NOVEL APPROACHES TO SUPPORT DECISIONS BY GUIDELINE PANELS

DEVELOPMENT OF NOVEL APPROACHES TO SUPPORT THE DECISION- MAKING PROCESS OF GUIDELINE PANELS

By GIAN PAOLO MORGANO, M.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree of Doctor of Philosophy

McMaster University DOCTOR OF PHILOSOPHY (2020)

Hamilton, Ontario (Health Research Methodology Program).

TITLE: Development of novel approaches to support the decision-making process of guideline panels.

AUTHOR: Gian Paolo Morgano, M.Sc. (McMaster University)

SUPERVISOR: Professor Holger J. Schünemann, M.D., Ph.D., M.Sc.

NUMBER OF PAGES: xxvi, 213

LAY ABSTRACT

Clinical practice guidelines assist health care professionals in selecting management options that can best improve the health outcomes of their patients. The development of trustworthy guidelines is a complex process that requires the contribution of several entities. The guideline panel, which typically comprises different experts (clinicians, patient representatives, experts in research methodologies) plays the key role in this process as it is responsible for selecting the most important questions to address in the guideline, reviewing the evidence supporting an option, agreeing on the recommendations, and endorsing the final guideline document. To ensure that the process of developing guidelines is transparent and that the recommendations are credible, it is important that panel activities are well documented and follow rigorous methods. Structured frameworks, such as the Grading of Recommendations Assessment, Development and Evaluation (GRADE) Evidence to Decision (EtD) approach, have been developed to systematically guide the panel members and to minimize the error that could be introduced while making decisions. In this thesis, I describe the development of an approach and its application for comprehensive guideline development by the Italian National Health Institute, to describe rigorous guideline development and propose two novel approaches to further assist panel members in enhancing their guideline development. The first of these two enhancements to guideline development describes how to derive a modelled estimate of the risk of having certain health conditions when this data is not directly available in the medical literature. The second of the two

enhancements is a method to support guideline panels in judging how substantial the desirable and undesirable effects of health interventions are. Both approaches were tailored to fit specific needs but can be adapted to inform the improvement of other steps in the guideline development process.

ABSTRACT

Trustworthy clinical practice guidelines assist health care professionals in selecting the management options that optimize patient health outcomes. The development of trustworthy guidelines requires the consideration of many aspects and the involvement of multiple contributors, often working in groups. The guideline panel plays the key role in the development process as it is responsible for prioritizing topics that should be covered as part of the guideline effort, formulating questions, reviewing the evidence, developing and agreeing on the recommendations, and endorsing the final guideline document. Ensuring transparency throughout the process by appropriately organizing and documenting panel activities is an essential standard that is used to assess the credibility of a developed guideline and its resulting recommendations. The adoption of conceptual frameworks that systematically guides panel members in their decision-making process (e.g. the Grading of Recommendations Assessment, Development and Evaluation (GRADE) Evidence to Decision (EtD) frameworks) can aid in the formulation of methodologically sound recommendations. In this dissertation, I used the example of a guideline on diagnosis and treatment of autism spectrum disorders to describe how rigorous research methods can support guideline panels in the development process from early stages to the formulation of recommendations. In another prominent guideline development effort with the American Society of Hematology, I have identified two steps in the process where panel members may benefit from further support and addressed these gaps by conceptualizing and developing novel approaches. The first approach

comprises modelling baseline risk estimates for patient-important outcomes when only surrogate data is available. The second approach proposes a method to estimate decision thresholds for judgments on health benefits and harms using the GRADE EtD framework. While these approaches are tailored to address specific guideline panel needs, guideline methodologists could use the underlying concepts to find solutions to aid guideline panels in other steps of the development process.

ACKNOWLEDGMENTS

My first thought goes to my mentor and inspiration, Dr. Paola Muti. She was the first person I met in the field of research and ever since has guided my trajectory. She supported me in every step of my career as well as helped me in facing and recovering from personal tragedies. Without her, I would have never had the opportunity to join McMaster University and reach such personal and professional growth.

My boundless gratitude goes also to my PhD supervisor, Dr. Holger J. Schünemann. With a *Michael-Jordanesque* drive and talent, he leads his team and makes his students better researchers and stronger people. He believed in me when I was repeatedly questioning my abilities and taught me to not underestimate the importance of time and experience. He brought me to stages that were beyond my wildest expectations. I feel blessed by being part of his team and having the possibility to learn right next to him.

I would like to express my gratitude to the members of my thesis committee: Dr. Jan Brozek, Dr. Nancy Santesso, and Dr. Feng Xie. They have been a significant part of this journey and with great passion, competence, and patience, shared with me their knowledge and guided my studies.

I want to thank all my friends and peers in the outstanding Health Research Methodology (HRM) program and department of Health Research Methods, Evidence, and Impact, for

making of 'GP' a much better 'GP'. Your friendship is a treasure that goes beyond any PhD.

Finally, I want to dedicate this achievement to my family and all the people I loved who are now looking at me from heaven. I know that without your support this would have been just impossible.

Gian Paolo Morgano

Hamilton, Ontario, August 2020

PREFACE

The work in this dissertation is presented as a "sandwich thesis" that includes three manuscripts which have been accepted, submitted for publication or prepared for submission. The manuscript in Chapter 2, "Introduction and methods of the evidence-based guidelines for the diagnosis and management of autism spectrum disorder by the Italian National Institute of Health", was published on 9 March 2020 in *Health and Quality of Life Outcomes*. The manuscript in Chapter 3, "A modeling approach to derive baseline risk estimates for GRADE recommendations: Concepts, development, and results of its application to the American Society of Hematology 2019 guidelines on prevention of venous thromboembolism in surgical hospitalized patients", was submitted for publication on 5 February 2020 in the *Journal of Clinical Epidemiology* and is currently under review. The manuscript in Chapter 4, "Defining Decision Thresholds for judgments on health benefits and harms using the Evidence to Decision Framework: concepts, methods, and preliminary results", will be submitted to the *Journal of Clinical Epidemiology*.

The manuscript presented in Chapter 2 was an effort conducted to describe the results of applying the new methodological standards for guideline development of the Italian National Institute of Health (ISS) to two guidelines on diagnosis and management of children/adolescents and adults with autism spectrum disorders. I contributed to the development of the methods for the ISS and then applied them in development of these

guidelines as methodologist. I facilitated the interaction between the groups involved in the process. I along with my supervisor, Dr. Holger J. Schünemann, drafted the manuscript which was circulated to co-authors. I incorporated feedback from the co-authors and submitted the manuscript for publication which was subsequently published. Chapter 3 was an effort conducted to inform the 2019 American Society of Hematology guidelines. My supervisor, Dr. Holger J. Schünemann, conceptualized the modeling approach to which I contributed. I performed initial selection of the sources of baseline risk for all guideline questions and calculated modeled estimates of baseline risk. I drafted the manuscript which was circulated to co-authors. I incorporated feedback from the co-authors and prepared the manuscript for submission. The work in Chapter 4 was conceived of and conducted under the supervision of Professor. J. Holger Schünemann and the methodological input of Professors Xie, Brożek, and Santesso. I drafted the survey including the scenario used to elicit data from content experts, analysed and interpreted the data. I drafted the manuscript, incorporated feedback from the co-authors, and prepared the manuscript for submission.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter two

Chapter three

Chapter four

LIST OF TABLES

Chapter two

Chapter three

LIST OF APPENDICES

Chapter two

Chapter four

LIST OF ABBREVIATIONS AND SYMBOLS

ASD: Autism Spectrum Disorder

ASH: American Society of Hematology

CI: Confidence Interval

CNEC: Centro Nazionale per l'Eccellenza Clinica, la Qualità e la Sicurezza delle Cure

COI: Conflict of Interest

CPGs: Clinical Practice Guidelines

DTs: Decision Thresholds

DVT: Deep Venous Thrombosis

EtD: Evidence-to-Decision

ERT: Evidence Review Team

GESI: Global Evidence Synthesis Initiative

GIN: Guidelines International Network

GRADE: Grading of Recommendations Assessment, Development and Evaluation

GRADEpro: GRADEpro Guideline Development Tool

HOD: Health Outcome Descriptor

HTA: Health Technology Assessment

ISS: Istituto Superiore di Sanità (Italian National Institute of Health)

NNT: Number Needed to Treat

PE: Pulmonary Embolism

PICO: population, intervention, comparator, outcomes

PROs: Participant Reported Outcomes

PRU: PICO Responsible Unit

RCT: Randomized Controlled Trial

RD: Risk Difference

RR: Risk Ratio

SNLG: Sistema Nazionale Linee Guida (Italian National Guidelines System)

SoF: Summary of Findings

VTE: Venous Thromboembolism

DECLARATION OF ACADEMIC ACHIEVEMENT

I declare that I, jointly with my supervisor, Professor Holger J. Schünemann, played the primary role in the conception, design, and conduction of the studies here included. We obtained feedback and advice from Professors Xie, Brozėk, and Santesso, as well as from methodologists from McMaster University.

This work is original research that I conducted. I am the principle contributor and first author of all the manuscripts contained in this dissertation.

I am responsible and made the following contributions in all projects included in this work: design, conception, and writing of materials; I contributed in the development process the autism guidelines including the questions prioritization process, synthesis of the evidence, formulation of recommendations, and interaction with stakeholders. I contributed to the final draft of the diagram outlining the assumptions around the distribution of outcome events made by the panel and calculated modeled estimates of baseline risk. I designed the prospective and retrospective collection of data to derive decision thresholds including the survey and the scenarios described. I conducted all analyses, designed figures and tables, and organized meetings. I wrote the manuscript with editorial advice and supervision of Professor Schünemann, and from feedback from Professors Xie, Brozėk, and Santesso. The co-authors on each paper contributed significantly with important comments and advice for the final manuscripts (details with each manuscript).

The first paper was published in *Health and Quality of Life Outcomes* in March 2020. The second paper was submitted to the *Journal of Clinical Epidemiology* in February 2020 and is under review. The target journal for third paper will be the *Journal of Clinical Epidemiology*.

