-5.15,-0.21)**	-9.00 (-22.11,4.11)	0.45 (-4.14,5.05)	1.75 (-4.02,7.53)	-2.86 (-10.30,4.58)	-6.19 (-14.01,1.62)
-	**SinglePrb & Lactoferrin**	-4.17 (-15.00,6.65)	-5.56 (-13.32,2.21)	-4.65 (-12.47,3.16)	-10.98 (-26.03,4.08)	-1.53 (-10.25,7.20)	-0.22 (-7.72,7.27)	-4.84 (-15.34,5.66)	-8.17 (-18.95,2.60)
36.3 (-170.8,243.4)	-	**SinglePrb & Fos**	-1.38 (-9.20,6.43)	-0.48 (-8.76,7.80)	-6.81 (-22.10,8.49)	2.65 (-6.49,11.78)	3.95 (-5.82,13.72)	-0.67 (-11.52,10.18)	-4.00 (-12.74,4.74)
51.9 (-36.1,139.9)	-	15.6 (-190.7,221.8)	**SinglePrb**	0.90 (-2.47,4.27)	-5.42 (-18.73,7.88)	4.03 (-1.11,9.16)	5.33 (-0.88,11.55)	0.71 (-7.07,8.50)	-2.62 (-10.36,5.13)
17.9 (-152.9,188.6)	-	-18.5 (-286.8,249.9)	-34.0 (-226.1,158.0)	**MultiPrb**	-6.32 (-19.66,7.01)	3.13 (-2.08,8.33)	4.43 (-1.85,10.71)	-0.19 (-8.02,7.65)	-3.52 (-11.73,4.69)
-	-	-	-	-	**MultiPrb & Fos & Gos**	9.45 (-4.44,23.34)	10.75 (-3.56,25.07)	6.14 (-8.93,21.21)	2.81 (-12.45,18.07)
-	-	-	-	-	-	**MultiPrb & Fos**	1.30 (-6.07,8.68)	-3.31 (-12.06,5.43)	-6.65 (-15.72,2.43)
33.8 (-182.3,249.9)	-	-2.5 (-301.8,296.7)	-18.1 (-251.4,215.2)	15.9 (-259.5,291.3)	-	-	**Lactoferrin**	-4.62 (-14.03,4.80)	-7.95 (-17.66,1.77)
59.3 (-119.3,238.0)	-	23.0 (-250.5,296.6)	7.5 (-191.6,206.6)	41.5 (-205.6,288.6)	-	-	25.6 (-254.8,306.0)	**Fos & Gos**	-3.33 (-14.13,7.46)
27.32 (-184.4,239.1)	-	-9.0 (-244.2,226.2)	-24.6 (-235.5,186.4)	9.46 (-262.5,281.5)	-	-	-6.5 (-309.0,296.1)	-32.0 (-309.1,245.1)	**Fos**

Footnote: Results are mean change (95% CI) from the network meta-analysis. Mean difference values less than 0 indicates the treatment in bottom right is better. Numbers in bold represent statistically significant results. Colors represent the certainty in evidence for each pairwise comparison. Fos = Fructo-oligosaccharides, Gos = galacto-oligosaccharides; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.

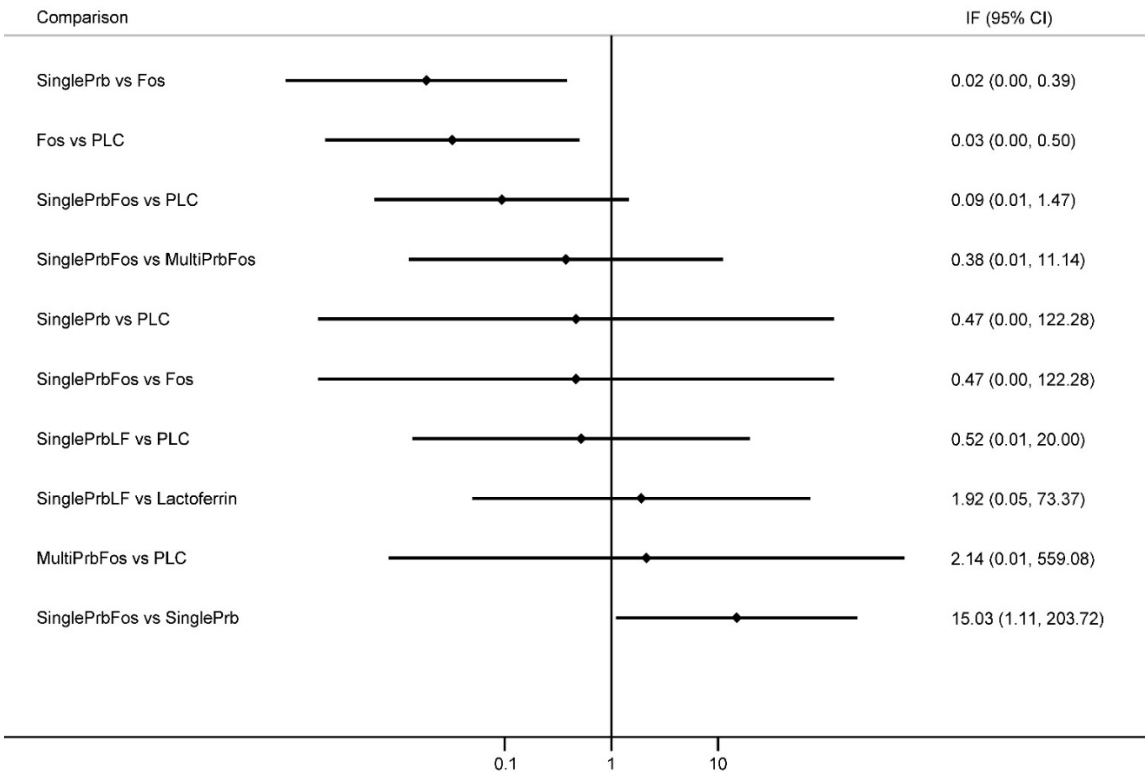
High certainty evidence	Moderate certainty evidence	Low certainty evidence	Very low certainty evidence

Appendix 6. 10. Results from evaluating incoherence in loops of evidence



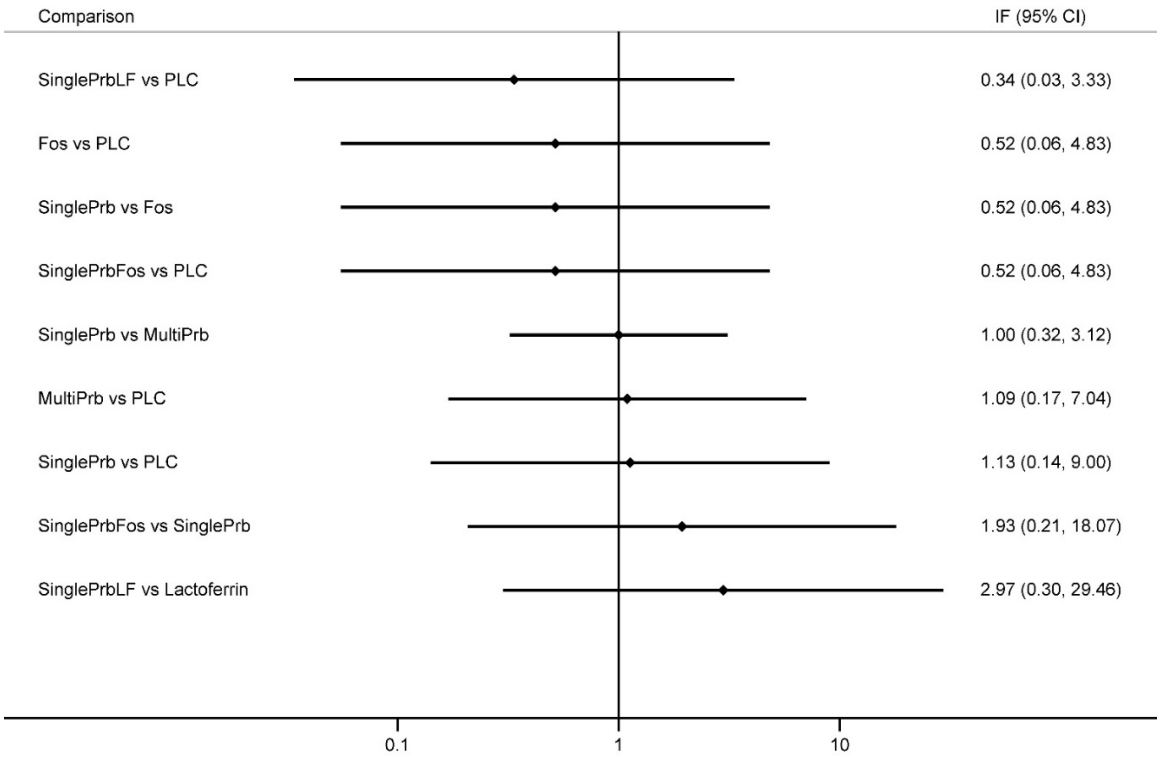
e-figure 6: Incoherence plot for all-cause mortality

Incoherence factors (IF) along with 95% confidence intervals (CI) are displayed. IFs are calculated as the ratio of the ORs from direct and indirect evidence in the loop. Comparisons that their lower CI limit does not reach the line of 1 are considered to present statistically significant incoherence. P value for the 'design-by-treatment' model (global test of incoherence) = 0.055. PLC = placebo; Fos = Fructo-oligosaccharides, Gos = galacto-oligosaccharides; LF = Lactoferrin; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.



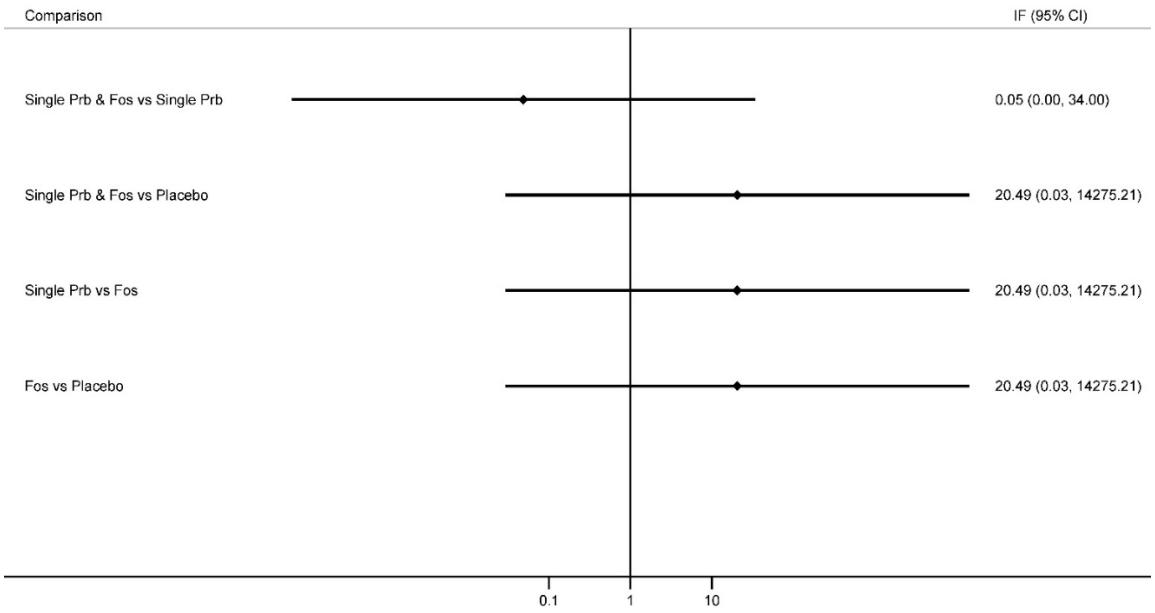
e-figure 7: Incoherence plot for severe NEC (Bell's stage II or more)

Incoherence factors (IF) along with 95% confidence intervals (CI) are displayed. IFs are calculated as the ratio of the ORs from direct and indirect evidence in the loop. Comparisons that their lower CI limit does not reach the line of 1 are considered to present statistically significant incoherence. P value for the 'design-by-treatment' model (global test of incoherence) = 0.057. PLC = placebo; Fos = Fructo-oligosaccharides, Gos = galacto-oligosaccharides; LF = Lactoferrin; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.



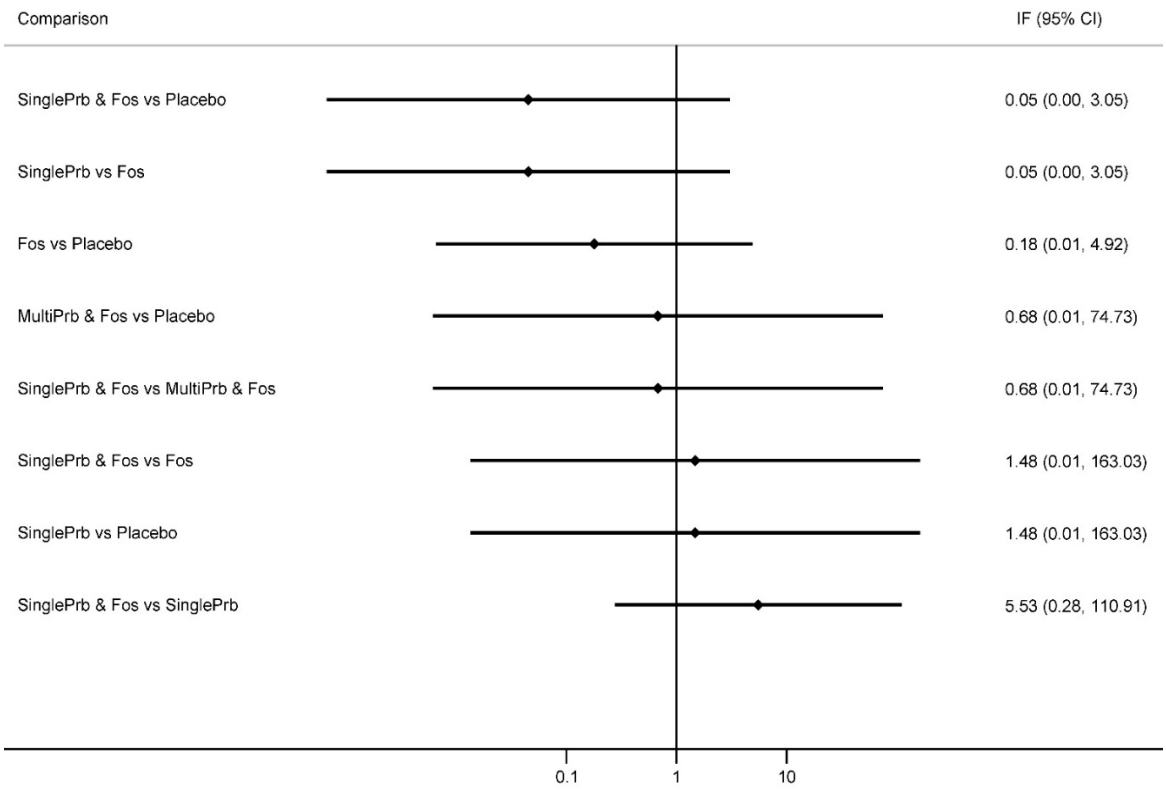
e-figure 8: Incoherence plot for culture proven sepsis

Incoherence factors (IF) along with 95% confidence intervals (CI) are displayed. IFs are calculated as the ratio of the ORs from direct and indirect evidence in the loop. Comparisons that their lower CI limit does not reach the line of 1 are considered to present statistically significant incoherence. P value for the 'design-by-treatment' model (global test of incoherence) = 0.881. PLC = placebo; Fos = Fructo-oligosaccharides, Gos = galacto-oligosaccharides; LF = Lactoferrin; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.



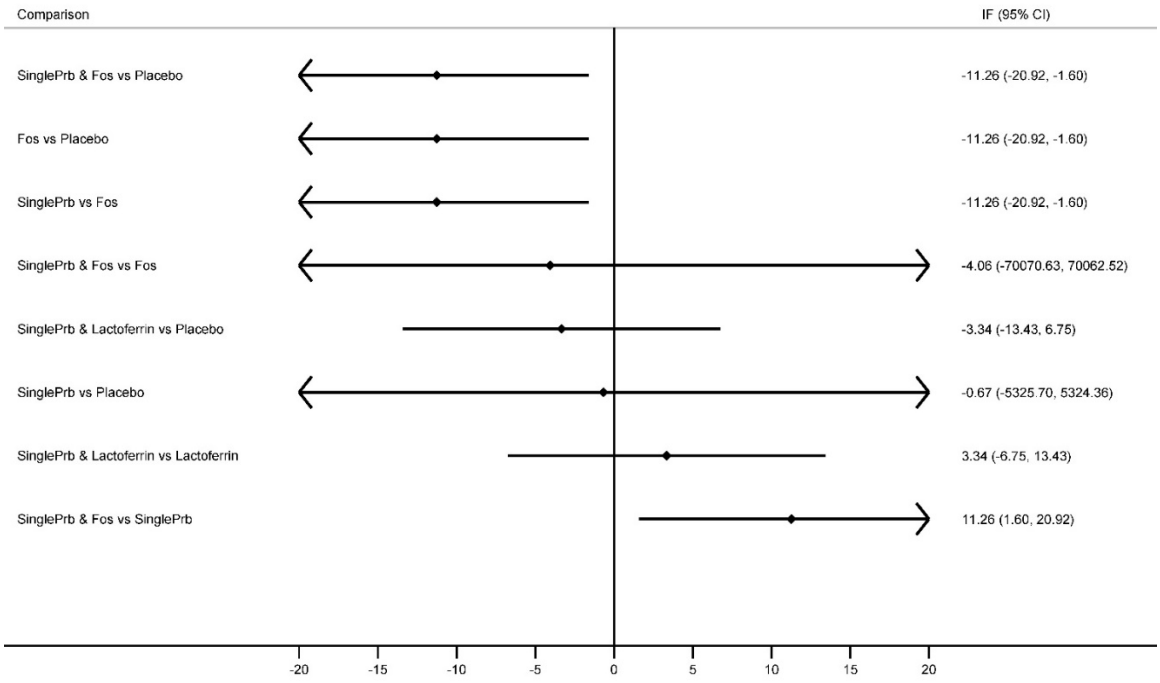
e-figure 9: Incoherence plot for NEC-related mortality

Incoherence factors (IF) along with 95% confidence intervals (CI) are displayed. IFs are calculated as the ratio of the ORs from direct and indirect evidence in the loop. Comparisons that their lower CI limit does not reach the line of 1 are considered to present statistically significant incoherence. P value for the 'design-by-treatment' model (global test of incoherence) = 0.366. Fos = Fructo-oligosaccharides, Multi Prb = Multiple-strain probiotics; Single Prb = Single-strain probiotics.



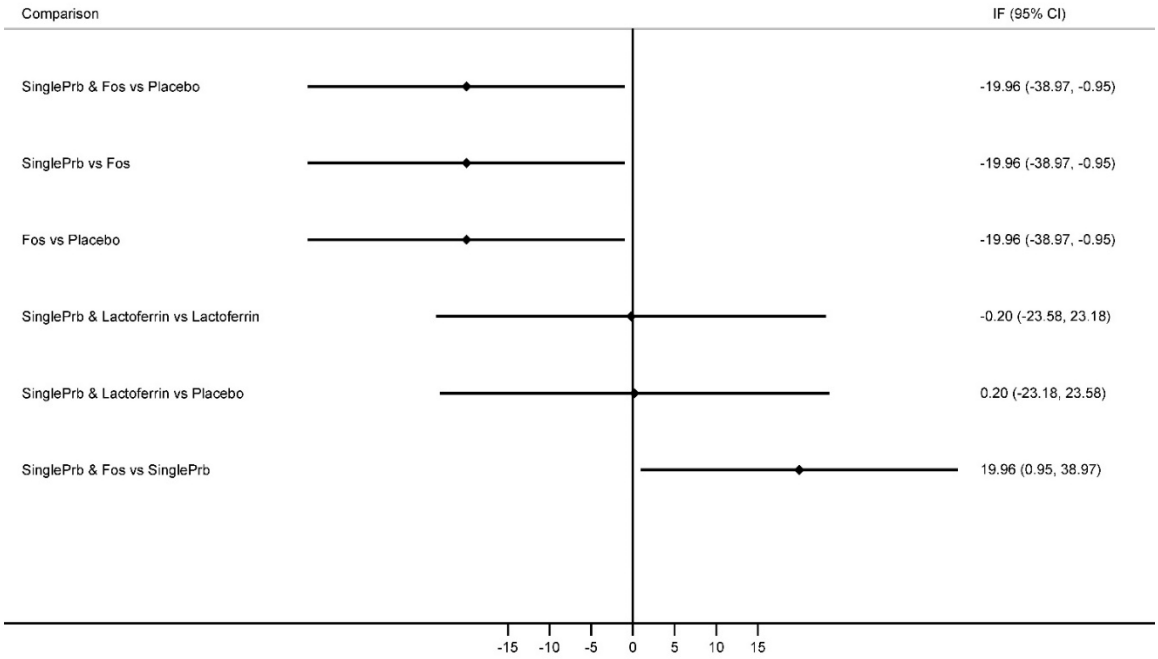
e-figure 10: Incoherence plot for feed intolerance

Incoherence factors (IF) along with 95% confidence intervals (CI) are displayed. IFs are calculated as the ratio of the ORs from direct and indirect evidence in the loop. Comparisons that their lower CI limit does not reach the line of 1 are considered to present statistically significant incoherence. P value for the 'design-by-treatment' model (global test of incoherence) = 0.354. Fos = Fructo-oligosaccharides; Gos = galacto-oligosaccharides; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.



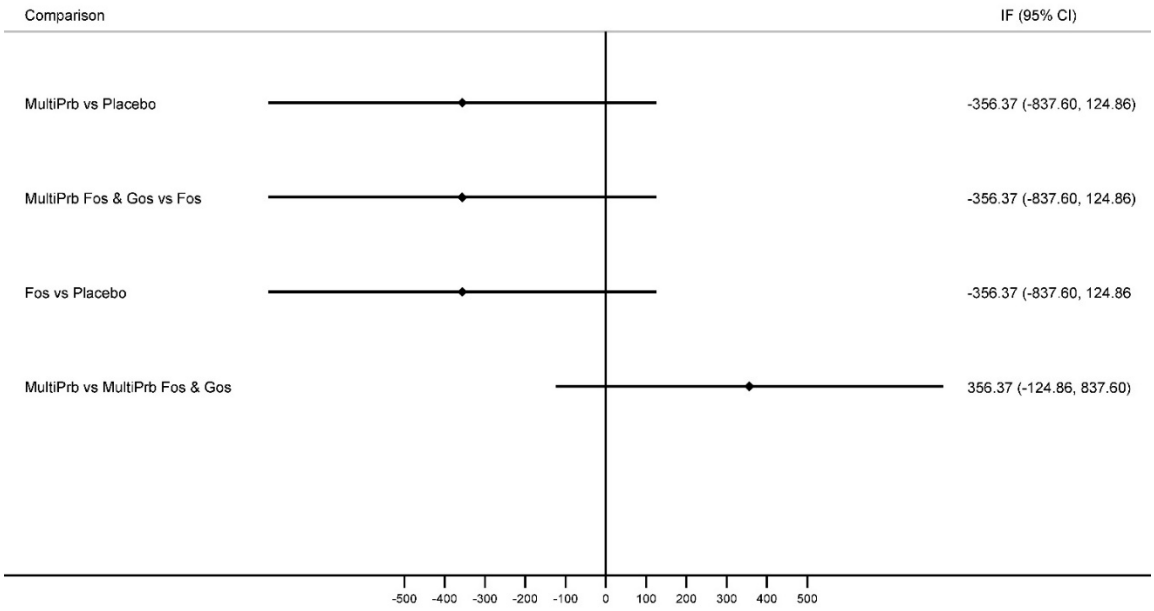
e-figure 11: Incoherence plot for time to reach full enteral feed (days)

Incoherence factors (IF) along with 95% confidence intervals (CI) are displayed. IFs are calculated as the absolute difference between direct and indirect estimates. Comparisons that their lower CI limit does not reach the zero line are considered to present statistically significant inconsistency. P value for the 'design-by-treatment' model (global test of incoherence) = 0.061. Fos = Fructo-oligosaccharides; Gos = galacto-oligosaccharides; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.



e-figure 12: Incoherence plot for duration of hospitalization (days)

Incoherence factors (IF) along with 95% confidence intervals (CI) are displayed. IFs are calculated as the absolute difference between direct and indirect estimates. Comparisons that their lower CI limit does not reach the zero line are considered to present statistically significant inconsistency. P value for the 'design-by-treatment' model (global test of incoherence) = 0.125. Fos = Fructo-oligosaccharides; Gos = galacto-oligosaccharides; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.



e-figure 13: Incoherence plot for weight change at discharge/37 weeks' postnatal age

Incoherence factors (IF) along with 95% confidence intervals (CI) are displayed. IFs are calculated as the absolute difference between direct and indirect estimates. Comparisons that their lower CI limit does not reach the zero line are considered to present statistically significant inconsistency. P value for the 'design-by-treatment' model (global test of incoherence) = 0.147. Fos = Fructo-oligosaccharides; Gos = galacto-oligosaccharides; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.

Appendix 6. 11. Results of the direct pairwise comparisons and GRADE assessments for all outcomes

e-Table 5: Results of direct pairwise comparisons with number trials and events for each trial arm and certainty of evidence (CoE) for all-cause mortality

Comparison	OR (95% CI)	# trials	# events C 1	n C 1	# events C 2	n C 2	I^2	P-bias*	GRADE CoE	Precision	Directness	Consistency	Overall RoB
SinglePrb & Lactoferrin vs PLC	0.52 (0.23, 1.19)	1	9	238	18	258	-	-	Moderate	Serious	Not serious	Not serious	Not serious
SinglePrb & Fos vs PLC	**0.23 (0.06, 0.83)**	1	3	100	12	100	-	-	Moderate	Serious	Not serious	Not serious	Not serious
SinglePrb vs PLC	0.83 (0.66, 1.04)	23	154	3030	187	2943	0.0	0.354	Moderate	Serious	Not serious	Not serious	Not serious
MultiPrb vs PLC	**0.68 (0.52, 0.87)**	30	117	4089	169	3993	0.0	0.946	Moderate	Serious	Not serious	Not serious	Not serious
MultiPrb & Fos & Gos vs PLC	**0.05 (0.01, 0.41)**	1	1	70	9	40	-	-	Very Low	Serious	Not serious	Not serious	Very serious
MultiPrb & Fos vs PLC	0.64 (0.10, 4.20)	2	10	208	12	210	45.1	-	Very Low	Not serious	Not serious	Serious	Serious
Lactoferrin vs PLC	0.69 (0.25, 1.88)	6	18	527	32	545	47.2	0.574	Low	Not serious	Not serious	Serious	Not serious
Fos & Gos vs PLC	0.95 (0.21, 4.42)	2	3	80	4	108	0.0	-	Very Low	Serious	Not serious	Not serious	Very serious
Fos vs PLC	**0.15 (0.03, 0.69)**	1	2	100	2	100	-	-	Moderate	Serious	Not serious	Not serious	Not serious
SinglePrb vs Fos	1.52 (0.25, 9.27)	1	3	100	2	100	-	-	Moderate	Serious	Not serious	Not serious	Not serious
SinglePrb & Lactoferrin vs Lactoferrin	1.90 (0.63, 5.76)	1	9	238	5	247	-	-	Moderate	Serious	Not serious	Not serious	Not serious
SinglePrb vs MultiPrb	1.96 (0.21, 18.0)	1	4	98	1	47	-	-	Moderate	Serious	Not serious	Not serious	Not serious
SinglePrb & Fos vs SinglePrb	1.00 (0.20, 5.08)	1	3	100	3	100	-	-	Moderate	Serious	Not serious	Not serious	Not serious

* P value of Harbord's test for small-study effects.

Results are Odds Ratio (95% CIs) from DerSimonian and Laird random-effects meta-analysis. Numbers in bold represent statistically significant results. "n" is number of infants randomized to each arm. C = comparison; RoB = overall risk of bias; PLC = Placebo; Fos = Fructo-oligosaccharides; Gos = galacto-oligosaccharides; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.

e-Table 6: Results of direct pairwise comparisons with number trials and events for each trial arm and certainty of evidence (CoE) for severe NEC

Comparison	OR (95% CI)	# trials	# events C 1	n C 1	# events C 2	n C 2	I^2	P-bias*	GRADE CoE	Precision	Directness	Consistency	Overall RoB
SinglePrb & Lactoferrin vs PLC	**0.04 (0.00, 0.60)**	1	0	238	14	259	-	-	Moderate	Serious	Not serious	Not serious	Not serious
SinglePrb & Fos vs PLC	**0.24 (0.08, 0.76)**	2	5	131	19	129	5.7	-	Low	Serious	Not serious	Not serious	Serious
SinglePrb vs PLC	**0.67 (0.53, 0.84)**	28	135	3116	205	3014	0.0	0.079	Moderate	Not serious	Not serious	Not serious	Serious
MultiPrb vs PLC	**0.35 (0.26, 0.47)**	28	73	3335	200	3226	0.0	0.170	Moderate	Not serious	Not serious	Not serious	Serious
MultiPrb & Fos & Gos vs PLC	**0.09 (0.02, 0.51)**	2	1	113	14	100	0.0	-	Very low	Serious	Not serious	Not serious	Very serious
MultiPrb & Fos vs PLC	**0.14 (0.04, 0.52)**	4	2	295	22	297	0.0	-	Low	Serious	Not serious	Not serious	Serious
Lactoferrin vs PLC	**0.42 (0.18, 0.97)**	4	8	368	22	382	0.0	-	Moderate	Serious	Not serious	Not serious	Not serious
Fos & Gos vs PLC	1.74 (0.69, 4.37)	4	12	238	8	271	0.0	-	Low	Serious	Not serious	Not serious	Serious
Fos vs PLC	0.62 (0.28, 1.37)	1	12	100	18	100	-	-	Moderate	Serious	Not serious	Not serious	Not serious
SinglePrb & Fos vs Fos	**0.31 (0.10, 0.98)**	1	4	100	12	100	-	-	Moderate	Serious	Not serious	Not serious	Not serious
SinglePrb vs Fos	**0.15 (0.03, 0.69)**	1	2	100	12	100	-	-	Moderate	Serious	Not serious	Not serious	Not serious
SinglePrb & Lactoferrin vs Lactoferrin	0.09 (0.01, 1.68)	1	0	238	5	247	-	-	Moderate	Serious	Not serious	Not serious	Not serious
SinglePrb & Fos vs MultiPrb & Fos	0.97 (0.06, 16.19)	1	1	31	1	30	-	-	Low	Serious	Not serious	Not serious	Serious
SinglePrb & Fos vs SinglePrb	2.04 (0.37, 11.41)	1	4	100	2	100	-	-	Moderate	Serious	Not serious	Not serious	Not serious

* P value of Harbord's test for small-study effects.