CHAPTER 1. INTRODUCTION

1. Clinical practice guidelines in context

Healthcare, defined as the provision of medical care to individuals or a community, is a complex universe navigated by healthcare professionals, patients, caregivers, researchers, and policy makers among others. [1] Clinical practice guidelines (or guidelines) serve many healthcare needs by acting as navigational stars that guide all those involved in their journey towards the common destination of improving health outcomes. Guidelines may support practitioners in their clinical practice, empower patients and caregivers during their decision-making process, guide researchers in identifying future research priorities, and inform policy-makers in deriving performance measures for quality improvement initiatives. Given their relevance and widespread use, it is essential that guidelines be based on sensible knowledge to appropriately inform healthcare practice.[2] Over the last three decades, epistemological principles of Evidence Based Medicine (EBM), have shown to be the solid foundations on which guidelines are produced.[3] EBM is defined as the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients.[4] According to the National Academy of Medicine of the United States (NAM, formerly known as the Institute of Medicine), the systematic evaluation and critical appraisal of the body of evidence available from scientific literature, without the biased selection of evidence that favours a particular claim, is now considered a requirement in guideline development. The NAM defines guidelines as

'*statements that include recommendations intended to optimize patient care that are informed by a systematic review of evidence and an assessment of the benefits and harms of alternative care options*'.[5]

2. Research methods to support guideline panels

The development of trustworthy guidelines requires the consideration of a large number of diverse topics and the involvement of multiple working groups. The guideline panel plays the key role in the guideline development process as they are responsible for prioritizing topics that should be covered as part of the guideline effort, formulating questions, developing and agreeing on the recommendations in the guideline, and endorsing the final guideline document.[6] Ensuring transparency throughout the process by appropriately organizing and documenting panel activities is an essential standard used to assess credibility of a developed guideline and its resulting recommendations.[5,7] Researchers have invested in conceptualizing, refining, and disseminating structured methods for the development of clinical practice guidelines. Most noticeably, the Grading of Recommendations Assessement, Development and Evaluation (GRADE) Working Group, a collaborative initiative with over 600 members from various countries, has developed a sensible and transparent approach that is currently adopted by more than 100 organizations including the Wold Health Organization (WHO), the European Commission (EU), the Italian National Institute of Health (ISS) and the American Society of Hematology (ASH).[8] This approach includes a system to assess the certainty of the

evidence and strength of recommendations and the use of Evidence to Decision (EtD) frameworks, developed by the GRADE Working Group, that guide the process of moving from evidence to recommendations.[9] The EtD framework represents a successful example of how health research methodologies can support the activities of guideline panels and improve the overall transparency and quality of the development process. The EtD framework requires guideline panels to be explicit in their judgments by making the basis for their decisions transparent to target audiences. Also, EtD frameworks may help ensure that decisions are informed by the best available evidence and that all relevant criteria for a decision, including resource use, equity, feasibility, and acceptability, are considered.[9]

3. Why is this research important?

The endorsement of research methods for guideline development can aid organizations in meeting the standard criteria for high-quality guidelines.[10] In this dissertation, we emphasize this notion by describing a guideline development process guided by the Italian National Institute of Health methodological manual to which I contributed and derived from the GIN- McMaster Guideline Development Checklist.[6,11] The adoption of a conceptual framework, such as the GRADE EtD that systematically guides panel members in their decision-making process, supports the development of methodologically sound recommendations. However, it is possible to identify steps in the process where guideline panels might benefit from further support and for which novel methods could

be developed. In this dissertation, we identify two such steps and propose solutions through the use of conceptualized novel approaches. While these approaches are tailored to address specific guideline panel needs, the underlying concepts should be applicable by guideline methodologists to find solutions that could assist guideline panels in other steps of the development process.

4. Goals and scope

This dissertation aims to:

a. Delineate how systematic and transparent methods for clinical practice guideline development can reduce the risk of bias posed by panel member activities. Also, to provide a reference standard for future guideline development efforts in the Italian setting.

b. Describe the conceptualization and application of a modeling approach to derive baseline risk estimates for GRADE recommendations. The goals of this approach are twofold: to provide guidance on how to derive modeled estimates of baseline risk in absence of direct data, and to increase transparency while reducing potential error in the decision-making process.

c. To derive Decision-Thresholds (DTs) from empirical data that can be used by guideline panels while making EtD judgements on the desirable and undesirable effects of a health intervention for a single, dichotomous, outcome. This work provides empirical evidence and represents a stepping-stone towards developing GRADE guidance on decision-thresholds for judgements across multiple outcomes.

We conducted three main studies to address each of the goals listed above. The first main study represents original work describing the application of the new standards of the Italian National Institute of Health to two guidelines on diagnosis and management of children/adolescents and adults with autism spectrum disorders. In describing a guideline

development process, led with my supervisor Holger J. Schünemann, we focused on the scoping of the guideline, panel composition, management of conflict of interest, generation and prioritization of research questions, and early stakeholders' involvement.

The second study entailed the development and application of a modeling approach to derive baseline risk estimates. We tested this approach in a guideline effort on the prevention of venous thromboembolism in hospitalized surgical patients that was developed in collaboration with the ASH. The approach includes guidance on how to calculate modeled estimates of baseline-risk and addresses potential bias of over- or underestimating anticipated absolute effects of interventions when using surrogate outcomes.

In the third, and final study, we propose Decision-Thresholds (DTs) that can be used by panel members to differentiate across EtD judgments and serve as references for interpreting findings. To achieve this, we conceptualized DTs as a joint measure of the magnitude of the effects, such as the proportion of people who would benefit, and the importance of the outcome, such as how much it is valued by the people affected. We carefully and meticulously designed the approach and began to survey decision-makers to collect empirical data to derive the DTs, and then verified our DTs by comparing them to DTs derived from a retrospective analysis of judgments made in existing guidelines.

5. Thesis overview

The overall aim of this dissertation is to improve methods to assist guideline panel members in their judgements and reduce potential error in their decision-making process. Chapter 2 provides an example of how systematic and transparent approaches in guideline development can reduce the risk of bias posed by panel member activities. It also investigates how guideline panels complied with methodologies for guideline development and suggests potential solutions for group training. Chapters 3 and 4 introduce novel approaches. The goal of the study presented in Chapter 3 was to develop an approach to model baseline risks for patient-important outcomes for guideline recommendations that could be used when only baseline risks for surrogate outcomes are available. We described the methods used for the development of this approach and its practical application in a guideline on prophylaxis of venous thromboembolism. In Chapter 4, we sought to derive DTs that could assist guideline panel members when evaluating how substantial the anticipated effects are for any potential combination of interventions' effects and outcome. Finally, Chapter 5 presents an overall summary of findings and a discussion including implications for future research and practice.

References

1. Adams O, Shengelia, B., Stilwell, B., Larizgoitia, I., Issakov, A., Kwankam, S. Y. & Jam, F. S. T. . Provision of Personal and Non-personal Health Services: Proposal For Monitoring. World Health Organization 2002.

2. Brennan C, Greenhalgh J, Pawson R. Guidance on guidelines: Understanding the evidence on the uptake of health care guidelines. J Eval Clin Pract 2018;24:105-16.

3. Djulbegovic B, Guyatt GH. Progress in evidence-based medicine: a quarter century on. Lancet 2017;390:415-23.

4. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ 1996;312:71-2.

5. Institute of Medicine Committee on Standards for Developing Trustworthy Clinical Practice G. Clinical Practice Guidelines We Can Trust. Washington (DC): National Academies Press (US); 2011.

6. Schünemann HJ, Wiercioch W, Etxeandia I, et al. Guidelines 2.0: systematic development of a comprehensive checklist for a successful guideline enterprise. CMAJ 2014;186:E123-E42.

7. Brouwers MC, Kho ME, Browman GP, et al. AGREE II: advancing guideline development, reporting and evaluation in health care. CMAJ 2010;182:E839-42.

8. Atkins D, Eccles M, Flottorp S, et al. Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches The GRADE Working Group. BMC Health Serv Res 2004;4:38. doi: 10.1186/472-6963-4-38.

9. Alonso-Coello P, Schunemann HJ, Moberg J, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction. BMJ 2016;353:i2016.

10. Qaseem A, Forland F, Macbeth F, et al. Guidelines International Network: toward international standards for clinical practice guidelines. Ann Intern Med 2012;156:525-31.

11. Manuale metodologico per la produzione di linee guida di pratica clinica. Centro Nazionale per l'Eccellenza Clinica, la Qualità e la Sicurezza delle Cure dell'Istituto Superiore di Sanità. 2018. (Accessed 30 January 2019, at https://snlg.iss.it/wp-content/uploads/2019/02/MM_v1.3_feb_2019.pdf.)

CHAPTER 2. INTRODUCTION AND METHODS OF THE EVIDENCE-BASED GUIDELINES FOR THE DIAGNOSIS AND MANAGEMENT OF AUTISM SPECTRUM DISORDER BY THE ITALIAN NATIONAL INSTITUTE OF HEALTH

Authors:

Gian Paolo Morgano[1], Francesca Fulceri[2], Franco Nardocci[2], Corrado Barbui[3], Giovanni Ostuzzi[3], Davide Papola[3,4], Laura Maria Fatta[2], Alice Josephine Fauci[5], Daniela Coclite[5], Antonello Napoletano[5], Franco De Crescenzo[6,7], Gian Loreto D'Alò[6,8], Laura Amato[6], Michela Cinquini[9], Primiano Iannone[5], Holger Jens Schünemann[1,10*], Maria Luisa Scattoni[2*] on behalf of the Italian National Institute of Health guideline working group on Autism Spectrum Disorder[#].

[1]Michael G DeGroote Cochrane Canada Centre and McMaster GRADE Centres, Department of Health Research Methods, Evidence and Impact (formerly Clinical Epidemiology and Biostatistics), McMaster University, Hamilton, Canada;

[2]Research Coordination and Support Service, Istituto Superiore di Sanità, Viale Regina Elena 299, 00161, Rome, Italy;

[3]WHO Collaborating Centre for Research and Training in Mental Health and Service Evaluation, Department of Neuroscience, Biomedicine and Movement Sciences, Section of Psychiatry, University of Verona, Verona, Italy;

[4]Department of Clinical, Neuro and Developmental Psychology; Faculty of Behavioural and Movement Sciences, Vrije Universiteit, Amsterdam, The Nederlands;

[5]Centro Eccellenza Clinica, Qualità e Sicurezza delle Cure, Istituto Superiore di Sanità, Viale Regina Elena 299, 00161, Rome, Italy;

[6]Department of Epidemiology, Lazio Regional Health Service, Rome, Italy;

[7]Department of Psychiatry, University of Oxford, Oxford, United Kingdom;

[8]School of Hygiene and Preventive Medicine, University of Rome Tor Vergata, Rome, Italy;

[9]Mario Negri Institute for Pharmacological Research IRCCS, Milan, Italy;

[10]Department of Medicine, McMaster University, Hamilton, Canada.

[*]These authors contributed equally

[#]The members of the working group are listed in "acknowledgments".

Corresponding authors:

Holger Schünemann, MD, PhD

Michael G DeGroote Cochrane Canada Centre and McMaster GRADE Centres, Department of Health Research Methods, Evidence and Impact (formerly Clinical

Epidemiology and Biostatistics), McMaster University, 1280 Main Street West,

Hamilton, Canada

Email: schuneh@mcmaster.ca

Maria Luisa Scattoni, PhD

Research Coordination and Support Service,

Istituto Superiore di Sanità,

Viale Regina Elena 299, 00161 Rome, Italy

Email: marialuisa.scattoni@iss.it

Word count

Abstract: 315

Body: 4504

Abstract

Background: Autism Spectrum Disorder (ASD) is a neuro-developmental disorder that affects communication and behavior with a prevalence of approximately 1% worldwide. Health outcomes of interventions for ASD are largely Participant Reported Outcomes (PROs). Specific guidelines can help support the best care for people with ASD to optimize these health outcomes but they have to adhere to standards for their development to be trustworthy.

Objective: The goal of this article is to describe the new methodological standards of the Italian National Institute of Health and novel aspects of this guideline development process. This article will serve as a reference standard for future guideline development in the Italian setting.

Methods: We applied the new standards of the Italian National Institute of Health to the two guidelines on diagnosis and management of children/adolescents and adults with ASD, with a focus on the scoping, panel composition, management of conflict of interest, generation and prioritization of research questions, early stakeholders' involvement, and PROs. Recommendations are based on the Grading of Recommendations Assessment, Development and Evaluation (GRADE) Evidence-to-Decision frameworks.

Results: Following a public application process, the ISS established two multidisciplinary panels including people with ASD and/or their caregivers. Seventy-nine

research questions were identified as potentially relevant for the guideline on children and adolescents with ASD and 31 for the one on adults with ASD. Questions deemed to have the highest priority were selected for inclusion in the guidelines. Other stakeholders valued their early involvement in the process which will largely focus on PROs. The panels then successfully piloted the development of recommendations using the methodological standards and process set by the ISS with a focus on PROs.

Conclusions: In this article, we describe the development of practice guidelines that focus on participant reported outcomes for the diagnosis and management of ASD based on novel methods for question prioritization and stakeholder involvement. The recommendations allow for the adoption or adaptation to international settings.

Keywords:

Autism Spectrum Disorder; Italian National Institute of Health; Italian National Guidelines System; GRADE approach; Healthcare decision; Diagnosis; Treatment; Recommendations; Guideline.