Results are Odds Ratio (95% CIs) from DerSimonian and Laird random-effects meta-analysis. Numbers in bold represent statistically significant results. "n" is number of infants randomized to each arm. C = comparison; RoB = overall risk of bias; PLC = Placebo; Fos = Fructo-oligosaccharides; Gos = galacto-oligosaccharides; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.

e-Table 7: Results of direct pairwise comparisons with number trials and events for each trial arm and certainty of evidence (CoE) for culture proven sepsis

Comparison	OR (95% CI)	# trials	# events C 1	n C 1	# events C 2	n C 2	I^2	P-bias*	GRADE CoE	Precision	Directness	Consistency	Overall RoB
SinglePrb & Lactoferrin vs PLC	**0.23 (0.10, 0.55)**	1	7	151	29	168	-	-	Moderate	Serious	Not serious	Not serious	Not serious
SinglePrb & Fos vs PLC	0.58 (0.23, 1.47)	1	8	100	13	100	-	-	Moderate	Serious	Not serious	Not serious	Not serious
SinglePrb vs PLC	**0.83 (0.69, 0.99)**	26	357	3192	397	3017	11.2	0.120	Low	Serious	Not serious	Not serious	Serious
MultiPrb vs PLC	**0.76 (0.59, 0.97)**	22	473	3081	561	3020	53.8	0.616	Very low	Serious	Not serious	Serious	Serious
MultiPrb & Fos & Gos vs PLC	0.79 (0.18, 3.44)	2	20	113	21	100	76.8	-	Low	Serious	Not serious	Serious	Not serious
MultiPrb & Fos vs PLC	0.59 (0.34, 1.02)	2	32	208	46	210	0.0	-	Very low	Serious	Not serious	Not serious	Very serious
Lactoferrin vs PLC	**0.43 (0.28, 0.67)**	6	36	432	78	454	0.0	0.241	High	Not serious	Not serious	Not serious	Not serious
Fos & Gos vs PLC	0.58 (0.33, 1.03)	3	31	180	53	208	0.0	-	Very low	Serious	Not serious	Not serious	Very serious
Fos vs PLC	0.74 (0.31, 1.78)	1	10	100	13	100	-	-	Moderate	Serious	Not serious	Not serious	Not serious
SinglePrb vs Fos	0.78 (0.30, 2.07)	1	8	100	10	100	-	-	Moderate	Serious	Not serious	Not serious	Not serious
SinglePrb & Fos vs Fos	0.78 (0.30, 2.07)	1	8	100	10	100	-	-	Moderate	Serious	Not serious	Not serious	Not serious
SinglePrb & Lactoferrin vs Lactoferrin	0.78 (0.28, 2.14)	1	7	151	9	153	-	-	Moderate	Serious	Not serious	Not serious	Not serious
SinglePrb vs MultiPrb	1.02 (0.41, 2.58)	1	17	98	8	47	-	-	Moderate	Serious	Not serious	Not serious	Not serious
SinglePrb & Fos vs SinglePrb	1.00 (0.36, 2.78)	1	8	100	8	100	-	-	Moderate	Serious	Not serious	Not serious	Not serious

* P value of Harbord's test for small-study effects.

Results are Odds Ratio (95% CIs) from DerSimonian and Laird random-effects meta-analysis. Numbers in bold represent statistically significant results. "n" is number of infants randomized to each arm. C = comparison; RoB = overall risk of bias; PLC = Placebo; Fos = Fructo-oligosaccharides; Gos = galacto-oligosaccharides; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.

e-Table 8: Results of direct pairwise comparisons with number trials and events for each trial arm and certainty of evidence (CoE) NEC-related mortality

Comparison	OR (95% CI)	# trials	# events C 1	n C 1	# events C 2	n C 2	I^2	P-bias*	GRADE CoE	Precision	Directness	Consistency	Overall RoB
SinglePrb & Fos vs PLC	5.10 (0.24, 107.62)	1	2	100	0	100	-	-	Low	Very serious	Not serious	Not serious	Not serious
SinglePrb vs PLC	0.71 (0.39, 1.27)	10	17	1843	26	1818	0.0	0.859	Moderate	Serious	Not serious	Not serious	Not serious
MultiPrb vs PLC	0.52 (0.26, 1.01)	16	8	2313	23	2250	0.0	0.535	Low	Serious	Not serious	Not serious	Serious
MultiPrb & Fos & Gos vs PLC	0.17 (0.02, 1.56)	2	0	113	4	100	0.0	-	Very low	Serious	Not serious	Not serious	Very serious
Lactoferrin vs PLC	1.13 (0.02, 59.49)	1	0	23	0	26	-	-	Low	Very serious	Not serious	Not serious	Not serious
Fos vs PLC	1.00 (0.02, 50.89)	1	0	100	0	100	-	-	Low	Very serious	Not serious	Not serious	Not serious
SinglePrb & Fos vs Fos	5.10 (0.24, 107.62)	1	2	100	0	100	-	-	Low	Very serious	Not serious	Not serious	Not serious
SinglePrb vs Fos	3.03 (0.12, 75.28)	1	1	100	0	100	-	-	Low	Very serious	Not serious	Not serious	Not serious
SinglePrb & Fos vs SinglePrb	1.68 (0.22, 12.99)	1	2	100	1	100	-	-	Low	Very serious	Not serious	Not serious	Not serious

* P value of Harbord's test for small-study effects.

Results are Odds Ratio (95% CIs) from DerSimonian and Laird random-effects meta-analysis. Numbers in bold represent statistically significant results. "n" is number of infants randomized to each arm. C = comparison; RoB = overall risk of bias; PLC = Placebo; Fos = Fructo-oligosaccharides; Gos = galacto-oligosaccharides; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.

e-Table 9: Results of direct pairwise comparisons with number trials and events for each trial arm and certainty of evidence (CoE) feed intolerance

Comparison	OR (95% CI)	# trials	# events C 1	n C 1	# events C 2	n C 2	I^2	P-bias*	GRADE CoE	Precision	Directness	Consistency	Overall RoB
SinglePrb & Fos vs PLC	0.42 (0.15, 1.18)	2	5	129	12	130	0.0	-	Low	Serious	Not serious	Not serious	Serious
SinglePrb vs PLC	**0.47 (0.36, 0.61)**	9	202	1272	307	1178	23.7	0.579	Moderate	Not serious	Not serious	Serious	Not serious
MultiPrb vs PLC	**0.48 (0.30, 0.77)**	3	53	242	93	271	15.0	-	Moderate	Not serious	Not serious	Not serious	Serious
MultiPrb & Fos & Gos vs PLC	**0.15 (0.06, 0.37)**	1	24	70	31	40	-	-	Very low	Serious	Not serious	Not serious	Very serious
MultiPrb & Fos vs PLC	0.12 (0.01, 2.52)	1	0	31	3	30	-	-	Low	Serious	Not serious	Not serious	Serious
Lactoferrin vs PLC	**0.21 (0.06, 0.77)**	1	3	63	13	67	-	-	Low	Serious	Not serious	Not serious	Serious
Fos & Gos vs PLC	1.00 (0.38, 2.66)	1	10	25	20	50	-	-	Low	Serious	Not serious	Not serious	Serious
Fos vs PLC	0.31 (0.08, 1.19)	1	3	100	9	100	-	-	Moderate	Serious	Not serious	Not serious	Not serious
SinglePrb & Fos vs Fos	1.35 (0.29, 6.18)	1	4	100	3	100	-	-	Moderate	Serious	Not serious	Not serious	Not serious
SinglePrb vs Fos	0.33 (0.03, 3.19)	1	1	100	3	100	-	-	Moderate	Serious	Not serious	Not serious	Not serious
SinglePrb & Fos vs MultiPrb & Fos	3.32 (0.13, 84.7)	1	1	29	0	31	-	-	Very low	Very serious	Not serious	Not serious	Serious
SinglePrb & Fos vs SinglePrb	4.13 (0.45, 37.6)	1	4	100	1	100	-	-	Low	Very serious	Not serious	Not serious	Not serious

* P value of Harbord's test for small-study effects.

Results are Odds Ratio (95% CIs) from DerSimonian and Laird random-effects meta-analysis. Numbers in bold represent statistically significant results. "n" is number of infants randomized to each arm. C = comparison; RoB = overall risk of bias; PLC = Placebo; Fos = Fructo-oligosaccharides; Gos = galacto-oligosaccharides; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.

e-Table 10: Results of direct pairwise comparisons with number trials and events for each trial arm and certainty of evidence (CoE) for time to reach full enteral feed (days)

Comparison	MD (95% CI)	# trials	n comparison 1	n comparison 2	I^2	P-bias*	GRADE CoE	Precision	Directness	Consistency	Overall RoB
SinglePrb & Lactoferrin vs PLC	**-1.40 (-2.27, -0.53)**	1	238	258	-	-	High	Not serious	Not serious	Not serious	Not serious
SinglePrb & Fos vs PLC	**-5.00 (-8.95, -1.05)**	1	100	100	-	-	Low	Serious	Not serious	Not serious	Serious
SinglePrb vs PLC	**-1.47 (-2.16, -0.77)**	19	2283	2078	60.8	0.137	Moderate	Not serious	Not serious	Serious	Not serious
MultiPrb vs PLC	**-2.00 (-3.35, -0.66)**	17	1708	1707	94.0	0.287	Very low	Not serious	Not serious	Serious	Very serious
MultiPrb & Fos & Gos vs PLC	-3.05 (-10.89, 4.79)	2	113	100	93.3	-	Very low	Serious	Not serious	Serious	Serious
MultiPrb & Fos vs PLC	0.27 (-0.54, 1.08)	2	157	158	0.0	-	Very low	Serious	Not serious	Not serious	Very serious
Lactoferrin vs PLC	**-1.99 (-3.26, -0.73)**	4	362	378	9.4	-	High	Not serious	Not serious	Not serious	Not serious
Fos & Gos vs PLC	**-1.03 (-1.70, -0.37)**	3	153	189	0.0	-	Moderate	Serious	Not serious	Not serious	Not serious
Fos vs PLC	**-8.00 (-11.64, -4.36)**	1	100	100	-	-	Low	Serious	Not serious	Not serious	Serious
SinglePrb & Fos vs Fos	**3.00 (0.10, 5.90)**	1	100	100	-	-	Low	Serious	Not serious	Not serious	Serious
SinglePrb vs Fos	1.00 (-1.18, 3.18)	1	100	100	-	-	Low	Serious	Not serious	Not serious	Serious
SinglePrb & Lactoferrin vs Lactoferrin	**0.90 (0.07, 1.73)**	1	238	247	-	-	Moderate	Serious	Not serious	Not serious	Not serious
SinglePrb & Fos vs SinglePrb	2.00 (-0.67, 4.67)	1	100	100	-	-	Low	Serious	Not serious	Not serious	Serious

* P value of Eggers's test for small-study effects.

Results are mean difference (95% CIs) from DerSimonian and Laird random-effects meta-analysis. Numbers in bold represent statistically significant results. "n" is number of infants randomized to each arm. RoB = overall risk of bias; PLC = Placebo; Fos = Fructo-oligosaccharides; Gos = galacto-oligosaccharides; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.

e-Table 11: Results of direct pairwise comparisons with number trials and events for each trial arm and certainty of evidence (CoE) for duration of hospitalization

Comparison	MD (95% CI)	# trials	n comparison 1	n comparison 2	I^2	P-bias*	GRADE CoE	Precision	Directness	Consistency	Overall RoB
SinglePrb & Lactoferrin vs PLC	2.00 (-1.88, 5.88)	1	238	258	-	-	Moderate	Serious	Not serious	Not serious	Not serious
SinglePrb & Fos vs PLC	**-8.00 (-14.89, -1.11)**	1	100	100	-	-	Moderate	Serious	Not serious	Not serious	Not serious
SinglePrb vs PLC	**-3.58 (-5.87, -1.29)**	18	2282	2197	87.2	0.648	Moderate	Not serious	Not serious	Serious	Not serious
MultiPrb vs PLC	**-3.00 (-4.84, -1.16)**	13	1527	1533	76.3	0.112	Low	Not serious	Not serious	Serious	Serious
MultiPrb & Fos & Gos vs PLC	-9.00 (-20.18, 2.18)	1	70	40	-	-	Very low	Serious	Not serious	Not serious	Very serious
MultiPrb & Fos vs PLC	0.02 (-0.92, 0.96)	3	265	268	0.0	-	Low	Not serious	Not serious	Not serious	Very serious
Lactoferrin vs PLC	1.80 (-1.67, 5.28)	3	332	343	0.0	-	Moderate	Not serious	Not serious	Not serious	Serious
Fos & Gos vs PLC	-2.99 (-8.23, 2.24)	2	80	108	0.0	-	Low	Serious	Not serious	Not serious	Serious
Fos vs PLC	**-12.00 (-18.81, -5.19)**	1	100	100	-	-	Moderate	Serious	Not serious	Not serious	Not serious
Fos vs SinglePrb & Fos	-4.00 (-9.44, 1.44)	1	100	100	-	-	Moderate	Serious	Not serious	Not serious	Not serious
SinglePrb vs Fos	-1.00 (-6.04, 4.04)	1	100	100	-	-	Moderate	Serious	Not serious	Not serious	Not serious
SinglePrb & Lactoferrin vs Lactoferrin	0.20 (-3.96, 4.36)	1	238	247	-	-	Moderate	Serious	Not serious	Not serious	Not serious
SinglePrb & Fos vs SinglePrb	5.00 (-0.15, 10.15)	1	100	100	-	-	Moderate	Serious	Not serious	Not serious	Not serious

* P value of Eggers's test for small-study effects.

Results are Odds Ratio (95% CIs) from DerSimonian and Laird random-effects meta-analysis. Numbers in bold represent statistically significant results. "n" is number of infants randomized to each arm. RoB = overall risk of bias; PLC = Placebo; Fos = Fructo-oligosaccharides; Gos = galacto-oligosaccharides; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.

e-Table 12: Results of direct pairwise comparisons with number trials and events for each trial arm and certainty of evidence (CoE) weight change

Comparison	MD (95% CI)	# trials	n comparison 1	n comparison 2	I^2	P-bias*	GRADE CoE	Precision	Directness	Consistency	Overall RoB
SinglePrb & Fos vs PLC	-44.00 (-141.65, 53.65)	1	100	100	-	-	Moderate	Serious	Not serious	Not serious	Not serious
SinglePrb vs PLC	51.72 (-32.56, 136.00)	8	635	616	85.1	0.948	Moderate	Serious	Not serious	Not serious	Not serious
MultiPrb vs PLC	14.40 (-58.01, 86.81)	2	598	601	0.0	-	Moderate	Serious	Not serious	Not serious	Not serious
Lactoferrin vs PLC	**33.77 (24.32, 43.23)**	1	58	55	-	-	Moderate	Not serious	Not serious	Not serious	Serious
Fos & Gos vs PLC	59.78 (-125.16, 244.71)	2	110	132	74.5	-	Very low	Serious	Not serious	Not serious	Very serious
Fos vs PLC	-53.00 (-160.20, 54.20)	1	100	100	-	-	Moderate	Serious	Not serious	Not serious	Not serious
SinglePrb & Fos vs Fos	9.00 (-84.45, 102.45)	1	100	100	-	-	Moderate	Serious	Not serious	Not serious	Not serious
SinglePrb vs Fos	-49.00 (-143.93, 45.93)	1	100	100	-	-	Moderate	Serious	Not serious	Not serious	Not serious
SinglePrb & Fos vs SinglePrb	58.00 (-26.00, 142.00)	1	100	100	-	-	Moderate	Serious	Not serious	Not serious	Not serious

* P value of Eggers's test for small-study effects.

Results are Odds Ratio (95% CIs) from DerSimonian and Laird random-effects meta-analysis. Numbers in bold represent statistically significant results. "n" is number of infants randomized to each arm. RoB = overall risk of bias; PLC = Placebo; Fos = Fructo-oligosaccharides; Gos = galacto-oligosaccharides; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.