1. Introduction

Clinical Practice Guidelines (CPGs) are statements containing recommendations for clinical practice or public health policy. A recommendation describes, for the intended end-user of the guideline, what he or she can or should do in specific situations to achieve the best health outcomes possible, individually or collectively.(1) Besides their primary objective to improve health outcomes through the promotion of evidence-based care and clinical pathways, CPGs also serve as a resource for patients, caregivers, policy-makers, researchers and regulatory bodies.

1.1 Clinical practice guidelines in Italy

Following novel national regulations on responsibilities of healthcare professionals (2), the Italian National Institute of Health (Istituto Superiore di Sanità, ISS) through the recently instituted Centre for Clinical Excellence, Quality and Safety of Care (Centro Nazionale per l'Eccellenza Clinica, la Qualità e la Sicurezza delle Cure, CNEC), is responsible for the governance of the Italian guidelines production process.(3) In this framework, the new Italian National Guidelines System (Sistema Nazionale Linee Guida, SNLG) was established as the pivotal instrument to promote an efficient production mechanism of good quality national guidelines, and the methodological standards recommended for the development and evaluation of CPGs were set. Based on international standards such as the Guidelines International Network (GIN) - McMaster guideline development tool, rigorous methods, combined with systematic and transparent

processes, are required by the ISS in its recently published methodological manual for CPGs development.(4-6) These regulations have not been previously applied in Italian national guidelines but are now a requirement in any CPG developed by ISS and, thus, in these two new ISS guidelines for managing ASD.

1.2 Autism spectrum disorder and current guidelines on its diagnosis and treatment in context

The essential behavioral features of ASD are persistent impairment in reciprocal social communication and social interaction and restricted, repetitive patterns of behavior, interests, or activities. These core symptoms are present from early childhood and limit or impair everyday functioning (7), are extremely heterogeneous both in terms of complexity and severity and vary over time. Recent systematic review and large observational research reported a prevalence of ASD in adults ranging from 0.7% to 1.1% (8, 9), while a recent study performed by the National Observatory for ASD (coordinated by ISS and the Ministry of Health) revealed that approximately 1.3% of children in the age range 7 to 9 have been diagnosed with ASD in Italy.(10) People with ASD frequently present co-occurring neurological, psychiatric and medical disorders that must be considered for the organization of the appropriate interventions. Outcomes related to interventions, both from tests and management strategies, are typically reported by people with ASD or caregivers. A considerable number of different approaches to diagnose and manage ASD have been proposed over the last 50 years. This reflects the complexity of

this condition, which requires a balance between medical, psychological, social, educational and even ethical and existential needs. Many of these approaches have been object of academic and public debate, often with overt disagreement between researchers, clinicians, people with ASD, family and caregivers, and other stakeholders.(11, 12)

These factors represent challenges for the development of evidence-based guidelines in this field. In 2011, the ISS published the first Italian Guideline on ASD entitled *'The treatment of children and adolescents with ASD'*. (13) The published recommendations have been very controversially debated by professionals, institutions and parents' associations.(14) In 2015, a new law demanded an update of these guidelines and the Italian Ministry of Health appointed the ISS to coordinate the development of national guidelines on management of ASD throughout the lifespan. As opposed to the previous version, these new guidelines will also include diagnostic questions and provide, separately, recommendations for the population of children and adolescents with ASD and for adults with ASD. Furthermore, these guidelines will have to adhere to new methods outlined by CNEC within the framework of the new SNLG and comply with its innovations such as a policy for the disclosure and management of Conflict Of Interest (COI), transparent stakeholder involvement and adoption of the GRADE approach.(15, 16) Yet, these methods have not been tested in real guideline development in this new legal framework.

1.3 Objectives of this article

This article introduces the methods and approach to guideline development at the ISS, laying out its innovative methods using the example of evidence-based guidelines for the diagnosis and management of ASD with a focus on PROs.

2. Methods

The guideline development process was guided by the ISS methodological manual (4) and derived from the GIN-McMaster Guideline Development Checklist (6) (https://cebgrade.mcmaster.ca/guidelinechecklistonline.html). It was intended to meet recommendations for trustworthy guidelines by the National Academy of Medicine (U.S.), formerly known as the Institute of Medicine, the World Health Organization and the GIN.(5, 17, 18)

2.1 Scope of the guideline

The Italian national law *'Provisions on prevention, treatment, and rehabilitation of people with autism spectrum disorders, and assistance to families'* (Law 134, approved by the Italian Government on August 2015) intends to ensure the health, the improvement of living conditions and the inclusion in social and working environments of individuals with ASD. The two guidelines described here will be developed in observance of the law

134. Their scope includes the diagnosis and management of ASD and requires describing the perspective, objectives, target population, and target audience.

2.2 Participants in the process

The guideline working group benefits from the contribution of several teams. We describe their roles and responsibilities here.

The ISS Steering Committee (SC) leads and oversees the development of the guideline, it defines the groups involved (chairs, developers, panel, evidence review team) and supports their productive interaction and it is responsible for the development process including budgeting, the definition of a timeline, and the management of COI. The SC, coordinated by a principal investigator and supported by scientific and technical secretariat, selected two chairs for each guideline: one content expert and one methodological expert. The chairs are included in the SC together with a quality assurance team that ensures the compliance of the development process with the ISS methods and regulations.

The guideline group or panel is responsible for prioritization of questions for the guideline, participation in group meetings and teleconferences, providing input on evidence and contextual factors, reviewing evidence summaries, making judgments and formulating recommendations in final panel meetings, reviewing and writing of final guideline report and support for dissemination.(19) Two separate panels have been

selected, each focusing on one of the ASD populations of interest: children/adolescents and adults. Considering that the management of ASD, from diagnosis to the delivery of comprehensive care, involves a heterogeneous group of professionals and competencies, the two panels were designed to be multidisciplinary and geographically representative of the entire Italian territory. Through a public process (20), we invited representatives of fields relevant to the guideline's scope with at least five years of experience and working for the Italian national healthcare system (either in the local health units or in the university/research hospitals) to voluntarily participate in the guidelines. The invitation included representativeness of people with ASD and/or their caregivers. Based on the analysis of their curriculum vitae, cover letter and years of personal or professional experience in the ASD field, sixteen panel members have been selected. All panel members have been invited to sign a declaration of commitment and confidentiality and fill in the COI form. The guideline methodologists or developers, trained in the GRADE approach and the use of the GRADEpro Guideline Development Tool (GRADEpro, https://gradepro.org), work closely with the guideline panel in prioritizing the relevant questions and outcomes, prepare background documents for the guideline panel and stakeholders, coordinate teleconferences and online voting processes, review comments. The Evidence Review Team (ERT) searches the literature and produces syntheses of the evidence. Following the GRADE approach, the ERT rates the certainty in the evidence, prepares the GRADE evidence tables and Evidence-to-Decision (EtD) frameworks that the panels use in formulating recommendations.

2.3 Management of conflict of interest

The ISS policy on the management of COI follows the GIN principles for disclosure of interests and management of COI in guidelines (21) and it is described in the ISS manual.(4) According to this policy, those involved in the guideline development, including panel members, the ERT, guideline developers and external referees, had to declare all financial, non-financial, personal and institutional interests relevant to the scope of the guidelines completing a standardized form. The SC evaluated each individual interest based on its nature and type, specificity with respect to the scope of guideline, financial value, period and duration. If a declared interest was deemed to represent a conflict, the following measures for the management of COI were applied: full participation, with public disclosure of interest; partial exclusion (e.g. exclusion from the works related to the declared interest and from the relevant decision-making process); total exclusion. We applied the policy throughout the entire process, including during panel members selection, generation and prioritization of research questions, and participation in the formulation of recommendations. We regularly monitored and updated declarations of COI.

2.4 Opening meeting and training of the guideline panel

The working group met for the first time in a two-day meeting held at the ISS headquarter. The following activities took place: the SC outlined the scope of the guideline; guideline developers presented the existing guidelines on ASD; the working

group discussed the resources and time available and agreed to produce recommendations on 16 research questions for each of the two ASD identified populations over an 18-month time period. The guideline quality assurance team presented the ISS policy on COI and collected COI disclosure forms from participants. The ERT introduced the GRADE methodology in two presentations. The first presentation served as introduction to the GRADE constructs of certainty in the evidence and strength of recommendations.(22, 23) The second focused on GRADE evidence tables, GRADE EtD frameworks and the importance of people's values and preferences in decision-making processes.(24-27) We shared links to training material, including the ISS manual and online resources on the GRADE approach to rating the certainty of evidence and the EtD frameworks to meeting participants. The meeting served for the members of the working group to get to know each other and to commence collaboration.

2.5 Selection of guideline questions

We implemented a two-step approach that allowed panels to identify and agree on the questions to be addressed in the guidelines using the module in GRADEpro that allows for the generation and prioritization of questions and health outcomes GRADEpro.(28)

2.6 Generation of questions

Guideline developers drafted a list of strategies and interventions addressed in existing CPGs on the diagnosis and management of ASD.(29-32) We discussed the list during the

opening meeting and invited panel members to identify items missing or deemed not applicable to the Italian context. Based on the output of the meeting, subgroups including guideline developers and members of the panel with specific expertise (content experts) generated a list of candidate questions framed using the PICO format (population, intervention, comparator, and outcomes).(33) To streamline the initial list, questions were organized by category (e.g. questions pertaining to the diagnosis, pharmacological, or psychosocial interventions) and, where appropriate, grouped together. The grouping was applied when interventions were assumed to share similar functioning or having similar effects on health outcomes (e.g. medications belonging to the same drug class) and for similar diagnostic instruments. We presented the list of candidate questions to the groups during two-hour recorded web-based conferences.

2.7 Prioritisation of questions

Once the list of candidate questions was finalized, we asked panels to rate the priority of questions on a 1 to 9 scale. We used surveys electronically generated in GRADEpro (Figure 1) and applied the following criteria: rating of 7 to 9, high priority question - should be addressed in the guideline; rating of 4 to 6, priority question but not of high priority - should be listed as priority in the guideline; rating of 1 to 3, not a priority question - it is acceptable to neither include nor mention it in the guideline.

We invited panel members to consider a brief list of factors that typically influence whether a question is relevant in the context of a CPG (Table 1).

We also provided supplementary materials including a glossary of the acronyms used to formulate the questions and articles related to the underpinning theoretical frameworks considered to organize the questions into categories. Following the rating exercise, we presented the results (means, median, minimum and maximum) to the groups in separate two-hour teleconferences using the mean rating score as a ranking criterion. We invited the groups to critically appraise the list and to evaluate its harmony. In particular, we asked to verify if any of the top-rated questions for inclusion could not be considered as exhaustively informative to the reader if not paired with another question that was not rated for inclusion. To achieve harmony, we also organized questions in sensible units, consisting of the smallest recommendation sets that would be informative or required for readers to avoid gaps and achieve rapid dissemination.(34) We used the sensible units to streamline the production and dissemination of recommendations and to create working sub-groups for each, also known as the PICO Responsible Unit (PRU), consisting of content experts and members of the ERT.(35)

2.8 Generation of outcomes

To determine the people-important outcomes to be addressed in the syntheses of the evidence, we first engaged the PRUs in drafting descriptions of potentially relevant desirable and undesirable outcomes. We created written definitions of outcomes, known also as *health outcome descriptors*, to reduce the risk of introducing error that could result when panel members have different understanding of the same outcomes. We then sent

GRADEpro surveys asking to add, for each question separately, potentially relevant people-important outcomes that were not yet included in the list drafted by the PRUs (Figure 2).

2.9 Prioritisation of outcomes

We elicited ratings of the relative importance of outcomes on a 1 to 9 scale (Figure 3) in the corresponding GRADEpro module. We asked the panels to rate outcomes separately for each question using the following criteria: a rating of 7 to 9, the outcome is critical for decision-making; 4 to 6, the outcome is important but not critical for decision-making; 1 to 3, the outcome is of low importance for decision-making.(36)

Similar to the question prioritization step, we provided guidance materials on the task and its underpinning concepts, available in appendix 1. We discussed the results of the rating exercise (means, minimum and maximum) in a face-to-face meeting using the mean rating score as the ranking criterion and considered only outcomes rated as critical or important for inclusion in systematic reviews and decision-making during formulation of recommendations. Once the list of outcomes was prioritized, we reached consensus on the final list of questions as described above (2.7).

2.10 Stakeholders' involvement

Public involvement in the development of ISS CPGs is guaranteed through the participation of lay members in the panel as well as also through a public consultation on

two key outputs of the process: draft list of guideline questions and draft recommendations. As for the former, we made the list of prioritized questions available for comments by stakeholders who met eligibility criteria.(37) The stakeholders were organized in six categories: scientific societies and health professions associations; family associations and advocacy organizations; national and regional public institutions (e.g. public universities); private institutions (e.g. foundations, private health facilities, private universities); industry (e.g. pharmaceutical companies); public and private research institutes.

Guideline panel members reviewed the comments that were collected electronically using a structured questionnaire (https://piattaformasnlg.iss.it) over a four-week period. Example of questions used in the questionnaire are available in appendix 2. This early involvement aims at increasing transparency and stakeholder engagement. Similarly, we will invite stakeholders to review and provide comments on the draft recommendations once they will become available. Our dissemination also includes a website (www.osservatorionazionaleautismo.it) where recommendations and the underlying evidence will be available for different user profiles, similar those of the European Commission Breast Guidelines.(35)

2.11 Piloting of the development of recommendations

With the goal to allow the working group to gain experience with the process of making a recommendation and to familiarize with the dynamics typical of guideline panels, we

identified two pilot research questions. The ERT conducted systematic reviews and shared the following materials in advance of panel discussion: GRADE Evidence Profiles and a Summary of Findings tables (26, 27) summarizing the effects of the interventions, an EtD framework with structured summaries of the evidence to address each criterion, the list of included and excluded studies, and forest plots where applicable. We piloted the decision-making process using both the in-person and the online approach. In the former, panels met in a meeting room equipped with a u-shaped table, microphones and a recording system. A projector was used by the ERT to present the synthesis of the evidence on a large screen and by the chairs to facilitate discussion and navigate through the various criteria of the EtD. Simultaneously, we also streamed the meeting online using Webex (Cisco Webex, https://www.webex.com/), to allow off-site participation and visualization of content on the screen of panel members' devices while discussion it. To pilot the online approach, we used the PanelVoice module of GRADEpro (https://gradepro.org/panel-voice/). Through electronic surveys that are integrated in the EtDs, PanelVoice enables guideline developers to facilitate the decision-making process electronically. The process starts with the collection of panel judgments on the EtD criteria (figure 4).

Results of the PanelVoice are reported and agreement reached through email interaction or other necessary. The panel is then asked to decide and agree on the direction and strength of the recommendation and to formulate the statements to be reported in the EtD

conclusions section (e.g. justification, implementation considerations, research priorities etc.) (Figure 5).

3. Results

3.1 Composition of the guideline panels

Between June and July 2018, the steering committee received 158 applications for the two multidisciplinary and multi-professional panels of independent experts. Twenty-six applicants were not eligible because employed in private healthcare facilities or universities, had undocumented declared professional competences, their professional profile was not requested in the public selection announcement, or they applied after the submission deadline. Among the 138 who met the requirements, the SC selected 16 applicants per panel on the basis of their professional and personal experience, expertise, healthcare setting (primary, secondary and tertiary care), and geographical representation. Table 2 shows the compositions of the two panels.

3.2 Management of Conflict of Interest

The SC reviewed the detailed declarations of interest of all the 158 candidate panelists. None of them was prevented from participating in the panel because of relevant COI, since all the interests declared were considered as manageable through measures such as

partial exclusion or public disclosure. Afterwards, the SC evaluated the panelists' declared and non-declared interests, the latter identified through surveillance of research projects or training activities in which experts are engaged. The SC did not identify any relevant COI that would have prevented guideline panelists from participating in the generation and selection of the research questions addressed.

3.3 Guideline questions

We abstracted interventions and management strategies from previous guidelines into 7 macro-areas to create an initial list: diagnosis and assessment of ASD core-symptoms, diagnosis and assessment of ASD associated features, comorbidities, differential diagnosis, pharmacological interventions, psychosocial interventions, other non-pharmacological interventions. Strategies within the same macro-area were categorized and grouped together by the PRU where applicable. Categorization was based on the target population (e.g. people with ASD versus their caregivers) and on the theory underpinning the interventions. Due to the availability of multiple theoretical frameworks related to non-pharmacological interventions for ASD, the latter categorization presented challenges that we solved through discussion. As for the population of children and adolescents with ASD, the process resulted in a list of 79 questions of which 27 were high-priority, 46 questions important, and 6 questions not important. As for adults, we generated a list of 31 questions of which 21 were high-priority and 10 questions important. For each population, we will develop recommendations to answer 16 research

questions whereas all other questions will be mentioned as not prioritized in the guideline. Table 3 and table 4 lists the questions prioritized for inclusion in the guidelines. The lists of all generated questions and their priority ratings are available in appendix 3.

3.4 Outcomes

The panel responsible for children and adolescents with ASD rated ASD core-symptoms as critical outcomes for all research questions. Impairments in social interaction and communication, and restricted and repetitive behaviors were considered as distinct core-symptoms of ASD and rated separately. Other critical and important outcomes included quality of life, adaptive functioning skills, and parenting stress. The panel responsible for adults with ASD prioritized outcomes related to quality of life and outcomes such as social inclusion, level of independency from the caregivers, overall functioning and professional competencies. Other important outcomes included core-symptoms, behavioral disturbances, psychotic symptoms and treatments' side effects. All outcomes were PROs.

3.5 Stakeholders' consultation on the research questions

Of the 129 stakeholders that requested to comment on the list of questions identified for inclusion in the guideline, 115 met the eligibility criteria. We excluded stakeholders for the following reasons: the application process was not completed or the relationship with

healthcare industries was not declared. Figure 6 shows the distribution of registered stakeholders.

The majority of comments pertained to potentially relevant subgroups and outcomes that were not considered in the prioritized guideline questions. Many stakeholders requested clarification regarding the meaning of "standard of care" which was used to phrase some of the questions. Based on the feedback received, we reviewed the comments and improved the wording of research questions and added new sub-groups, where necessary.

4. Piloting of the development of recommendations

The questions identified for piloting the process focused on the impact of polyunsaturated fatty acids on PROs in children and adolescents with ASD. The body of evidence consisted of Randomized Controlled Trials (RCTs) and the overall certainty in the estimated effects was rated as very low owing to serious indirectness and very serious imprecision. Based on the very low certainty in the evidence of effects and uncertainty in other judgments on EtD criteria, the panel made conditional recommendations. Further details on the pilot questions, including the EtD framework with panel judgments, are available in appendix 4.

5. Discussion

We have described the methods and processes for guideline development at the ISS using the diagnosis and management of people with ASD as an example. It is the first guideline that follows the new ISS standards and has posed a number of methodological challenges that we addressed using novel guideline development approaches.(4) A challenge particularly relevant to ASD is its focus on PROs in people living in the ASD spectrum and their caregivers.

5.1 Challenges encountered during the development process

The heterogenous composition of the panels, which includes health professionals and stakeholders across a broad spectrum, reflects the complexity of the condition being addressed. The management of such large guideline groups, which encompass different professionals and potentially heterogeneous viewpoints, requires particular ability by chairpersons to conduct effective meetings. Given the broad interest in this guideline by many and diverse stakeholder groups, the process requires maximum possible transparency and we tackled this challenge through the use of GRADE EtDs, the early involvement of key stakeholders and press releases. Applying the ISS COI policy revealed the need for a cultural change. In fact, experts are often not aware that having published on the topic of interest or carried out research or professional activities in the field constitute an interest to be declared. This is important not only for disclosure purposes but also for allowing an assessment of potential conflicts and for determining

measures to manage them. We provided guidance to experts in this process to enable them to recognize and declare any circumstance in which a secondary interest could interfere with the impartial performance of their duties, functions and tasks.

5.2 Strengths and innovations of this guideline development process

We created a large multidisciplinary panel which include people with ASD and their caregivers and operate under a transparent policy on COI. Our process for prioritization, using a structured and transparent approach, granted equal voices to panel members and focused on PROs. Our process is supported by independent systematic reviews by the ERT which include an assessment of the certainty in the evidence according to the GRADE approach. We used health outcome descriptors to minimize the bias and improve the overall transparency of the process. Using the GRADE EtD framework, criteria and judgments that yield recommendations are transparent and allow targeting to different user profiles.(35) Through training and piloting exercises, we allowed the panel acquired familiarity with the GRADE approach, the use of the EtD framework, and the summaries of evidence provided to make informed judgments and reach recommendations.

We used information technology to streamline the development process and improve efficiency. Indeed, web-based decision-making and communication tools such as of GRADEpro, StarLeaf, and Webex facilitated work logistics and decreased costs associated with in-person meetings while increasing panel members' involvement. We promoted stakeholders' involvement from early stages of the process. The ISS SNLG web

platform ensured a transparent and participative process in which stakeholders are empowered to provide valuable feedback in several phases of the process.

5.3 Limitations of this guideline development process

Guideline development requires advanced methodological skills and understanding of evidence. Although panel members are not formally required to know the details of methodology, they must get acquainted with the relevant principles in order to understand the process flow; an ability which demands appropriate training. Human and time resources to develop the syntheses of the evidence that are used to inform the guideline are a very relevant component of the development process, but these resources are small compared to the cost of treatment and primary research in this area.

6. Conclusions

We have described the new Italian national guideline development process during its first application in recommendations about the diagnosis and management of ASD. The process seems feasible and acceptable to key stakeholders, including guideline panel members, those synthesizing the evidence and the public. The guideline working group is now developing recommendations that will be disseminated and adopted in Italy. This guideline aims to serve as a reference standard for future guideline development in the

Italian setting, and it will allow the adoption or adaptation to various settings, including international jurisdictions.

List of abbreviations

ASD: Autism Spectrum Disorder; **PROs:** Participant Reported Outcomes; **GRADE**: Grading of Recommendations Assessment, Development and Evaluation; **CPGs:** Clinical Practice Guidelines; **ISS:** Istituto Superiore di Sanità (Italian National Institute of Health); **CNEC:** Centro Nazionale per l'Eccellenza Clinica, la Qualità e la Sicurezza delle Cure (Centre for Clinical Excellence, Quality and Safety of Care); **SNLG:** Sistema Nazionale Linee Guida (Italian National Guidelines System); **COI:** Conflict of Interest; **GIN:** Guidelines International Network; **GRADEpro:** GRADEpro Guideline Development Tool; **ERT:** Evidence Review Team**; EtD:** Evidence-to-Decision; **PICO:** population, intervention, comparator, outcomes); **PRU:** PICO Responsible Unit; RCT: Randomized Controlled Trial;

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Competing interests

Members of the GDG do not receive financial compensation for their work on the panel but receive payment or reimbursement by the ISS for travel related expenses for the guideline meetings. GPM and LA are members and HJS is co-chair of the GRADE working group.

Funding

Italian Ministry of Health Project 'I disturbi dello spettro autistico: attività previste dal decreto ministeriale del 30.12.2016 – capitolo 4395 (*articolo 1, comma 401, legge 28 dicembre 2015, n. 208, recante "Disposizioni per la formazione del bilancio annuale e pluriennale dello Stato (legge di stabilità 2016*)".