Appendix 6. 12. SUCRA and cumulative probability plots

e-Table 13: The surface under the cumulative ranking (SUCRA) values and mean ranks for primary outcomes

Interventions	All-cause mortality		NEC (Bell's stage ≥ II)		Culture proven sepsis	
	SUCRA value	*Mean rank*	*SUCRA value*	*Mean rank*	*SUCRA value*	*Mean rank*
MultiPrb Fos & Gos	96.8	1.3	84.4	2.4	43.6	6.1
MultiPrb & Fos	25.2	7.7	79.9	2.8	54.9	5.1
MultiPrb	48.9	5.6	57.0	4.9	39.6	6.4
SinglePrb & lactoferrin	48.4	5.6	91.6	1.8	93.1	1.6
SinglePrb & Fos	74.3	3.3	64.9	4.2	49.7	5.5
SinglePrb	31.0	7.2	33.2	7.0	35.1	6.8
Lactoferrin	49.4	5.6	48.7	5.6	79.1	2.9
Fos & Gos	29.5	7.3	3.2	9.7	61.2	4.5
Fos	82.6	2.6	23.4	7.9	32.9	7.0
Placebo	13.9	8.7	13.6	8.8	10.9	9.0

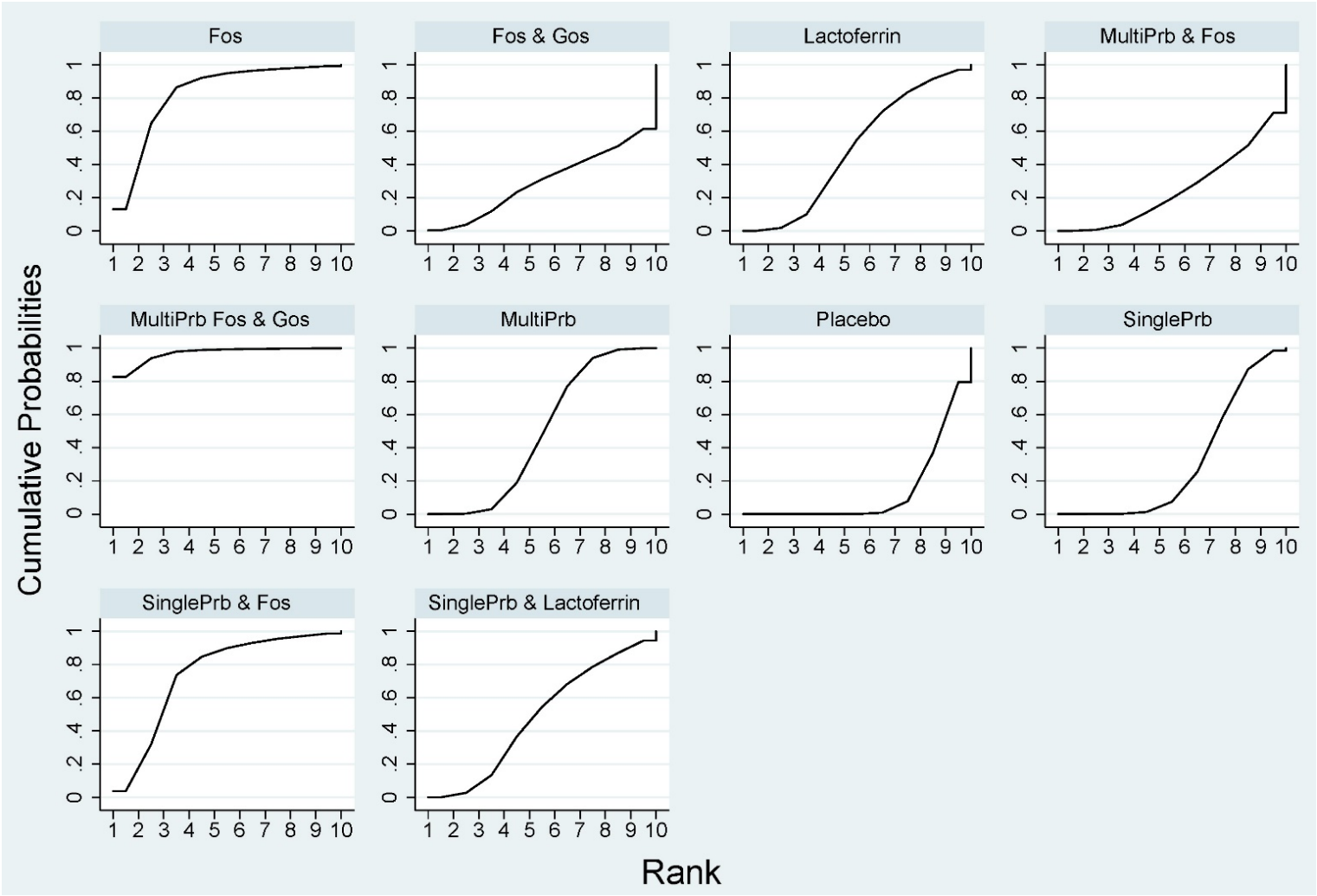
Fos = Fructo-oligosaccharides, Gos = galacto-oligosaccharides; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.

e-Table 14: The surface under the cumulative ranking (SUCRA) values and mean ranks for secondary outcomes

Interventions	NEC-related mortality		Feed intolerance		Time to reach full feed		Duration of hospitalization	
	SUCRA value	*Mean rank*	*SUCRA value*	*Mean rank*	*SUCRA value*	*Mean rank*	*SUCRA value*	*Mean rank*
MultiPrb Fos & Gos	83.3	2.0	86.1	2.1	79.8	2.8	83.5	2.5
MultiPrb & Fos	-	-	76.3	2.9	16.7	8.5	28.4	7.4
MultiPrb	63.9	3.2	44.3	5.5	64.0	4.2	59.7	4.6
SinglePrb & lactoferrin	-	-	-	-	40.7	6.3	20.9	8.1
SinglePrb & Fos	20.4	5.8	44.2	5.5	47.0	5.8	51.1	5.4
SinglePrb	49.6	4.0	45.9	5.3	52.9	5.2	68.8	3.8
Lactoferrin	38.7	4.7	75.0	3.0	49.8	5.5	20.2	8.2
Fos & Gos	-	-	13.9	7.9	43.3	6.1	57.5	4.8
Fos	65.6	3.1	54.8	4.6	89.0	2.0	79.9	2.8
Placebo	28.5	5.3	9.4	8.2	16.8	8.5	30.0	7.3

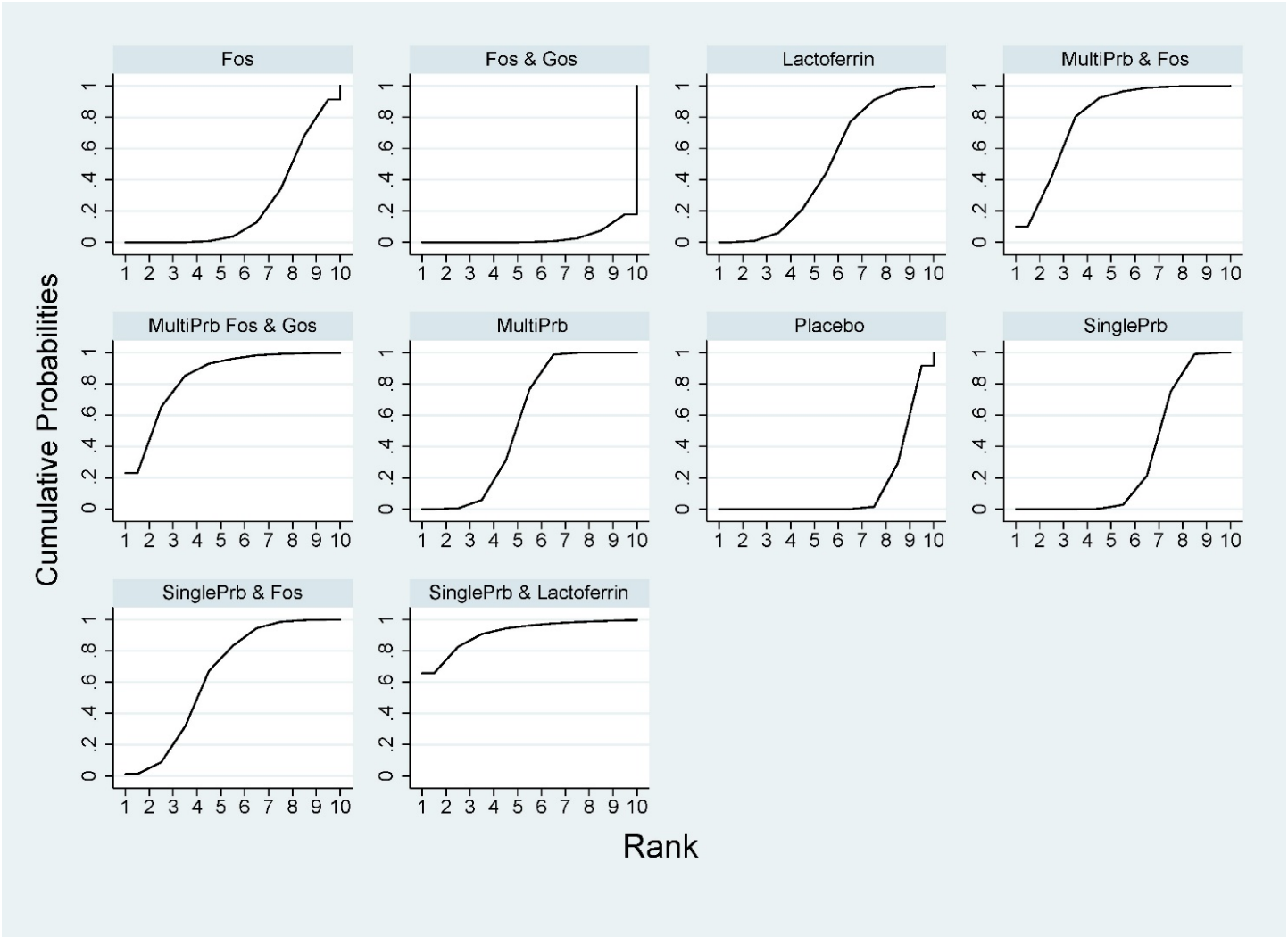
Fos = Fructo-oligosaccharides, Gos = galacto-oligosaccharides; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.

e-Figure 14: Plots of the surface under the cumulative ranking curves for all treatments for the 'all-cause mortality'



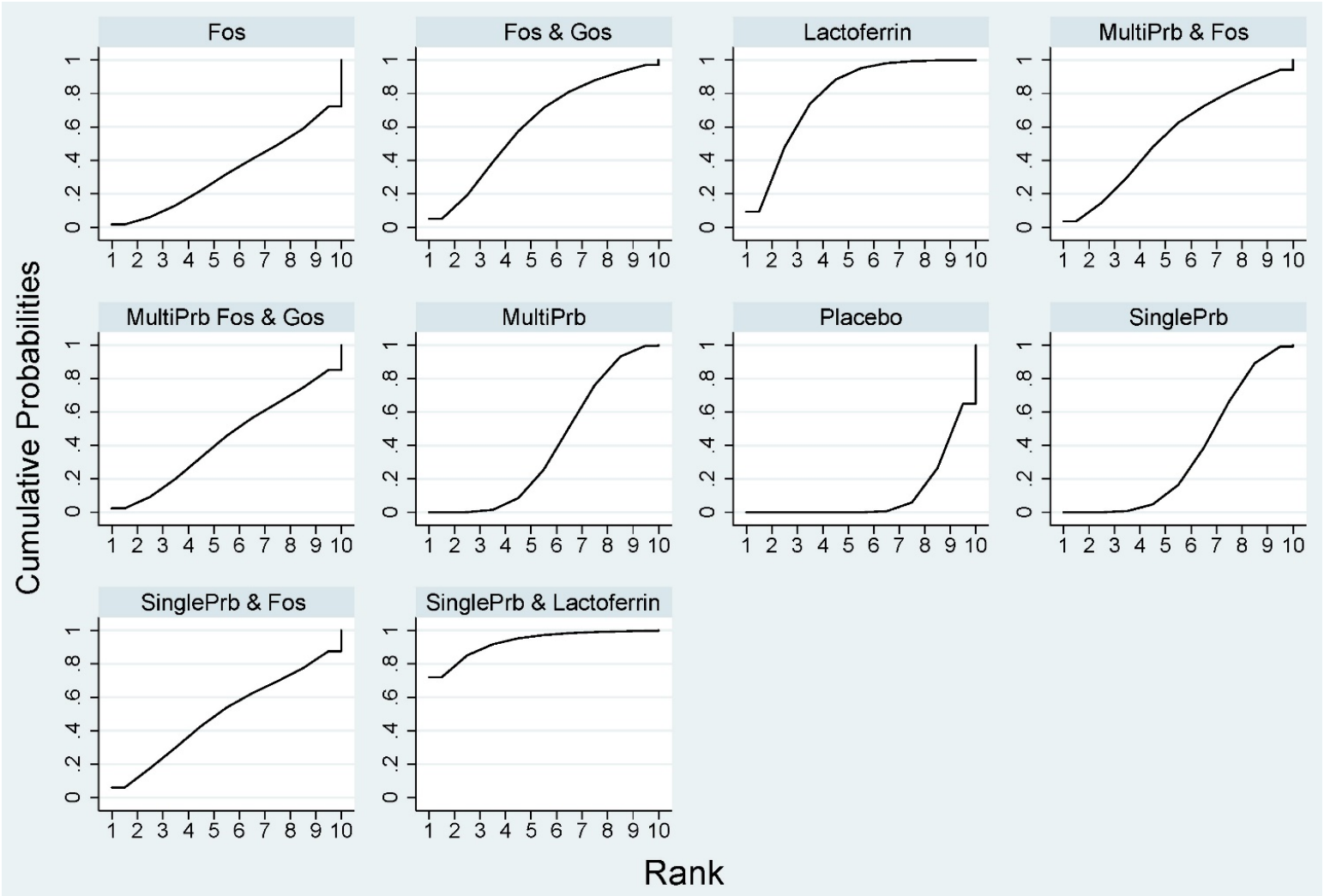
Fos = Fructo-oligosaccharides; Gos = galacto-oligosaccharides; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.

e-Figure 15: Plots of the surface under the cumulative ranking curves for all treatments for the 'NEC stage II or more'



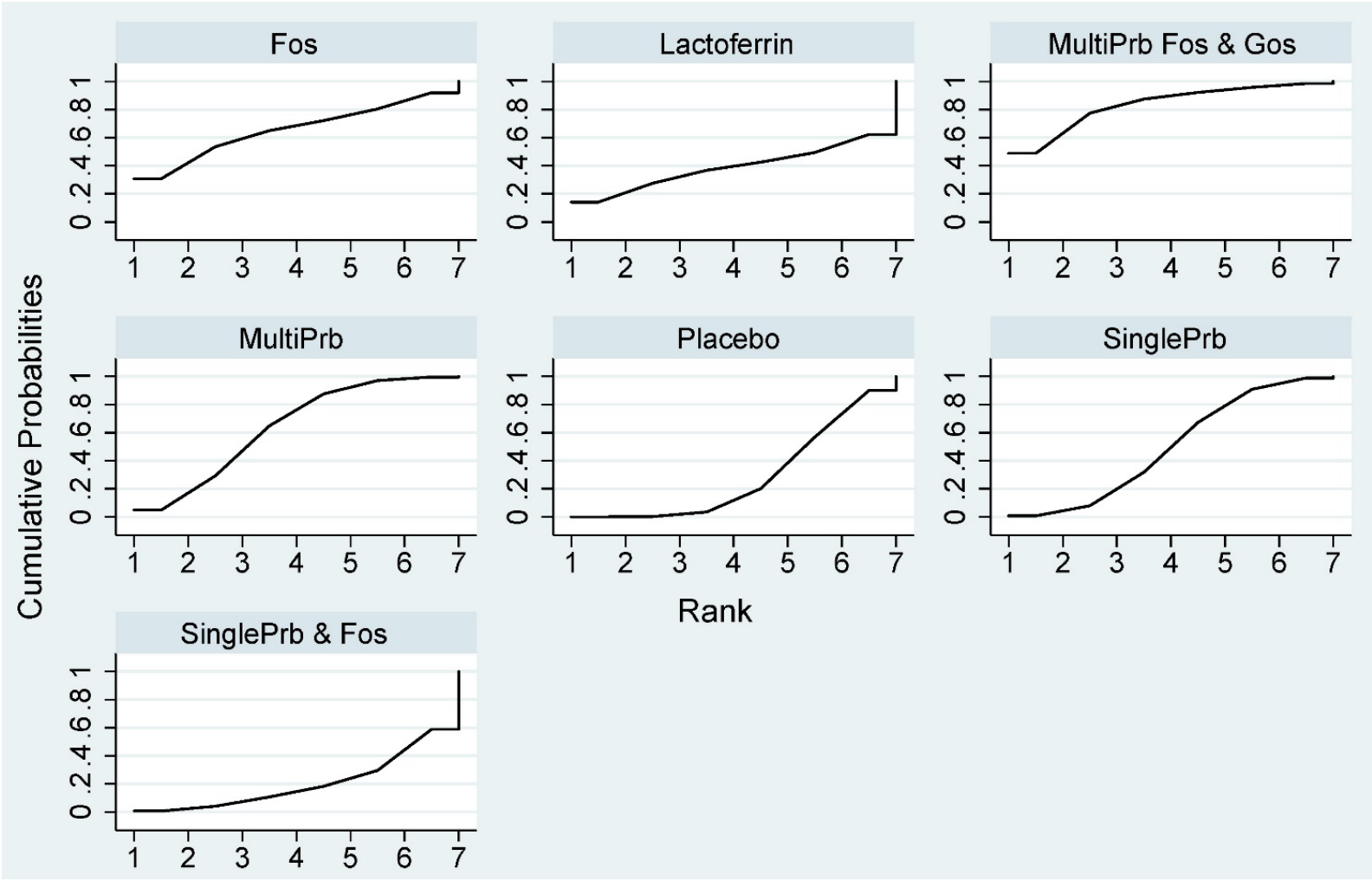
Fos = Fructo-oligosaccharides; Gos = galacto-oligosaccharides; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.