Authors' contributions

GPM and HJS outlined the manuscript content. GPM drafted the initial manuscript. HJS and MLS helped developing subsequent drafts, contributing equally to its production. FN, CB, AF and FF critically reviewed it and provided important input. LA, FDC, GLD produced the synthesys of the evidence that were used to pilot the process of making a

recommendation. DP, GO, LF, PI, DC, AN, and MC reviewed the manuscript and provided important intellectual contributions leading to the finalization of the manuscript. The members of the guideline working group contributed to various parts of the work that allowed the development of the guideline. All authors read and approved the final manuscript.

Acknowledgements

Collaborating authors of the Italian National Institute of Health guideline working group on Autism Spectrum Disorder:

Members of the panel on children and adolescents with ASD:
Maurizio Arduino, Cristina Bellosio, Sandra Biasci, Serafino Buono, Corrado Cappa, Concetta Cordò, Emanuela Di Tommaso, Clelia Maria Duff, Claudia Felici, Angelo Massagli, Massimo Molteni, Franco Nardocci, Laura Reali, Holger J. Schünemann, Raffaella Tancredi, Giovanni Valeri, Lorella Venturini, Alessandro Zuddas.

Members of the panel on adults with ASD:

Mauro Andreoli, Corrado Barbui, Consuelo Bergamin, Marco Bertelli, Danilo Catania, Roberto Cavagnola, Pietro Cirrincione, Serafino Corti, Marusca Crognale, Raffaella Faggioli, Anna Maria Giogoli, Serenella Grittani, Roberto Keller, Pasqualina Pace, Pierluigi Politi, Holger J. Schünemann, Fabrizio Starace, Marco Valenti.

Members of the Evidence Review Team:

Laura Amato, Sara Balduzzi, Michele Basile, Michela Cinquini, Fabio Cruciani, Gian Loreto D'Alò, Roberto D'Amico, Marina Davoli, Franco De Crescenzo, Marien Gonzalez Lorenzo, Silvia Minozzi, Zuzana Mitrova, Ivan Moschetti, Vanna Pistotti, Rosella Saulle, Simona Vecchi.

The authors would like to thank Wojtek Wiercioch (McMaster University, Canada) for his critical review of this manuscript, and Giulia Galati (Italian National Institute of Health) for her administrative assistance.

References

1. World Health Organization. WHO handbook for guideline development 2014.

2. Miconi V. La responsabilità professionale dei sanitari: come cambia dopo la legge Gelli-Bianco. La Rivista Italiana della Medicina di Laboratorio - Italian Journal of Laboratory Medicine. 2018;14(3):184-90.

3. Iannone P, Coclite D, Fauci AJ, Graziano G, Napoletano AM. [Italian guidelines in accordance with the new National Guidelines System: critical issues and perspectives.]. Recenti progressi in medicina. 2017;108(9):360-2.

4. Iannone P, Coclite D, Napoletano A, Fauci A. Manuale metodologico per la produzione di linee guida di pratica clinica: Centro Nazionale per l'Eccellenza Clinica, la Qualità e la Sicurezza delle Cure dell'Istituto Superiore di Sanità. 2018 [Available from: https://snlg.iss.it/wp-content/uploads/2019/02/MM_v1.3_feb_2019.pdf.

5. Qaseem A, Forland F, Macbeth F, Ollenschlager G, Phillips S, van der Wees P. Guidelines International Network: toward international standards for clinical practice guidelines. Annals of internal medicine. 2012;156(7):525-31.

6. Schunemann HJ, Wiercioch W, Etxeandia I, Falavigna M, Santesso N, Mustafa R, et al. Guidelines 2.0: systematic development of a comprehensive checklist for a successful

guideline enterprise. CMAJ: Canadian Medical Association journal = journal de l'Association medicale canadienne. 2014;186(3):E123-42.

7. Arlington VA. Diagnostic and statistical manual of mental disorders American Psychiatric Association. 2013.

8. Baxter AJ, Brugha TS, Erskine HE, Scheurer RW, Vos T, Scott JG. The epidemiology and global burden of autism spectrum disorders. Psychol Med. 2015;45(3):601-13. doi: 10.1017/S003329171400172X. Epub 2014 Aug 11.

9. Brugha TS, Spiers N, Bankart J, Cooper SA, McManus S, Scott FJ, et al. Epidemiology of autism in adults across age groups and ability levels. Br J Psychiatry. 2016;209(6):498-503. doi: 10.1192/bjp.bp.115.174649. Epub 2016 Jul 7.

10. Sanità ISd. Giornata mondiale della Consapevolezza dell'Autismo: in Italia un bimbo ogni 77 2019 [Available from: https://www.iss.it/?p=3421.

11. Hrdlicka M, Dudova I. Controversies in autism: is a broader model of social disorders needed? Child Adolesc Psychiatry Ment Health. 2013;7(1):9. doi: 10.1186/753-2000-7-9.

12. Masi A, DeMayo MM, Glozier N, Guastella AJ. An Overview of Autism Spectrum Disorder, Heterogeneity and Treatment Options. Neurosci Bull. 2017;33(2):183-93. doi: 10.1007/s12264-017-0100-y. Epub 2017 Feb 17.

13. Sistema Nazionale Linee Guida - ISS. The treatment of children and adolescents with ASD 2011 [Available from: http://old.iss.it/auti/index.php?lang=1&id=437&tipo=5.

14. Zappella M, Bertelli MO. Guidelines of the Italian National Institute of Health on "The treatment of autism spectrum disorders in children and adolescents": limits and implications. Journal of Psychopathology. 2012(18):322-7.

15. Atkins D, Eccles M, Flottorp S, Guyatt GH, Henry D, Hill S, et al. Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches The GRADE Working Group. BMC Health Serv Res. 2004;4(1):38. doi: 10.1186/472-6963-4-38.

16. 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. doi: 10.1136/bmj.39489.470347.AD.

17. In: Graham R, Mancher M, Miller Wolman D, Greenfield S, Steinberg E, editors. Clinical Practice Guidelines We Can Trust. Washington DC: 2011 by the National Academy of Sciences; 2011.

18. Schunemann HJ, Al-Ansary LA, Forland F, Kersten S, Komulainen J, Kopp IB, et al. Guidelines International Network: Principles for Disclosure of Interests and Management of Conflicts in Guidelines. Annals of internal medicine. 2015;163(7):548-53.

19. Piggot T, Schunemann HJ. Guideline Participant Orientation Tool (GPOT) (v6) 2016 [Available from: https://heigrade.mcmaster.ca/guideline-development/guideline-participants.

20. Istituto Superiore di Sanità. ISS Guideline for the diagnosis and treatment of ASD in children and adolescents - Public call for panel members [Available from: https://snlg.iss.it/wp-content/uploads/2018/05/Selezione-pubblica-LG-ASD.pdf.

21. Schunemann HJ, Al-Ansary LA, Forland F, Kersten S, Komulainen J, Kopp IB, et al. Guidelines International Network: Principles for Disclosure of Interests and Management of Conflicts in Guidelines. Ann Intern Med. 2015;163(7):548-53.

22. Andrews JC, Schunemann 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. Journal of clinical epidemiology. 2013;66(7):726-35.

23. Balshem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. Journal of clinical epidemiology. 2011;64(4):401-6.

24. Zhang Y, Coello PA, Guyatt GH, Yepes-Nunez JJ, Akl EA, Hazlewood G, et al. GRADE guidelines: 20. Assessing the certainty of evidence in the importance of outcomes

or values and preferences-inconsistency, imprecision, and other domains. Journal of clinical epidemiology. 2019;111:83-93.

25. Alonso-Coello P, Schunemann HJ, Moberg J, Brignardello-Petersen R, Akl EA, Davoli M, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction. BMJ. 2016;353:i2016.

26. Guyatt GH, Oxman AD, Santesso N, Helfand M, Vist G, Kunz R, et al. GRADE guidelines: 12. Preparing summary of findings tables-binary outcomes. J Clin Epidemiol. 2013;66(2):158-72. doi: 10.1016/j.jclinepi.2012.01.012. Epub May 18.

27. Guyatt GH, Thorlund K, Oxman AD, Walter SD, Patrick D, Furukawa TA, et al. GRADE guidelines: 13. Preparing summary of findings tables and evidence profiles-continuous outcomes. J Clin Epidemiol. 2013;66(2):173-83. doi: 10.1016/j.jclinepi.2012.08.001. Epub Oct 30.

28. Schunemann HJ, Cushman M, Burnett AE, Kahn SR, Beyer-Westendorf J, Spencer FA, et al. American Society of Hematology 2018 guidelines for management of venous thromboembolism: prophylaxis for hospitalized and nonhospitalized medical patients. Blood Adv. 2018;2(22):3198-225. doi: 10.1182/bloodadvances.2018022954.

29. (SIGN) SIGN. Assessment, diagnosis and interventions for autism spectrum disorders. Edinburgh: SIGN; 2016. SIGN publication. 2016;145.

30. National Institute for Health and Clinical Excellence. NICE Clinical Guidelines [CG 170] - Autism spectrum disorder in under 19s: support and management 2013 [Available from: https://www.nice.org.uk/guidance/cg170.

31. National Institute for Health and Clinical Excellence. NICE Clinical Guidelines [CG142] - Autism spectrum disorder in adults: diagnosis and management 2016 [Available from: https://www.nice.org.uk/guidance/cg142.

32. National Institute for Health and Clinical Excellence. NICE Clinical Guidelines [CG 128] - Autism spectrum disorder in under 19s: recognition, referral and diagnosis

2017 [Available from: https://www.nice.org.uk/guidance/cg128.

33. O'Sullivan D, Wilk S, Michalowski W, Farion K. Using PICO to align medical evidence with MDs decision making models. Studies in health technology and informatics. 2013;192:1057.

34. Schünemann HJ. Getting trustworthy guideline recommendations into practice: thebmjopinion; 2018 [Available from: https://blogs.bmj.com/bmj/2018/12/31/holger-j-schunemann-getting-trustworthy-guideline-recommendations-into-practice/.

35. Schunemann HJ, Lerda D, Dimitrova N, Alonso-Coello P, Grawingholt A, Quinn C, et al. Methods for Development of the European Commission Initiative on Breast Cancer

Guidelines: Recommendations in the Era of Guideline Transparency. Annals of internal medicine. 2019;23(2738921):M18-3445.

36. Guyatt GH, Oxman AD, Kunz R, Atkins D, Brozek J, Vist G, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. Journal of clinical epidemiology. 2011;64(4):395-400.

37. Istituto Superiore Sanità. Avviso di consultazione pubblica sulle LG autismo 2018 [Available from: [https://snlg.iss.it/?p=1113].

Figures

Figure 1

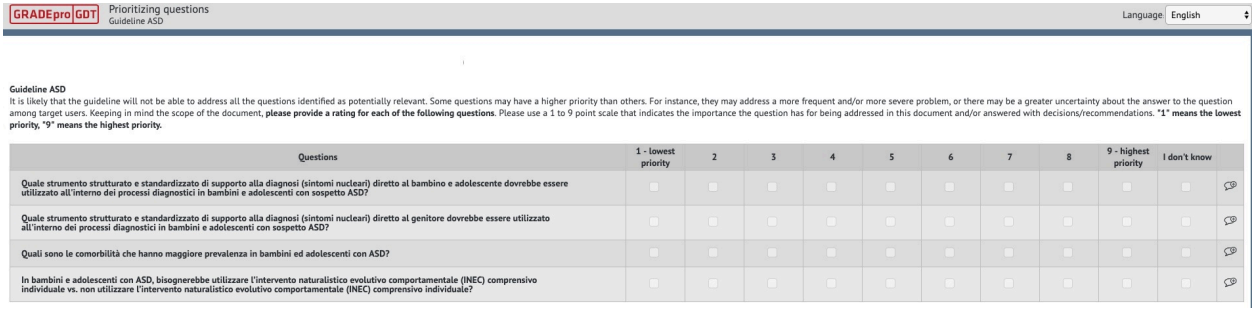


Figure 1: Rating question importance using GRADEpro. Panel members rate the importance of candidate guideline questions on a 1 to 9 scale. Lower ratings are indicative of lower importance.

Figure 2

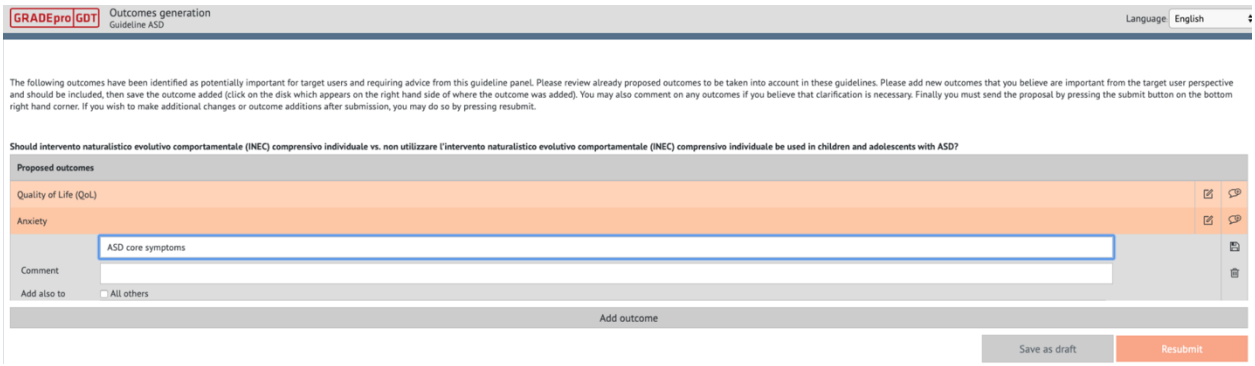


Figure 2: Generation of outcomes using GRADEpro. GRADEpro interface. Panel members suggest, separately for each question, any people-important outcomes that should be considered during the rating of the relative importance of outcomes.

Figure 3

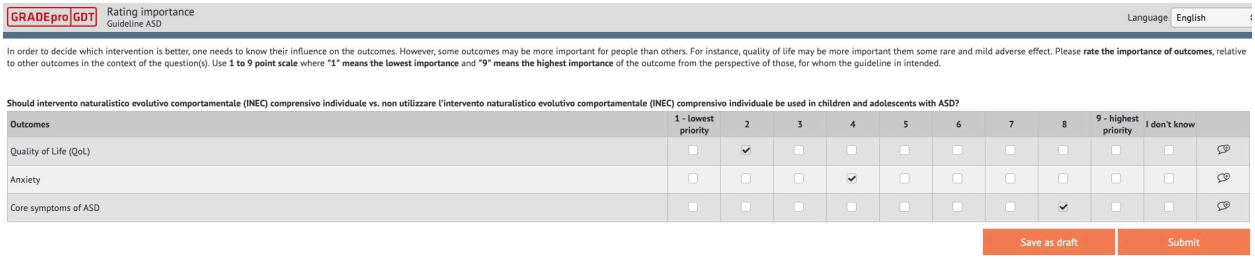


Figure 3 Rating relative importance of outcomes using GRADEpro. GRADEpro interface. Panel members rate the importance of people-important outcomes on a 1 to 9 scale. Lower ratings are indicative of lower importance.

Figure 4

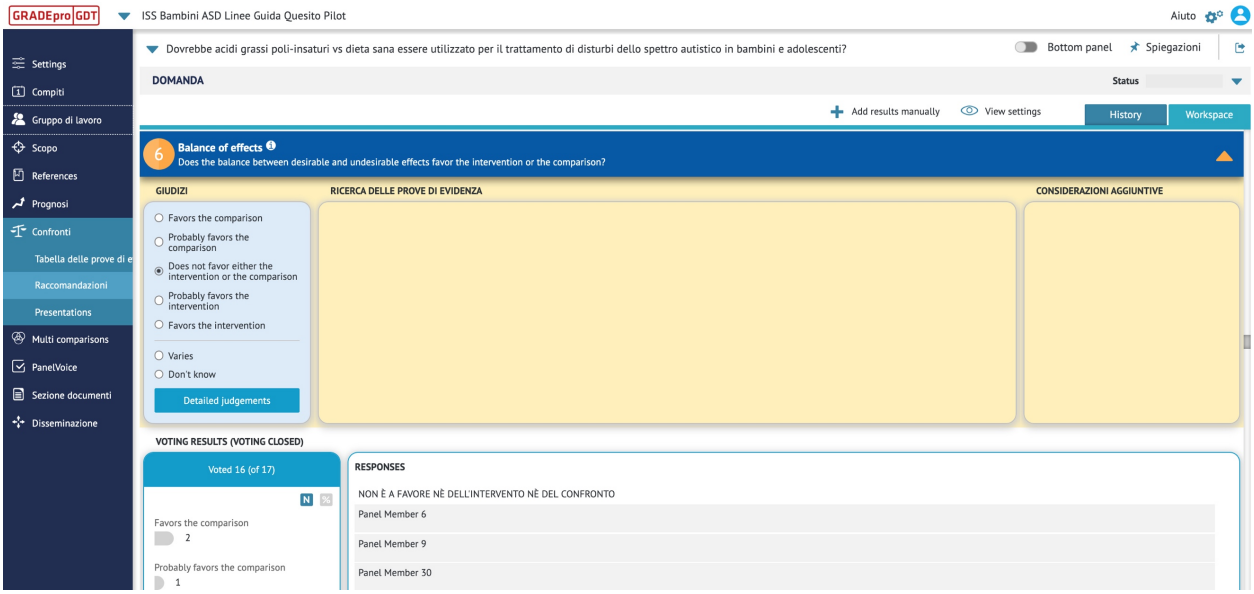


Figure 4: Collection of EtD judgments using PanelVoice. PanelVoice/GRADEpro interface. Judgments on the EtD criteria submitted by panel members are visible to guideline developers and can be used to facilitate the decision-making process online.

Figure 5

Figure 5: Collection of votes on the strength and direction of recommendations using PanelVoice. PanelVoice/GRADEpro interface. Voting results for the direction and strength of the recommendation are visible to guideline developers and can be used to reach online agreement about the final recommendation.

Figure 6

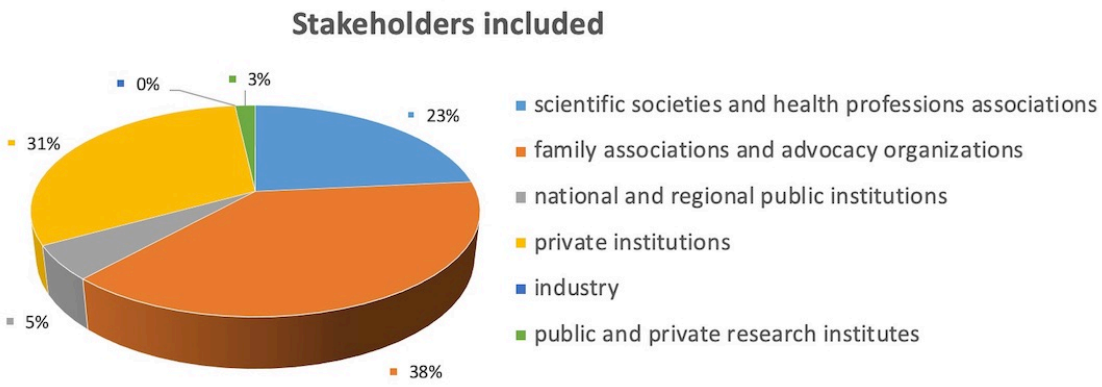


Figure 6: Distribution of registered stakeholders. Pie chart reporting affiliations of the stakeholders participating in the public consultation.

Tables

Table 1

Factors that influence if a question is important in the context of a guideline
Common question in practice?
Uncertainty in practice?
New evidence to consider?
Variation in practice?
Consequences for resource use/cost?
Not previously or sufficiently addressed?

Table 1: Factors that should be considered while deciding which the research questions to be included in a guideline

Table 2

	children and adolescents	**adults**
Expertise	**n.**	**n.**
Child Neurologist and Psychiatrist	4	1
Psychiatrist	1	4
Psychologist	2	2
Pshycopharmacologist	1	1
Childhood neuro and psychomotricity therapist	1	-
Speech therapist	1	-
Pedagogues	1	1

Social worker	-	1
Educational therapist	1	1
Occupational therapist	-	2
Expert in the management of healthcare systems	1	1
General practitioner	1	1
Pediatrician	1	-
Methodologist	1	1
Parent of child or adolescents with ASD	2	1
Person with ASD	-	1

Table 2: Composition of the guideline working groups

Table 3

	Question	**Macro-area**
1	Should structured diagnostic instruments (to the children?) be added to the clinical assessment from a multidisciplinary team to diagnose ASD core symptoms?	Diagnosis
2	Should structured diagnostic instruments (to the parents or caregivers?) be added to the clinical assessment from a multidisciplinary team to diagnose ASD core symptoms?	Diagnosis
3	Which are the most prevalent comorbidities in children and adolescents with ASD?	Diagnosis
4	Should INEC comprehensive individual vs no intervention or treatment as usual be used for children and adolescents with ASD?	Psychosocials interventions
5	Should ABA comprehensive vs no intervention or treatment as usual be used for children and adolescents with ASD?	Psychosocials interventions
6	Should Educational comprehensive individual vs no intervention or treatment as usual be used for children and adolescents with ASD?	Psychosocials interventions

7	Should interventions with parents/caregivers vs no intervention or treatment as usual be used for children and adolescents with ASD?	Psychosocials interventions
8	Should INEC focalized vs no intervention or treatment as usual be used for children and adolescents with ASD?	Psychosocials interventions
9	Should ABA focalized vs no intervention or treatment as usual be used for children and adolescents with ASD?	Psychosocials interventions
10	Should INEC focalized vs no intervention or treatment as usual be used for children and adolescents with ASD?	Psychosocials interventions
11	Should mood stabilizers vs no intervention be used in children and adolescents with ASD?	Pharmacological interventions
12	Should SSRIs and/or SNRIs vs no SSRIs and/or SNRIs be used in children and adolescents with ASD?	Pharmacological interventions
13	Should D2 blockers vs no D2 blockers be used in children and adolescents with ASD?	Pharmacological interventions
14	Should psychostimulants and/or atomoxetine vs no psychostimulants and/or atomoxetine be used in children and adolescents with ASD?	Pharmacological interventions
15	Should communicative interventions for social communication and interaction vs no intervention or treatment as usual be used in children and adolescents with ASD?	Other interventions
16	Should interventions for specific behaviours vs no intervention or treatment as usual be used in children and adolescents with ASD?	Other interventions

Table 3: List of questions included in the guideline on children and adolescents with ASD

Table 4

	Question	**Macro-area**
1	Should structured diagnostic instruments be added to routine clinical assessment to diagnose ASD in adults?	Diagnosis
2	Should structured diagnostic instruments to assess psichoeducative and adaptive profile be added to the clinical assessment of the adults with ASD?	Diagnosis
3	Should structured diagnostic instruments to assess neuropsychological and cognitive profile be added to the clinical assessment of the adults with ASD?	Diagnosis