e-Figure 16: Plots of the surface under the cumulative ranking curves for all treatments for the 'culture proven sepsis'



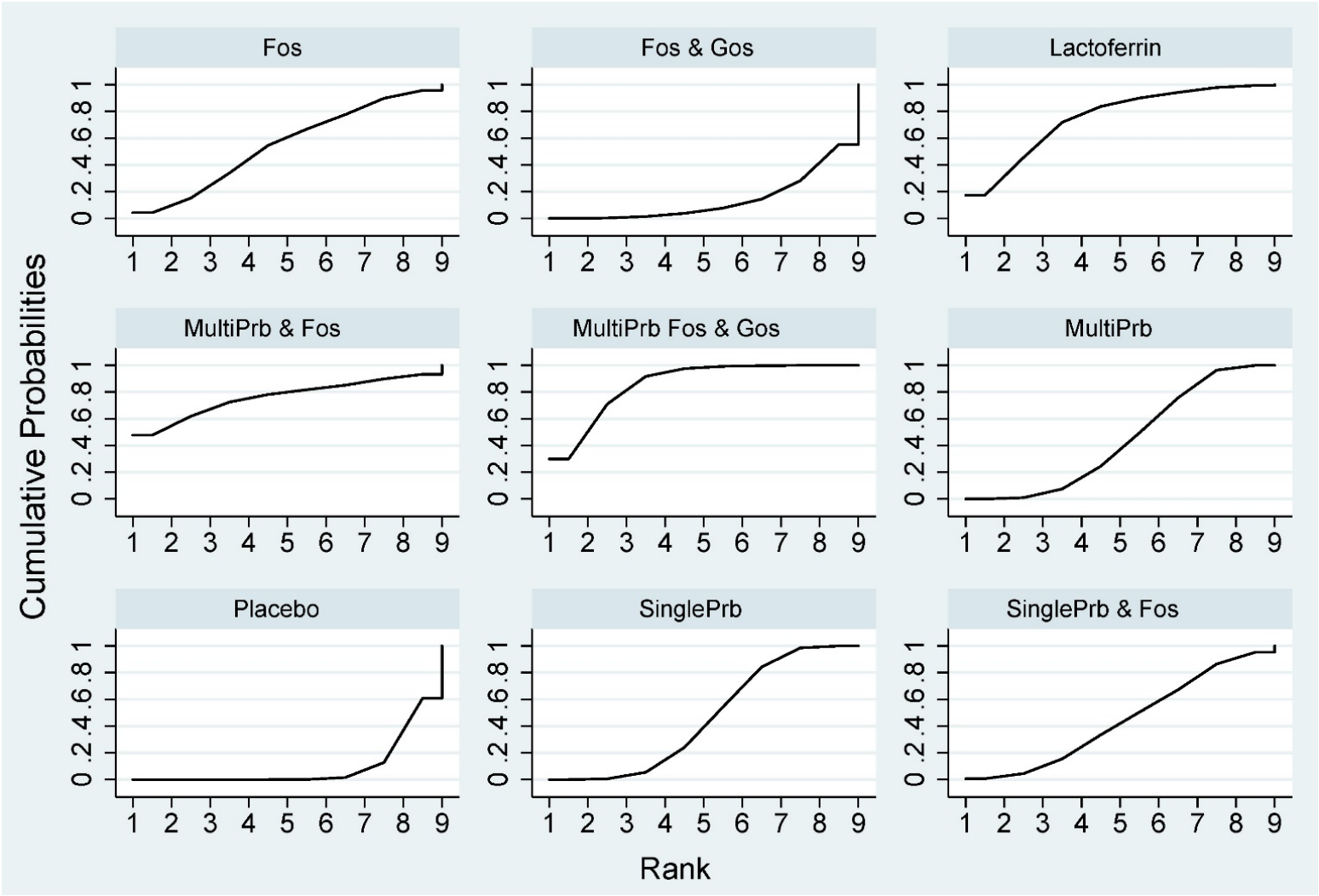
Fos = Fructo-oligosaccharides; Gos = galacto-oligosaccharides; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.

e-Figure 17: Plots of the surface under the cumulative ranking curves for all treatments for the 'NEC-related mortality'



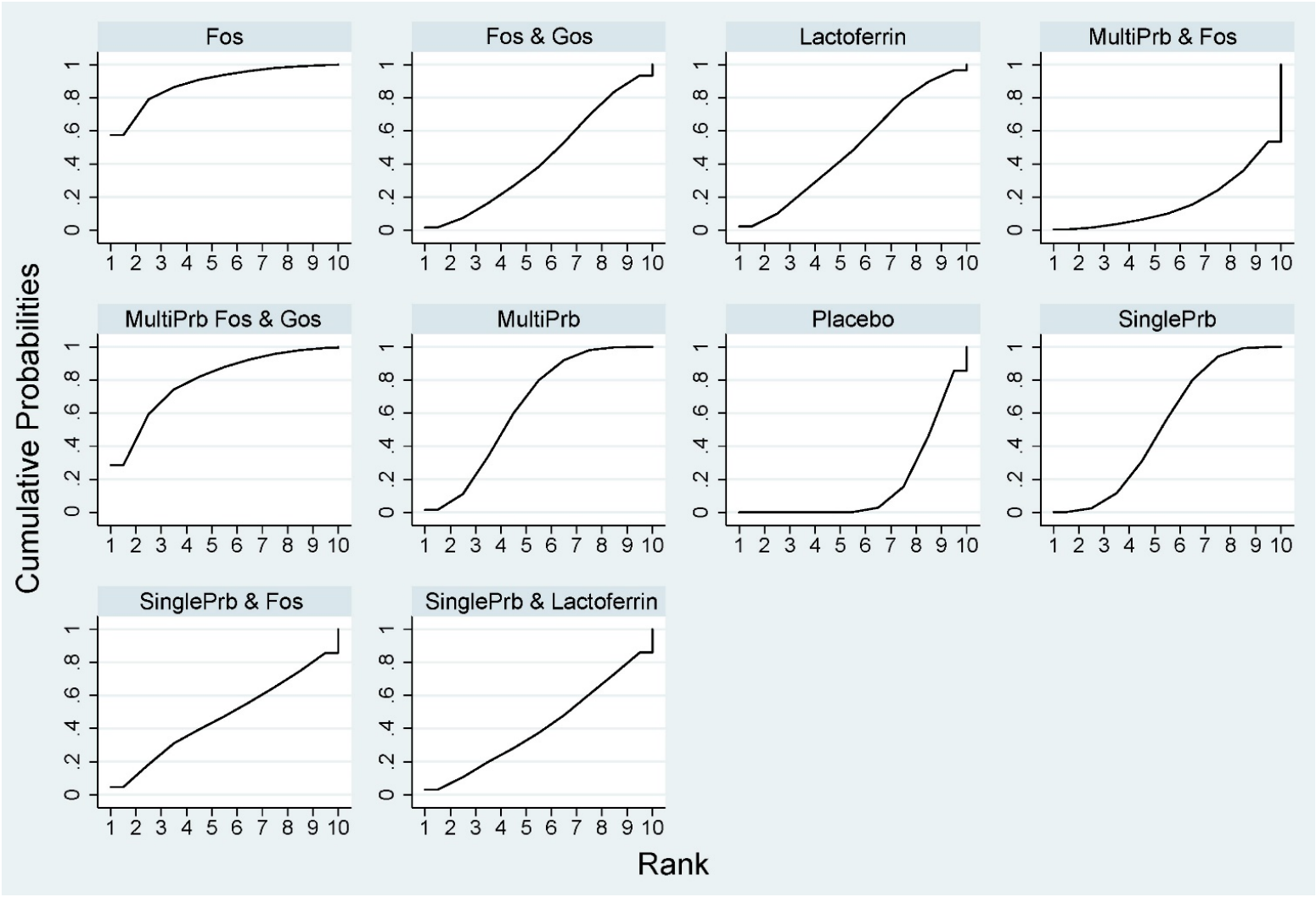
Fos = Fructo-oligosaccharides; Gos = galacto-oligosaccharides; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.

e-Figure 18: Plots of the surface under the cumulative ranking curves for all treatments for the 'feed intolerance'



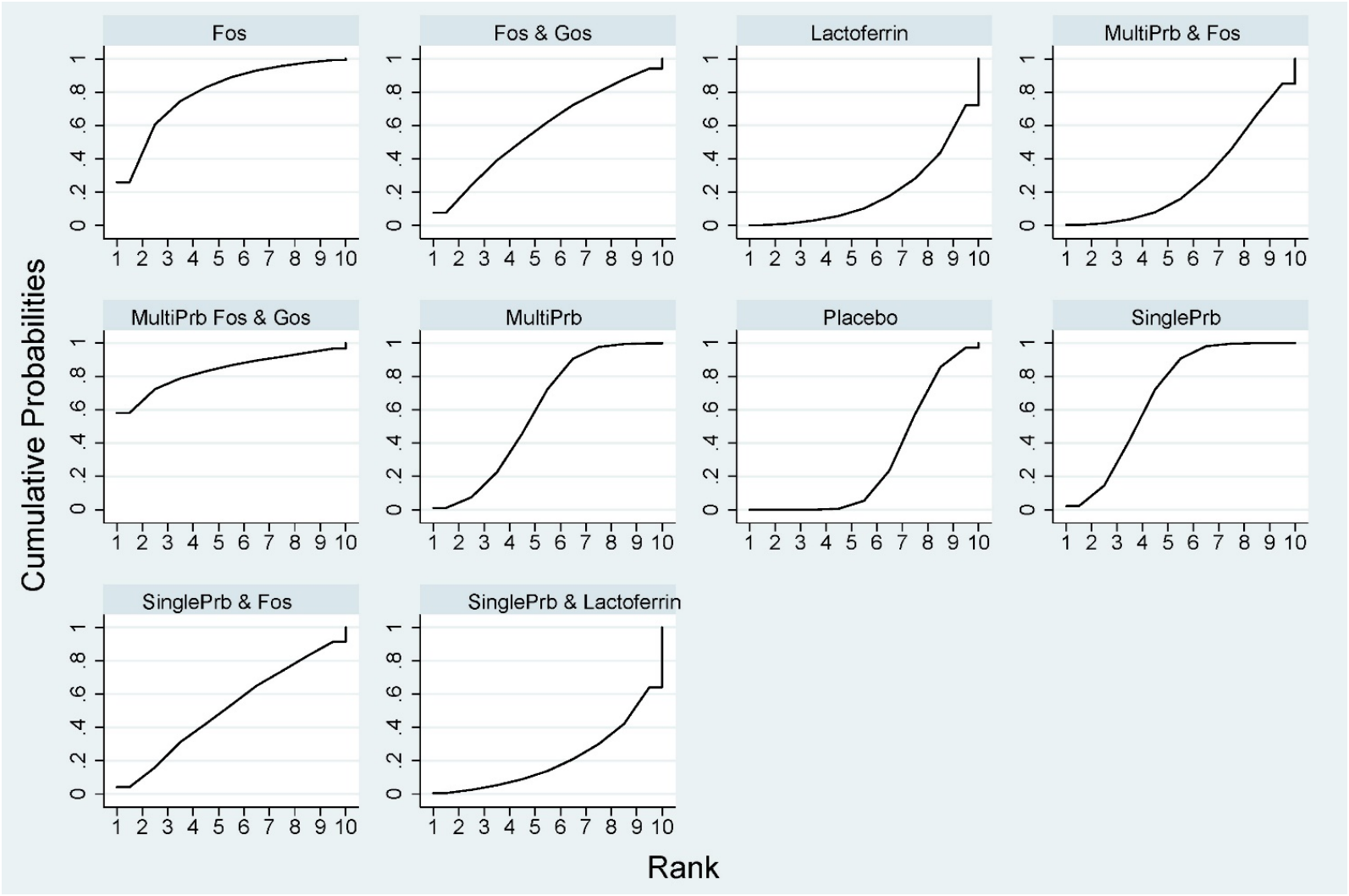
Fos = Fructo-oligosaccharides; Gos = galacto-oligosaccharides; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.

e-Figure 19: Plots of the surface under the cumulative ranking curves for all treatments for the 'time to reach full enteral feed'



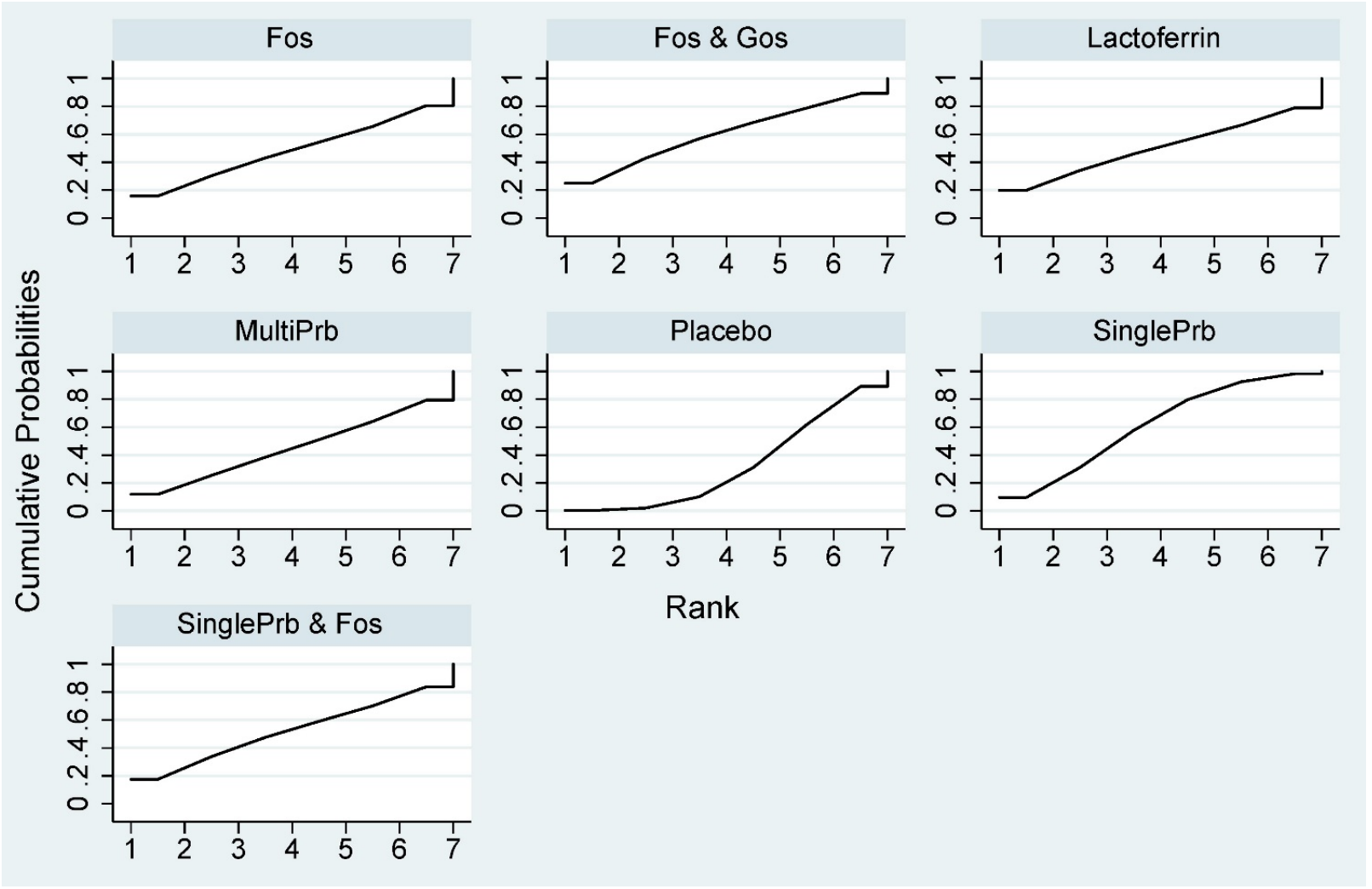
Fos = Fructo-oligosaccharides; Gos = galacto-oligosaccharides; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.

e-Figure 20: Plots of the surface under the cumulative ranking curves for all treatments for the 'duration of hospitalization'



Fos = Fructo-oligosaccharides; Gos = galacto-oligosaccharides; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.