4	Should tests or diagnostic examinations be used to identify psychiatric, neurologic and/or selected medical comorbidities in adults with ASD?	Diagnosis
5	Should standardized instruments to rate the quality of life be used in clinical routine for adults with ASD?	Psychosocials interventions
6	Should standardized preference procedures be used to plan the “life project” of adults with ASD?	Psychosocials interventions
7	Should community-based services and housing support be taken into consideration for adults with ASD?	Psychosocials interventions
8	Should psychoeducative programs be implemented in adults with ASD?	Psychosocials interventions
9	Should information/support campaigns for family members, caregivers and other public figures be accomplished in support of adults with ASD?	Psychosocials interventions
10	Should interventions in support of occupational activities be implemented in adults with ASD?	Psychosocials interventions
11	Should psychological interventions be implemented in adults with ASD?	Psychosocials interventions
12	Should antipsychotics vs no antipsychotics be used in adults with ASD?	Pharmacological interventions
13	Should mood stabilizers vs no mood stabilizers be used in adults with ASD?	Pharmacological interventions
14	Should antidepressants vs no antidepressants be used in adults with ASD?	Pharmacological interventions
15	Should stimulants vs no stimulants be used in adults with ASD?	Pharmacological interventions
16	Should “other drugs” vs no "other drugs" be used in adults with ASD?	Pharmacological interventions

Table 4: List of questions included in the guideline on adults with ASD

Appendices

Appendix 1

Interpretation of ratings on the 1 to 9 scale for prioritisation of questions	Interpretation of ratings on the 1 to 9 scale for prioritisation of outcomes
7 to 9 the question has high priority. It should be addressed in the guideline. **4 to 6** the question is considered a priority but not having high priority relative to other questions. It should be listed as a priority question but not addressed in the guideline. **1 to 3** the question is not a priority. It is acceptable to neither include nor mention it in the guideline.	**7 to 9** the outcome is *critical* for decision making **4 to 6** the outcome is *important* but not critical for decision making **1 to 3** the outcome is of *low importance*

Appendix 1: Interpretations of ratings for research questions and outcomes

Appendix 2

Example of questions included in the questionnaire for stakeholders
Are the population and its sub-populations clearly described? Is there any relevant sub-group of the population that was not listed?
Is the intervention clearly described? Is there any relevant intervention related to the type of interventions being addressed that was not listed?
Is the comparison clearly described? Is there any relevant comparison related to the type of interventions being addressed that was not listed?
Are the outcomes clearly described? Is there any relevant outcome that was not listed?

Appendix 2: Example of questions included in the questionnaire for stakeholder

Appendix 3

Questions on children and adolescents with ASD

Rank	Section	Question	Mean
1	TREATMENT	[Sezione 5 - Psicosociali/Genitori-Caregivers] Per bambini e adolescenti con ASD, bisognerebbe utilizzare parent training vs. intervento di sostegno psicoeducativo con i genitori/caregiver?	8.43
2	TREATMENT	[Sezione 5 - Psicosociali/Comprensivi/Bambino] In bambini e adolescenti con ASD, bisognerebbe utilizzare l'intervento INEC comprensivo individuale vs. nessun intervento o treatment as usual?	8.31
3	TREATMENT	[Sezione 5 - Altri Interventi] In bambini e adolescenti con ASD, bisognerebbe utilizzare interventi comunicativi per la comunicazione sociale e l'interazione (include social stories, interventi che utilizzano le nuove tecnologie, interventi mediati dai coetanei, training sulla teoria della mente) nessun intervento o treatment as usual? Se si quali?	8.21
4	DIAGNOSIS	[Sezione 1- Sintomi Core/Bambino] Per la diagnosi di ASD (sintomi core) in bambini e adolescenti è utile l'utilizzo di strumenti strutturati standardizzati di supporto alla diagnosi diretti al bambino, in aggiunta all'osservazione e al colloquio clinico (OCC), verso il solo OCC? Se si, quale?	8.20
5	TREATMENT	[Sezione 5 - Psicosociali/Comprensivi/Bambino] In bambini e adolescenti con ASD, bisognerebbe utilizzare l'intervento Evolutivo comprensivo individuale vs. nessun intervento o treatment as usual?	8.08
6	TREATMENT	[Sezione 5 - Psicosociali/Comprensivi/Bambino] In bambini e adolescenti con ASD, bisognerebbe utilizzare l'intervento ABA comprensivo individuale vs. nessun intervento o treatment as usual?	8.00
7	DIAGNOSIS	[Sezione 1- Sintomi Core/Genitori] Per la diagnosi di disturbo dello spettro autistico (ASD) (sintomi core) in bambini e adolescenti è utile l'utilizzo di strumenti strutturati standardizzati di supporto alla diagnosi diretti ai genitori, in aggiunta all'osservazione e al colloquio clinico (OCC), verso il solo OCC? Se si, quale?	7.93
8	TREATMENT	[Sezione 5 - Psicosociali/Genitori-Caregivers] Per bambini e adolescenti con ASD, bisognerebbe utilizzare trattamenti con i genitori/caregiver vs. nessun intervento o treatment as usual? Se si, quale?	7.93
9	TREATMENT	In bambini e adolescenti con ASD, bisognerebbe utilizzare gli stabilizzanti dell'umore vs. placebo o nessun intervento?	7.92
10	TREATMENT	[Sezione 5 - Altri Interventi] In bambini e adolescenti con ASD, bisognerebbe utilizzare interventi per comportamenti specifici (include programmi sulle abilità sociali, social skill group (nice), lego therapy, sulp, junior detective training program) nessun intervento o treatment as usual? Se si quali?	7.85
11	TREATMENT	[Sezione 5 - Psicosociali/Focalizzati/Bambino] In bambini e adolescenti con ASD, bisognerebbe utilizzare l'intervento INEC focalizzato individuale vs. nessun intervento o treatment as usual?	7.83
12	DIAGNOSIS	[Sezione 3 - Comorbilità] Quali sono le comorbilità che hanno maggiore prevalenza in bambini ed adolescenti con ASD e che dovrebbero essere prese in considerazione durante il processo di valutazione?	7.60
13	TREATMENT	In bambini e adolescenti con ASD, bisognerebbe utilizzare gli inibitori del reuptake della serotonina (SSRI) vs. placebo o nessun intervento?	7.54

14	TREATMENT	In bambini e adolescenti con ASD, bisognerebbe utilizzare gli antipsicotici vs. placebo o nessun intervento?	7.54
15	TREATMENT	[Sezione 5 - Psicosociali/Focalizzati/Bambino] In bambini e adolescenti con ASD, bisognerebbe utilizzare l'intervento ABA focalizzato individuale vs. nessun intervento o treatment as usual?	7.50
16	TREATMENT	[Sezione 5 - Psicosociali/Comprensivi/Bambino] In bambini e adolescenti con ASD, bisognerebbe utilizzare l'intervento Educativo comprensivo individuale vs. nessun intervento o treatment as usual?	7.46
17	TREATMENT	In bambini e adolescenti con ASD, bisognerebbe utilizzare gli inibitori del reuptake della norepinefrina vs. placebo o nessun intervento?	7.38
18	TREATMENT	[Sezione 5 - Psicosociali/Focalizzati/Bambino] In bambini e adolescenti con ASD, bisognerebbe utilizzare l'intervento Evolutivo focalizzato individuale vs. nessun intervento o treatment as usual?	7.36
19	TREATMENT	[Sezione 5 - Altri Interventi] In bambini e adolescenti con ASD, bisognerebbe utilizzare la Cognitive Behavioural Therapy (CBT) vs. nessun intervento o treatment as usual?	7.36
20	DIAGNOSIS	[Sezione 2.1 - Sintomi Non-Core/Cognitivo/Bambino] Per la diagnosi di ASD (sintomi non core_dominio cognitivo) in bambini e adolescenti è utile l'utilizzo di strumenti strutturati standardizzati di supporto alla diagnosi diretti al bambino, in aggiunta all'osservazione e al colloquio clinico (OCC), verso il solo OCC? Se si, quale?	7.33
21	DIAGNOSIS	[Sezione 4 - Diagnosi Differenziale] In bambini e adolescenti che vengono riferiti per sospetto ASD, quali patologie, oltre l'ASD, vengono maggiormente diagnosticate?	7.27
22	DIAGNOSIS	[Sezione 2.2 - Valutazione Globale/Neuropsicologico/Bambino] Per la valutazione globale (dominio neuropsicologico) in bambini e adolescenti con ASD è utile l'utilizzo di strumenti strutturati standardizzati di supporto alla diagnosi diretti al bambino, in aggiunta all'osservazione e al colloquio clinico (OCC), verso il solo OCC? Se si, quale?	7.21
23	DIAGNOSIS	[Sezione 1- Sintomi Core] Per la diagnosi di ASD (sintomi core) in bambini e adolescenti, l'utilizzo combinato di ADOS e ADI-R offre maggiore accuratezza se confrontato all'uso singolo dei due strumenti?	7.20
24	TREATMENT	In bambini e adolescenti con ASD, bisognerebbe utilizzare gli inibitori del reuptake della serotonina e noradrenalina (SNRI) vs. placebo o nessun intervento?	7.18
25	DIAGNOSIS	[Sezione 2.1 - Sintomi Non-Core/Adattivo/Bambino] Per la diagnosi di ASD (sintomi non core_dominio adattivo/diretto al bambino) in bambini e adolescenti è utile l'utilizzo di di Vineland Adaptive Behavior Scales (VABS), in aggiunta all'osservazione e al colloquio clinico (OCC), verso il solo OCC?	7.14
26	TREATMENT	In bambini e adolescenti con ASD, bisognerebbe utilizzare gli psicostimolanti vs. placebo o nessun intervento?	7.08
27	TREATMENT	[Sezione 6 - Terapie del sonno] In bambini e adolescenti con ASD, bisognerebbe utilizzare trattamenti per la gestione del sonno (include CBT, melatonina e invio ad uno specialista del sonno) vs. placebo o nessun intervento?	7.07
28	DIAGNOSIS	[Sezione 2.2 - Valutazione Globale/Comorbilità/Genitori] Per la valutazione globale (dominio comorbilità/diretto ai genitori) in bambini e adolescenti con ASD è utile l'utilizzo della Child Behaviour Checklist (CBCL), in aggiunta all'osservazione e al colloquio clinico (OCC), verso il solo OCC?	6.87