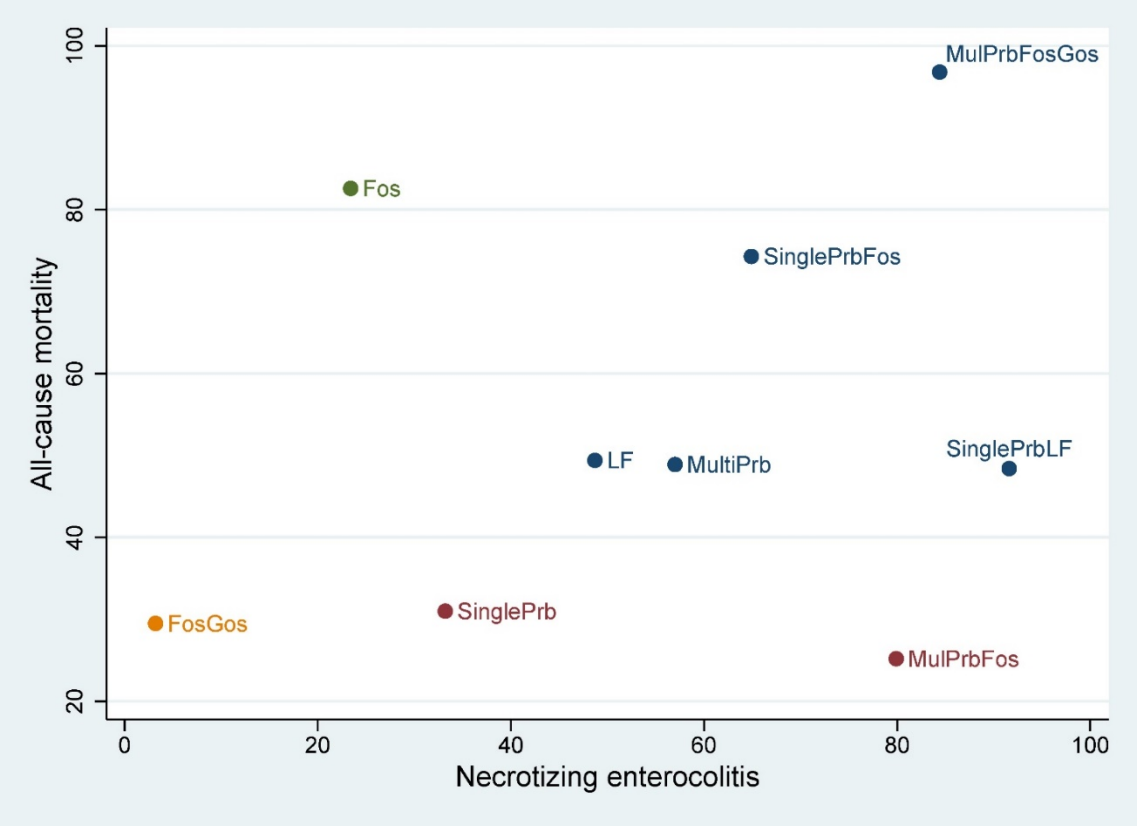
e-Figure 21: Plots of the surface under the cumulative ranking curves for all treatments for the 'weight change'



Fos = Fructo-oligosaccharides; Gos = galacto-oligosaccharides; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.

Appendix 6. 13. Results of hierarchical cluster analysis for primary outcomes

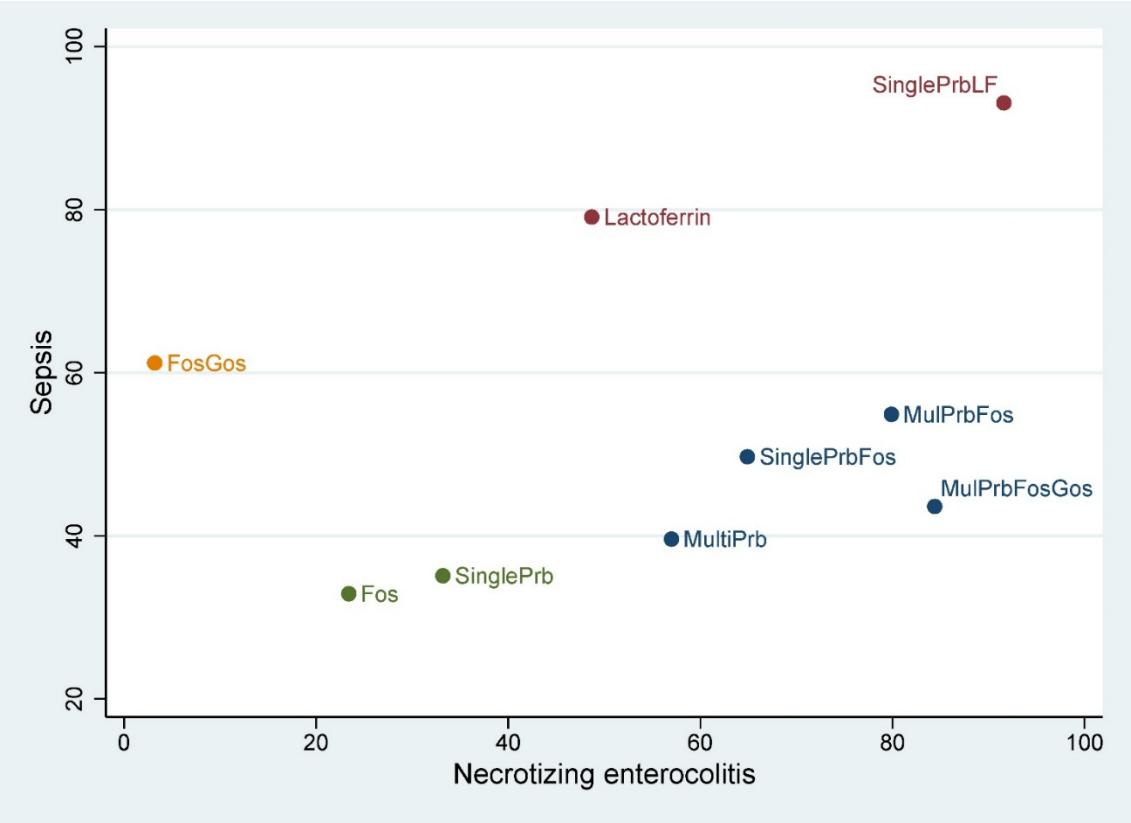
e-Figure 22: Clustered ranking based on SUCRA values for all-cause mortality and severe NEC (stage II or more)



Hierarchical cluster analysis is performed to group the competing treatments. Different colors represent different treatment groups considering joint relative ranking for each two outcomes (treatments with the same color can be considered as effective with respect to both outcomes). Treatments lying on the upper right-hand side are more effective for both outcomes.

Fos = Fructo-oligosaccharides, Gos = galacto-oligosaccharides; LF = Lactoferrin; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.

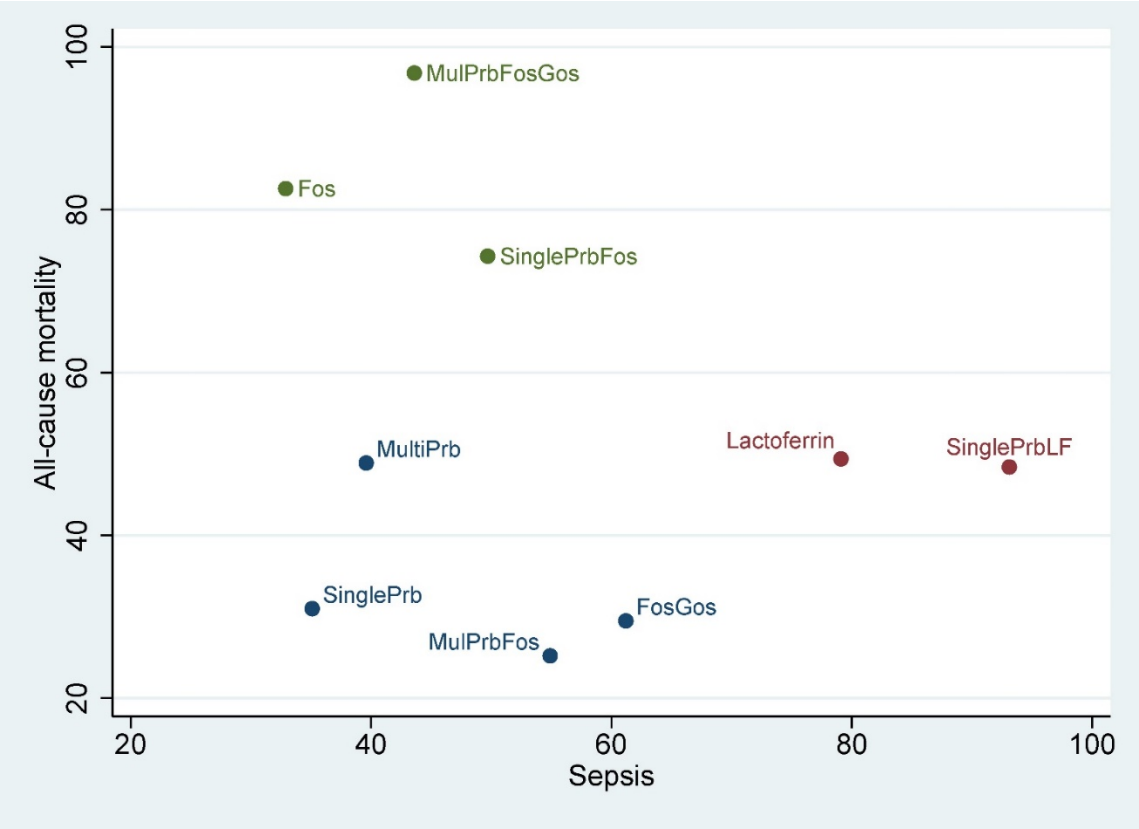
e-Figure 23: Clustered ranking based on SUCRA values for severe NEC (stage II or more) and Culture proven sepsis



Hierarchical cluster analysis is performed to group the competing treatments. Different colors represent different treatment groups considering joint relative ranking for each two outcomes (treatments with the same color can be considered as effective with respect to both outcomes). Treatments lying on the upper right-hand side are more effective for both outcomes.

Fos = Fructo-oligosaccharides, Gos = galacto-oligosaccharides; LF = Lactoferrin; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.

e-Figure 24: Clustered ranking based on SUCRA values for all-cause mortality and Culture proven sepsis

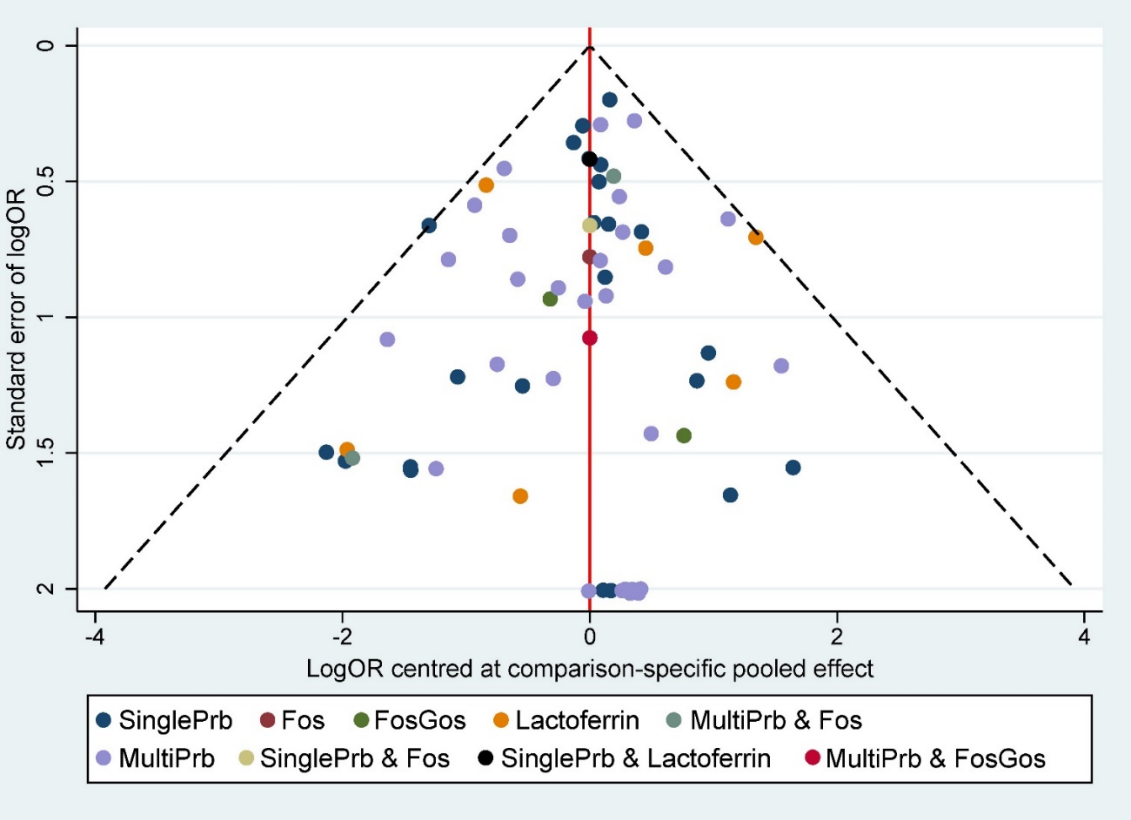


Hierarchical cluster analysis is performed to group the competing treatments. Different colors represent different treatment groups considering joint relative ranking for each two outcomes (treatments with the same color can be considered as effective with respect to both outcomes). Treatments lying on the upper right-hand side are more effective for both outcomes.

Fos = Fructo-oligosaccharides, Gos = galacto-oligosaccharides; LF = Lactoferrin; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.

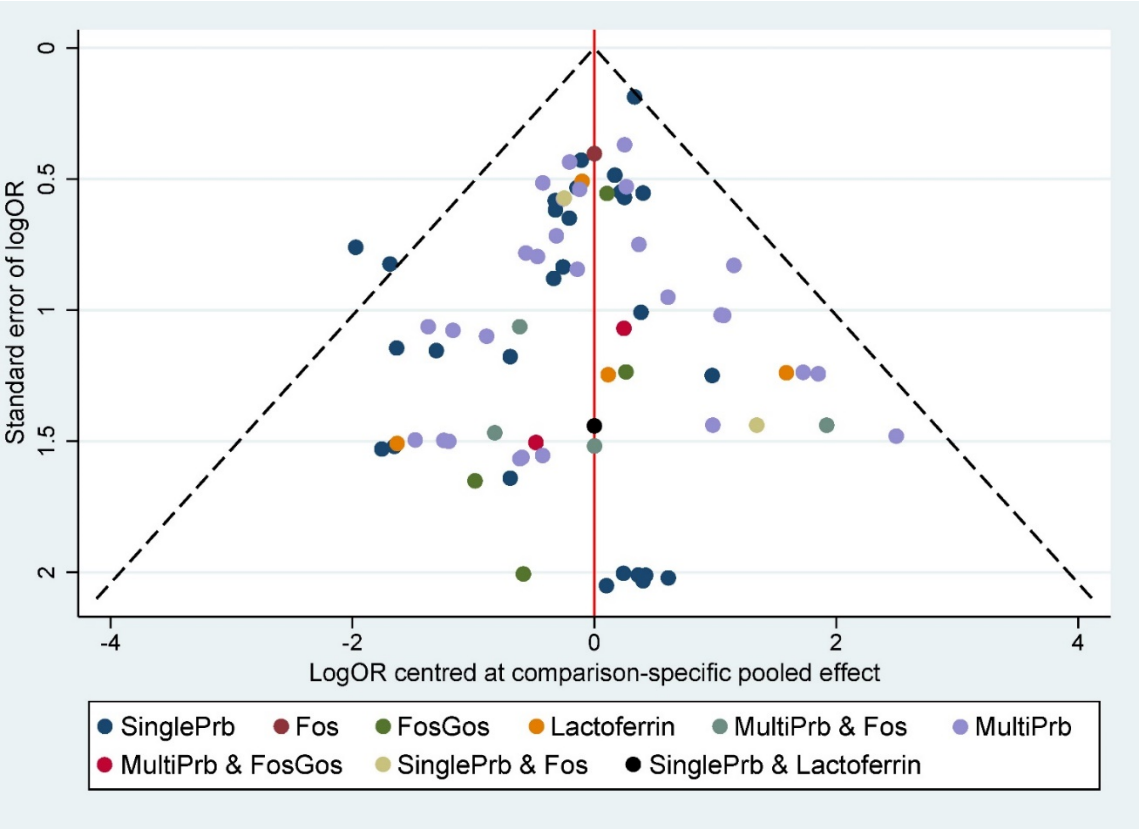
Appendix 6. 14. Comparison-adjusted funnel plots for the network of all outcomes

e-Figure 25: Funnel plot with pseudo 95% confidence limits including all the comparisons of the active treatment vs. placebo for all-cause mortality.



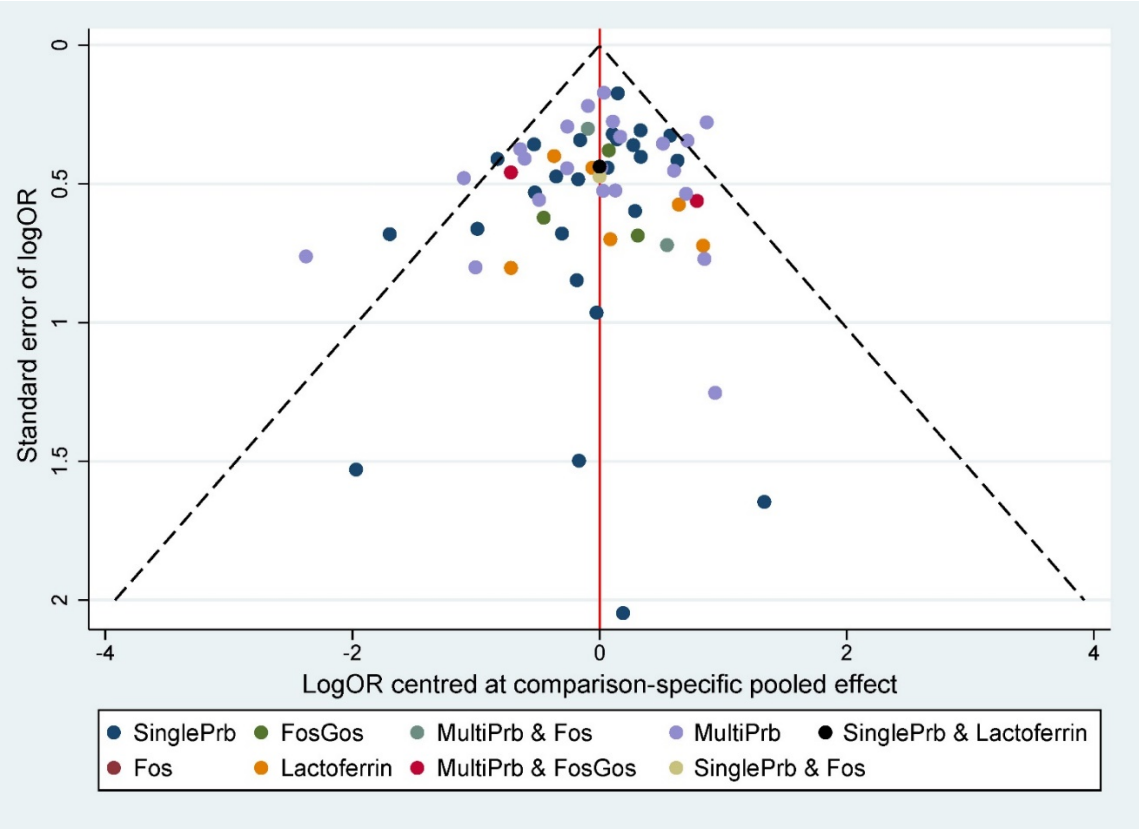
Fos = Fructo-oligosaccharides; Gos = galacto-oligosaccharides; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.

e-Figure 26: Funnel plot with pseudo 95% confidence limits including all the comparisons of the active treatment vs. placebo for severe NEC (stage II or more).



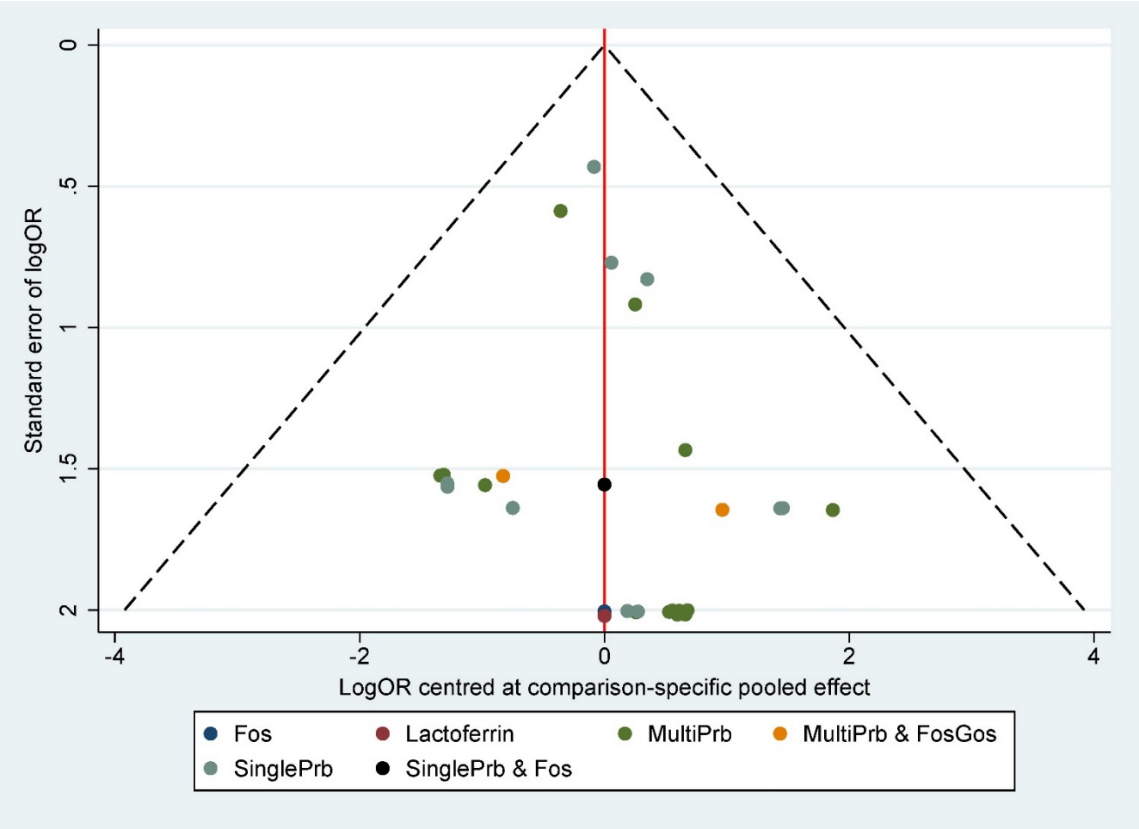
Fos = Fructo-oligosaccharides; Gos = galacto-oligosaccharides; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.

e-Figure 27: Funnel plot with pseudo 95% confidence limits including all the comparisons of the active treatment vs. placebo for culture proven sepsis.



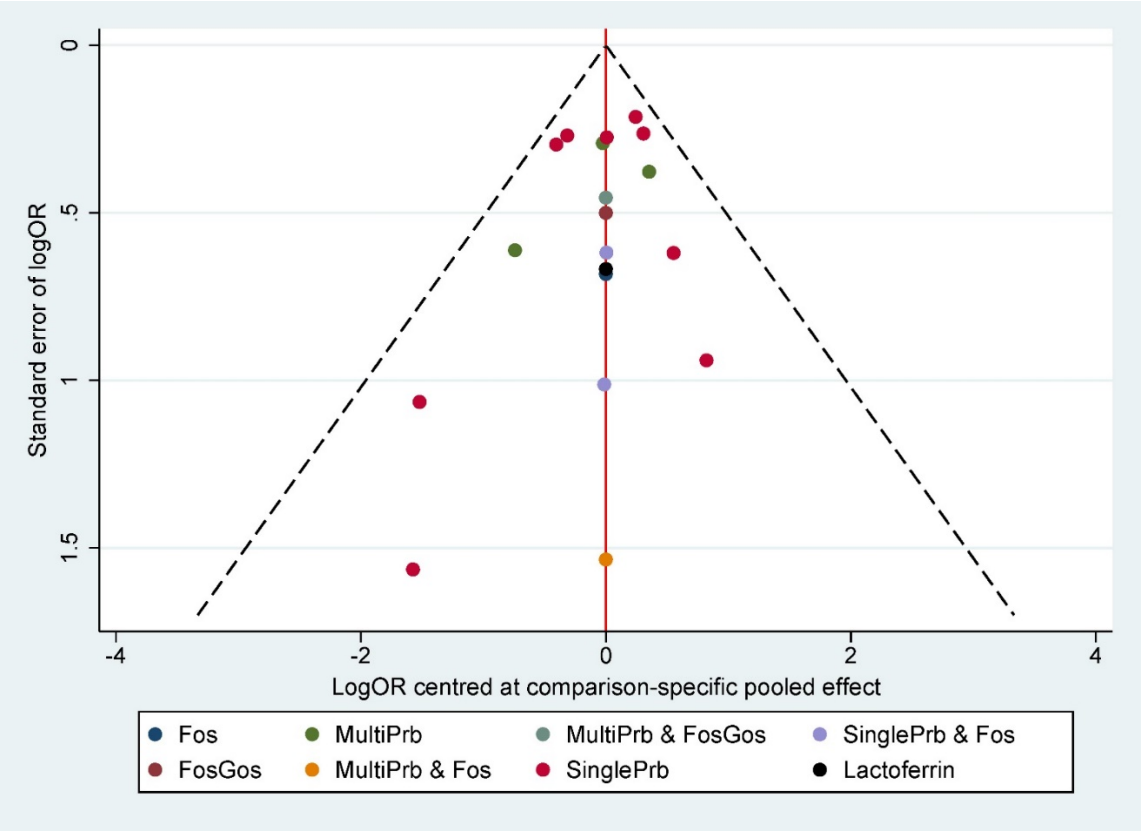
Fos = Fructo-oligosaccharides; Gos = galacto-oligosaccharides; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.

e-Figure 28: Funnel plot with pseudo 95% confidence limits including all the comparisons of the active treatment vs. placebo for NEC-related mortality.



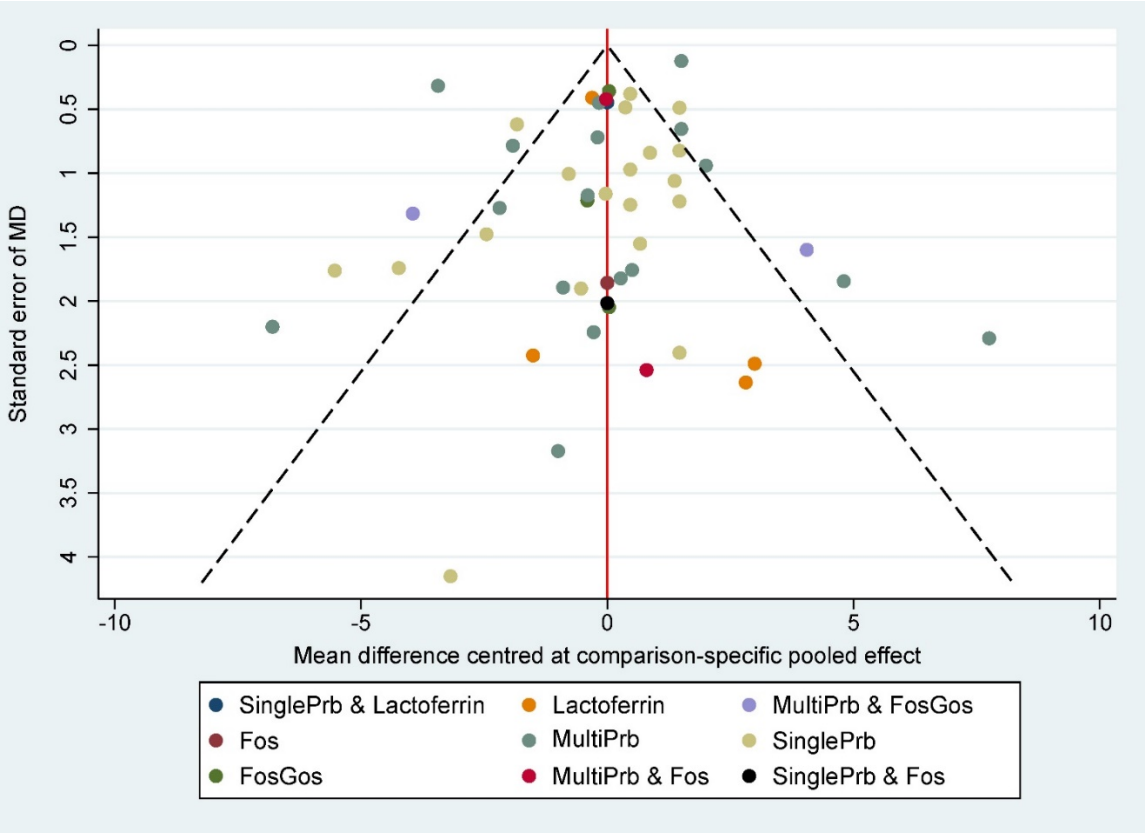
Fos = Fructo-oligosaccharides; Gos = galacto-oligosaccharides; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.

e-Figure 29: Funnel plot with pseudo 95% confidence limits including all the comparisons of the active treatment vs. placebo for feed intolerance.



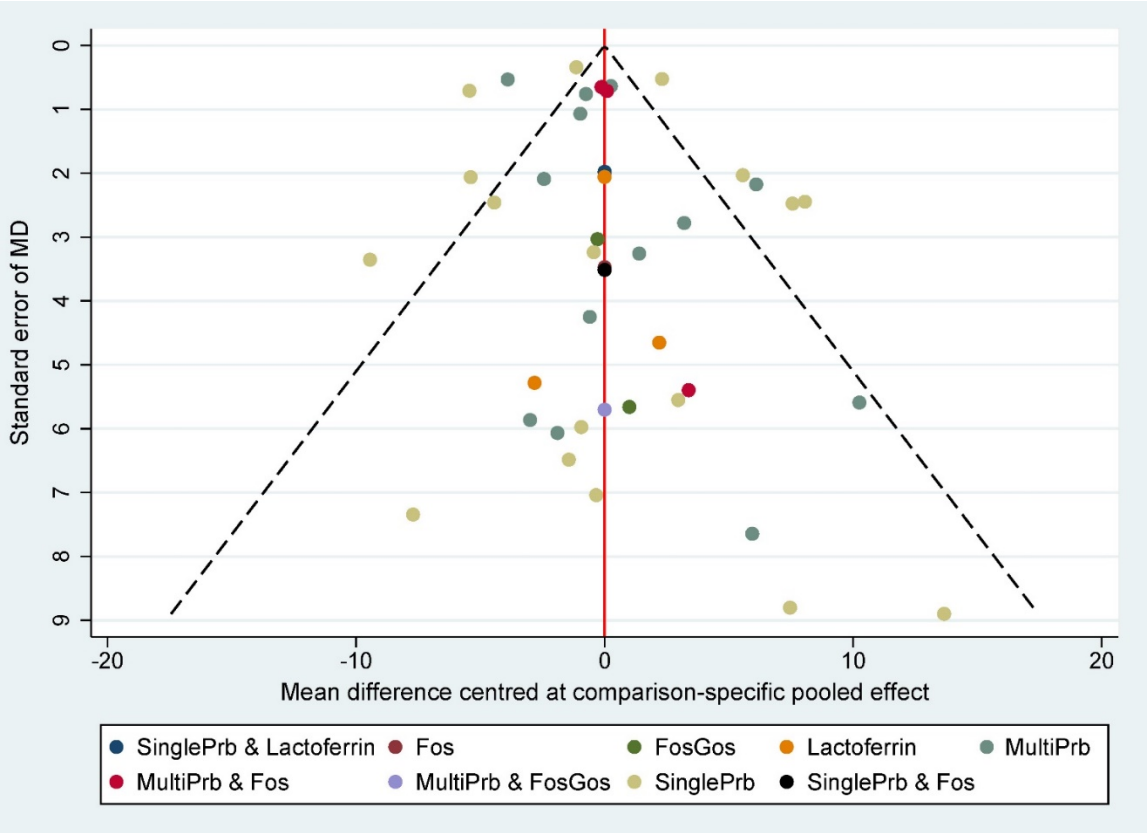
Fos = Fructo-oligosaccharides; Gos = galacto-oligosaccharides; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.

e-Figure 30: Funnel plot with pseudo 95% confidence limits including all the comparisons of the active treatment vs. placebo for time to reach full enteral feed



Fos = Fructo-oligosaccharides; Gos = galacto-oligosaccharides; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.

e-Figure 31: Funnel plot with pseudo 95% confidence limits including all the comparisons of the active treatment vs. placebo for duration of hospitalization



Fos = Fructo-oligosaccharides; Gos = galacto-oligosaccharides; MultiPrb = Multiple-strain probiotics; SinglePrb = Single-strain probiotics.

Supplementary appendices references

1. Bell MJ, Ternberg JL, Feigin RD, et al. Neonatal necrotizing enterocolitis. Therapeutic decisions based upon clinical staging. *Ann Surg* 1978; **187**(1): 1-7.
2. Walsh M, Kliegman R. Necrotizing enterocolitis: treatment based on staging criteria. *Pediatr Clin North Am* 1986; **33**(1): 179-201.
3. Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011; **343**: d5928.
4. Akl EA, Sun X, Busse JW, et al. Specific instructions for estimating unclearly reported blinding status in randomized trials were reliable and valid. *J Clin Epidemiol* 2012; **65**(3): 262-7.
5. Higgins JP, Green S. Cochrane Handbook for Systematic Reviews of Interventions, version [5.1.0] (updated March 2011). The Cochrane Collaboration; 2011.
6. Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. *BMC Med Res Methodol* 2005; **5**: 13.
7. Chaimani A, Higgins JP, Mavridis D, Spyridonos P, Salanti G. Graphical tools for network meta-analysis in STATA. *PloS one* 2013; **8**(10): e76654.
8. Chaimani A, Salanti G. Visualizing assumptions and results in network meta-analysis: The network graphs package. *Stata Journal* 2015; **15**(4): 905-50.
9. Harbord RM, Egger M, Sterne JA. A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. *Stats Med* 2006; **25**(20): 3443-57.
10. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; **327**(7414): 557-60.
11. Lu G, Ades AE. Assessing evidence inconsistency in mixed treatment comparisons. *J Am Stat Assoc* 2006; **101**(474): 447-59.
12. Higgins JP, Jackson D, Barrett JK, Lu G, Ades AE, White IR. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. *Res Synth Methods* 2012; **3**(2): 98-110.
13. Hays S, Jacquot A, Gauthier H, et al. Probiotics and growth in preterm infants: A randomized controlled trial, PREMAPRO study. *Clin Nutr* 2016; **35**(4): 802-11.
14. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008; **336**(7650): 924-6.
15. Puhan MA, Schunemann HJ, Murad MH, et al. A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. *BMJ* 2014; **349**: g5630.

16. Brignardello-Petersen R, Bonner A, Alexander PE, et al. Advances in the GRADE approach to rate the certainty in estimates from a network meta-analysis. *J Clin Epidemiol* 2018; **93**: 36-44.

17. Baker SG, Kramer BS. The transitive fallacy for randomized trials: if A bests B and B bests C in separate trials, is A better than C? *BMC Med Res Methodol* 2002; **2**: 13.

18. Cipriani A, Higgins JP, Geddes JR, Salanti G. Conceptual and technical challenges in network meta-analysis. *Ann Internal Med* 2013; **159**(2): 130-7.

19. Donegan S, Williamson P, D'Alessandro U, Tudur Smith C. Assessing key assumptions of network meta-analysis: a review of methods. *Res Synth Methods* 2013; **4**(4): 291-323.

Chapter 7: Discussion and Future Directions

Summary of findings

This work presents four main pieces of research. The main findings can be summarized as the following:

1. Review of dietary sugar guidelines revealed poor quality of evidence synthesis methods in addition to poor quality of reporting. Further, overall quality of evidence to support recommendations in these guidelines was low to very low. Optimal guidelines should be developed with increased rigor, and recommendations should be specific and transparent. These guidelines rarely followed GRADE guidance as intended.

2. Nutritional interventions are complex and designing randomized trials investigating their effects can be very challenging, particularly if topic areas swerve towards food science and behavioral modification interventions such as the impact of food advertising. The majority of trials are designed poorly and have small sample sizes. All of the above leads to considerable variability in population, intervention, comparison, and outcome assessments across trials, which complicates evidence synthesis.

3. Although standard meta-analysis of direct comparisons is an important and useful tool to assess pooled estimates of the effectiveness and safety of interventions, it has limitations. Nutrition research can benefit from indirect treatment comparison and network meta-analysis as powerful tools that allow multiple treatment comparisons.

4. NMA is a powerful tool that offers the opportunity to synthesize large amounts of data and might improve the precision of the effect estimates in the field of nutrition by using both direct and indirect estimates of effect. Comparison of multiple interventions in a network of trials add complexity to the interpretation of the results from an NMA. As the number of treatments being compared increases, interpretation and presentation of results of NMA becomes more challenging. The challenges increase further when the NMA deals with multiple outcomes. Although, NMA allows ranking interventions in terms of their relative efficacy, ranking has important limitations (1). Efforts thus far have failed to produce a fully satisfactory approach to interpretation and presentation of NMA results under these conditions. Therefore, we have developed and apply an approach that combines relative effect estimates, together with certainty of evidence to help NMA authors presents results in nutrition NMAs in ways that will facilitate optimal interpretation of study findings and optimal understanding in target audiences.

Reflections on an effort to facilitate interpretation and presentation of network meta-analysis results

In addition to pooled effect estimates, rating the certainty or quality of a body of evidence helps with appropriate interpretation of results from an evidence synthesis (2, 3). The GRADE approach is now an established methodology for standard systematic reviews of interventions in medicine that provide direct comparisons. Its methodology in the context of indirect comparisons and NMA was introduced in 2014 (4, 5) and the

methods continue to advance. The most recent advances addressed strategies to handle incoherence or spurious judgments of imprecision in sparse networks (6-8).

Although it is too early to assess the uptake of new advances in GRADE methodology, it has been documented that the rating of certainty of evidence is not commonly performed in standard nutrition reviews, and when it is, authors often neglect to follow GRADE guidelines as intended (9, 10). In the filed of nutrition, assessing the certainty of evidence is rarely used for indirect comparisons and NMAs (11, 12). This may be because NMA methods offer easier alternatives such as probability ranking and SUCRA values, which often can be misleading, for interpreting results and making conclusion. In addition, there is no guidance on how to make conclusions from an NMA, a gap in NMA methods that is important to the field of nutrition.

NMA not only provides an estimate of effect for every possible comparison between pairs of interventions in the network, but also can rank interventions in terms of their relative efficacy (12, 13). Probability rankings and SUCRA values can provide attractive information for users, but can often be misleading for several reasons including ignoring the uncertainty of effect estimates. Ignoring the certainty of the evidence can lead to suggesting treatment option(s) that have very low or low certainty evidence as the best (12, 14, 15).

From a GRADE perspective, this is already an issue in the field of nutrition based on standard direct comparisons based on observational studies that are typically low quality evidence because of residual confounding, and direct comparisons based on

RCTs that often suffer from imprecision (too few patients and too few events) and indirectness (use of surrogate outcomes rather than patient-important outcomes). In addition, by experience, we have realized that in large networks, very seldom- if ever- will an NMA will establish an intervention clearly superior to all others. It is also possible that treatment option(s) can be among the best in effectiveness (benefit) outcomes, but among the worst in safety/harm outcomes. Above all, as the number of treatments being compared increases, interpretation and presentation of results of NMA becomes more challenging.

In this work, we realized the need for further methodological development on how to draw conclusions regarding which treatments are more likely to be superior or inferior to others in terms of effectiveness and harm, considering the estimates of effects, certainty of evidence, and rankings, and to ascertain an intuitive understanding of NMA results. This issue is not unique to the field of nutrition.

To improve the interpretation of NMAs in nutrition, we built upon previous work done by our group to make conclusions from NMA for a single outcome (16). This approach has two guiding principles and 5 steps: the first guiding principle is considering categories of intervention based on their effectiveness (superior, intermediate, and inferior); the second is that categorizing interventions should be based on relative effect estimates and their certainty of evidence, and secondarily considering probability rankings and SUCRA values. The detailed description on the 5 steps can be found in the methods section of Chapter 6 and the resulting tables are presented in Appendix 6.8.

The approach explained above facilitates drawing conclusion from NMA of single outcome, but clinical or policy decision making requires understanding of the effects of interventions across multiple outcomes and the balance of benefits and harms. Given all the statistical and methodological complexities of NMA, providing clinicians and decision makers with the results of this approach does not obviate the need for judgements that might not be simple. Therefore, we developed a new visual presentation of results that combines the outputs of this approach and categorize treatments across outcomes based on certainty of evidence and relative effect estimates, with presentation order informed by the importance of the outcomes.

To create such visual presentation of NMA results (an example can be found in Table 1 - Chapter 6), after applying methods explained above and producing relevant tables for all outcomes, we rated outcomes based on their perceived importance to patients and families. Assessing the importance of selected outcomes for decision making can be based on formal or less formal approaches. In the next step, we identify treatments that are clearly superior and then intermediate in terms effectiveness and harms with moderate-to-high certainty evidence. The next step will be categorizing those treatments with low-to-very low certainty evidence as being clearly superior versus intermediate. Treatment in the middle of table will be those with no clear benefit compared to control/placebo. Those treatments associated with significant harms and no clear benefit compared to control/placebo will be placed at the bottom of the table. The order of outcomes and treatments can be flexible based on values and preferences. We believe this revolutionary approach which places a high emphasis in simplicity and

applicability can be used by authors of NMAs in nutrition to gain a more intuitive understanding of results that consider certainty (quality) of evidence, balance of effectiveness and harms, and importance of outcomes.

Future directions

The use of the GRADE approach to direct comparisons, and to indirect and NMA comparisons in the field of nutrition, and the methodological advances presented here offer new approaches to summarizing evidence that enhance interpretability. However, there is a need to test our new GRADE NMA summary tables across multiple subject areas in future projects. Currently, efforts are underway to user test multiple graphical presentations of our approach for NMA. In addition, we are working to apply current methods in several other NMAs in different subject areas, including popular dietary programs for weight-loss and cardiovascular risk factors, to ascertain the applicability and flexibility of these methods. Although the need for making conclusions from NMA results is evident, our methods require further research based on experience with NMAs across fields, and in particular, in the field of nutrition. Collectively, this research should move us forward in the field, allowing developers and users' of NMA to make results more interpretable for decision making.

References

1. Brignardello-Petersen R, Johnston BC, Jadad AR, Tomlinson G. Using decision thresholds for ranking treatments in network meta-analysis results in more informative rankings. J Clin Epidemiol. 2018;98:62-9.
2. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol. 2011;64(4):383-94.
3. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008;336(7650):924-6.
4. Puhan MA, Schunemann HJ, Murad MH, Li T, Brignardello-Petersen R, Singh JA, et al. A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. BMJ. 2014;349:g5630.
5. Salanti G, Del Giovane C, Chaimani A, Caldwell DM, Higgins JP. Evaluating the quality of evidence from a network meta-analysis. PLoS One. 2014;9(7):e99682.
6. Brignardello-Petersen R, Bonner A, Alexander PE, Siemieniuk RA, Furukawa TA, Rochwerg B, et al. Advances in the GRADE approach to rate the certainty in estimates from a network meta-analysis. J Clin Epidemiol. 2018;93:36-44.
7. Brignardello-Petersen R, Murad MH, Walter SD, McLeod S, Carrasco-Labra A, Rochwerg B, et al. GRADE approach to rate the certainty from a network meta-analysis: avoiding spurious judgments of imprecision in sparse networks. J Clin Epidemiol. 2019;105:60-7.
8. Brignardello-Petersen R, Mustafa RA, Siemieniuk RAC, Murad MH, Agoritsas T, Izcovich A, et al. GRADE approach to rate the certainty from a network meta-analysis: addressing incoherence. J Clin Epidemiol. 2019;108:77-85.
9. Meerpohl JJ, Naude CE, Garner P, Mustafa RA, Schunemann HJ. Comment on "Perspective: NutriGrade: A Scoring System to Assess and Judge the Meta-Evidence of Randomized Controlled Trials and Cohort Studies in Nutrition Research". Adv Nutr. 2017;8(5):789-90.
10. Schwingshackl L, Knuppel S, Schwedhelm C, Hoffmann G, Missbach B, Stelmach-Mardas M, et al. Perspective: NutriGrade: A Scoring System to Assess and Judge the Meta-Evidence of Randomized Controlled Trials and Cohort Studies in Nutrition Research. Adv Nutr. 2016;7(6):994-1004.
11. Schwingshackl L, Buyken A, Chaimani A. Network meta-analysis reaches nutrition research. Eur J Nutr. 2019;58(1):1-3.
12. Schwingshackl L, Schwarzer G, Rucker G, Meerpohl JJ. Perspective: Network Meta-analysis Reaches Nutrition Research: Current Status, Scientific Concepts, and Future Directions. Adv Nutr. 2019:1-16.

13. Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. Res Synth Methods. 2012;3(2):80-97.
14. Mbuagbaw L, Rochwerg B, Jaeschke R, Heels-Andsell D, Alhazzani W, Thabane L, et al. Approaches to interpreting and choosing the best treatments in network meta-analyses. Syst Rev. 2017;6(1):79.
15. Yepes-Nunez JJ, Li SA, Guyatt G, Jack SM, Brozek JL, Beyene J, et al. Development of the Grading of Recommendations Assessment, Development and Evaluation (GRADE) Summary of Findings (SoF) Table for Network Meta-analysis. J Clin Epidemiol. 2019.
16. Florez ID, Veroniki AA, Al Khalifah R, Yepes-Nunez JJ, Sierra JM, Vernooij RWM, et al. Comparative effectiveness and safety of interventions for acute diarrhea and gastroenteritis in children: A systematic review and network meta-analysis. PLoS One. 2018;13(12):e0207701.