29	DIAGNOSIS	[Sezione 2.1 - Sintomi Non-Core/Adattivo/Genitori] Per la diagnosi di ASD (sintomi non core_dominio adattivo) in bambini e adolescenti è utile l'utilizzo di strumenti strutturati di supporto standardizzati alla diagnosi diretti ai genitori, in aggiunta all'osservazione e al colloquio clinico (OCC), verso il solo OCC? Se si, quale?	6.86
30	TREATMENT	[Sezione 5 - Psicosociali/Comprensivi/Bambino] In bambini e adolescenti con ASD, bisognerebbe utilizzare l'intervento INEC comprensivo individuale vs. Evolutivo comprensivo individuale?	6.85
31	TREATMENT	[Sezione 5 - Psicosociali/Focalizzati/Bambino] In bambini e adolescenti con ASD, bisognerebbe utilizzare l'intervento INEC focalizzato individuale vs. Evolutivo focalizzato individuale?	6.77
32	DIAGNOSIS	[Sezione 2.1 - Sintomi Non-Core/Linguaggio/Genitori] Per la diagnosi di ASD (sintomi non core_dominio linguaggio/diretto ai genitori) in bambini e adolescenti è utile l'utilizzo delle MacArthur-Bates Communicative Development Inventories (MB-CDIs), in aggiunta all'osservazione e al colloquio clinico (OCC), verso il solo OCC?	6.73
33	DIAGNOSIS	[Sezione 2.1 - Sintomi Non-Core/Linguaggio/Bambino] Per la diagnosi di ASD (sintomi non core_dominio linguaggio) in bambini e adolescenti è utile l'utilizzo di strumenti strutturati standardizzati di supporto alla diagnosi diretti al bambino, in aggiunta all'osservazione e al colloquio clinico (OCC), verso il solo OCC? Se si, quale?	6.64
34	DIAGNOSIS	[Sezione 2.1 - Sintomi Non-Core/Adattivo/Educatori] Per la diagnosi di ASD (sintomi non core_dominio adattivo/diretto agli educatori) in bambini e adolescenti è utile l'utilizzo di Adaptive Behavior Assessment System (ABAS), in aggiunta all'osservazione e al colloquio clinico (OCC), verso il solo OCC?	6.64
35	TREATMENT	[Sezione 5 - Psicosociali/Comprensivi/Bambino] In bambini e adolescenti con ASD, bisognerebbe utilizzare l'intervento ABA comprensivo individuale vs. Evolutivo comprensivo individuale?	6.62
36	DIAGNOSIS	[Sezione 2.2 - Valutazione Globale/Contesto di vita/Educatori] Per la valutazione globale (contesto di vita/diretto agli educatori) in bambini e adolescenti con ASD è utile l'utilizzo di Vineland Adaptive Behavior Scales (VABS), in aggiunta all'osservazione e al colloquio clinico (OCC), verso il solo OCC?	6.60
37	TREATMENT	[Sezione 5 - Psicosociali/Comprensivi/Bambino] In bambini e adolescenti con ASD, bisognerebbe utilizzare l'intervento ABA comprensivo individuale vs. INEC comprensivo individuale?	6.54
38	TREATMENT	[Sezione 5 - Psicosociali/Comprensivi/Bambino] In bambini e adolescenti con ASD, bisognerebbe utilizzare l'intervento INEC comprensivo individuale vs. Educativo comprensivo individuale?	6.54
39	TREATMENT	[Sezione 5 - Altri Interventi] In bambini e adolescenti con ASD, bisognerebbe utilizzare interventi di integrazione sensoriale vs. nessun intervento o treatment as usual?	6.29
40	TREATMENT	[Sezione 5 - Psicosociali/Comprensivi/Bambino] In bambini e adolescenti con ASD, bisognerebbe utilizzare l'intervento ABA comprensivo individuale vs. Educativo comprensivo individuale?	6.23
41	TREATMENT	[Sezione 5 - Altri Interventi] In bambini e adolescenti con ASD, bisognerebbe utilizzare la terapia occupazionale vs. nessun intervento o treatment as usual?	6.21

42	TREATMENT	[Sezione 5 - Psicosociali/Focalizzati/Bambino] In bambini e adolescenti con ASD, bisognerebbe utilizzare l'intervento ABA focalizzato individuale vs. Evolutivo focalizzato individuale?	6.14
43	DIAGNOSIS	[Sezione 2.2 - Valutazione Globale/Neuropsicologico/Genitori] Per la valutazione globale (dominio neuropsicologico/diretto ai genitori) in bambini e adolescenti con ASD è utile l'utilizzo di Behavior Rating Inventory of Executive Function (BRIEF), in aggiunta all'osservazione e al colloquio clinico (OCC), verso il solo OCC?	6.08
44	DIAGNOSIS	[Sezione 2.2 - Valutazione Globale/Contesto di vita/Genitori] Per la valutazione globale (contesto di vita) in bambini e adolescenti con ASD è utile l'utilizzo di strumenti strutturati standardizzati di supporto alla diagnosi diretti ai genitori, in aggiunta all'osservazione e al colloquio clinico (OCC), verso il solo OCC? Se si, quale?	6.07
45	DIAGNOSIS	[Sezione 1- Sintomi Core] Per la diagnosi di ASD (sintomi core) in bambini e adolescenti, l'utilizzo di strumenti strutturati standardizzati di supporto alla DIAGNOSIS diretti ai genitori (include strumenti diretti ai genitori diversi da ADI-R) offre maggiore accuratezza dell'uso di ADI-R?	5.93
46	DIAGNOSI	[Sezione 2.2 - Valutazione Globale/Comorbilità/Educatori] Per la valutazione globale (dominio comorbidità/diretto agli educatori) in bambini e adolescenti con ASD è utile l'utilizzo del Teacher Report Form (TRF) in aggiunta all'osservazione e al colloquio clinico (OCC), verso il solo OCC?	5.93
47	TREATMENT	[Sezione 5 - Psicosociali/Comprensivi/Bambino] In bambini e adolescenti con ASD, bisognerebbe utilizzare l'intervento Evolutivo comprensivo individuale vs. Educativo comprensivo individuale?	5.92
48	TREATMENT	[Sezione 5 - Psicosociali/Focalizzati/Bambino] In bambini e adolescenti con ASD, bisognerebbe utilizzare l'intervento ABA focalizzato individuale vs. INEC focalizzato individuale?	5.92
49	DIAGNOSIS	[Sezione 2.2 - Valutazione Globale/Comorbilità/Bambino] Per la valutazione globale (dominio comorbidità/diretto al bambino) in bambini e adolescenti con ASD è utile l'utilizzo dello Youth Self Report (YSR) in aggiunta all'osservazione e al colloquio clinico (OCC), verso il solo OCC?	5.85
50	TREATMENT	In bambini e adolescenti con ASD, bisognerebbe utilizzare gli ormoni peptidici (ossitocina, secretina) vs. placebo o nessun intervento?	5.85
51	DIAGNOSIS	[Sezione 1- Sintomi Core/Educatori] Per la diagnosi di ASD (sintomi core) in bambini e adolescenti è utile l'utilizzo di strumenti strutturati standardizzati di supporto alla diagnosi diretti agli educatori, in aggiunta all'osservazione e al colloquio clinico (OCC), verso il solo OCC? Se si, quale?	5.73
52	TREATMENT	[Sezione 6 - Interventi Nutrizionali] In bambini e adolescenti con ASD, bisognerebbe utilizzare acidi grassi poli-insaturi vs. placebo o nessun intervento?	5.64
53	TREATMENT	[Sezione 5 - Altri Interventi] In bambini e adolescenti con ASD, bisognerebbe utilizzare l' Emotion Recognition Training (ERT) vs. nessun intervento o treatment as usual?	5.62
54	TREATMENT	[Sezione 5 - Altri Interventi] In bambini e adolescenti con ASD, bisognerebbe utilizzare il Face Recognition Training (FRT) vs.nessun intervento o treatment as usual?	5.62
55	TREATMENT	In bambini e adolescenti con ASD, bisognerebbe utilizzare gli α agonisti vs. placebo o nessun intervento?	5.55
56	TREATMENT	In bambini e adolescenti con ASD, bisognerebbe utilizzare gli agenti gabaergici vs. placebo o nessun intervento?	5.54

57	TREATMENT	[Sezione 5 - Altri Interventi] In bambini e adolescenti con ASD, bisognerebbe utilizzare la musicoterapia vs. nessun intervento o treatment as usual?	5.43
58	DIAGNOSIS	[Sezione 2.2 - Valutazione Globale/Neuropsicologico/Educatori] Per la valutazione globale (dominio neuropsicologico/diretto agli educatori) in bambini e adolescenti con ASD è utile l'utilizzo di Behavior Rating Inventory of Executive Function (BRIEF), in aggiunta all'osservazione e al colloquio clinico (OCC), verso il solo OCC?	5.38
59	TREATMENT	[Sezione 6 - Interventi Nutrizionali] In bambini e adolescenti con ASD, bisognerebbe utilizzare dieta senza glutine e/o prodotti caseari vs. placebo o nessun intervento?	5.33
60	TREATMENT	[Sezione 5 - Altri Interventi] In bambini e adolescenti con ASD, bisognerebbe utilizzare attività fisica vs. treatment as usual?	5.23
61	TREATMENT	In bambini e adolescenti con ASD, bisognerebbe utilizzare gli antidepressivi triciclici vs. placebo o nessun intervento?	5.23
62	TREATMENT	[Sezione 6 - Interventi Nutrizionali] In bambini e adolescenti con ASD, bisognerebbe utilizzare supplementazione di vitamine e minerali (fatta eccezione per la vitamina k, il ferro e rame) vs. placebo o nessun intervento ?	5.15
63	TREATMENT	In bambini e adolescenti con ASD, bisognerebbe utilizzare agenti glutammatergici vs. placebo o nessun intervento?	5.08
64	TREATMENT	[Sezione 6 - Interventi Nutrizionali] In bambini e adolescenti con ASD, bisognerebbe utilizzare L-Carnosine/L-Carnitine vs. placebo o nessun intervento?	5.00
65	TREATMENT	In bambini e adolescenti con ASD, bisognerebbe utilizzare gli ormoni steroidei vs. placebo o nessun intervento?	5.00
66	TREATMENT	In bambini e adolescenti con ASD, bisognerebbe utilizzare CX516 vs. placebo o nessun intervento?	5.00
67	TREATMENT	In bambini e adolescenti con ASD, bisognerebbe utilizzare gli antistaminici vs. placebo o nessun intervento?	4.83
68	TREATMENT	[Sezione 5 - Altri Interventi] In bambini e adolescenti con ASD, bisognerebbe utilizzare la terapia assistita con gli animali vs. nessun intervento o treatment as usual? Se si quale?	4.64
69	TREATMENT	In bambini e adolescenti con ASD, bisognerebbe utilizzare gli antiossidanti vs. placebo o nessun intervento?	4.58
70	TREATMENT	In bambini e adolescenti con ASD, bisognerebbe utilizzare la minociclina vs. placebo o nessun intervento?	4.58
71	TREATMENT	[Sezione 5 - Altri Interventi] In bambini e adolescenti con ASD, bisognerebbe utilizzare la terapia psicodinamica vs. nessun intervento o treatment as usual?	4.50
72	TREATMENT	[Sezione 6 - Terapie Complementari] In bambini e adolescenti con ASD, bisognerebbe utilizzare neurofeedback vs. nessun intervento o treatment as usual?	4.46
73	TREATMENT	In bambini e adolescenti con ASD, bisognerebbe utilizzare la fenfluramina vs. placebo o nessun intervento?	4.31
74	TREATMENT	[Sezione 5 - Altri Interventi] In bambini e adolescenti con ASD, bisognerebbe utilizzare la comunicazione facilitata vs. nessun intervento o treatment as usual?	3.93
75	TREATMENT	[Sezione 6 - Terapie Complementari] In bambini e adolescenti con ASD, bisognerebbe utilizzare chelazione lungo termine, chelazione breve termine vs. placebo o nessun intervento?	3.73

76	TREATMENT	[Sezione 6 - Terapie Complementari] In bambini e adolescenti con ASD, bisognerebbe utilizzare agopuntura, elettro-agopuntura, agopressione vs. placebo o nessun intervento?	3.56
77	TREATMENT	[Sezione 6 - Terapie Complementari] In bambini e adolescenti con ASD, bisognerebbe utilizzare auditory integration training vs. nessun intervento o treatment as usual?	3.54
78	TREATMENT	[Sezione 6 - Terapie Complementari] In bambini e adolescenti con ASD, bisognerebbe utilizzare kata exercise training vs. nessun intervento o treatment as usual?	3.50
79	TREATMENT	[Sezione 6 - Terapie Complementari] In bambini e adolescenti con ASD, bisognerebbe utilizzare Qigong massage training vs. nessun intervento o treatment as usual?	3.08

Questions on adults with ASD

ranking	Section	Domanda	Mean
1	TREATMENT	Negli adulti con diagnosi di ASD la pianificazione e la valutazione degli interventi basata su quality of life (e altre person-centred outcome measures) vs la pianificazione e la valutazione degli interventi basata su misure di esito tradizionali. non person-centred (es. sintomi. integrità morfologica e di funzionamento. etc) migliora (OUTCOME)?	8.18
2	TREATMENT	Negli adulti con ASD gli interventi psicoeducativi sono efficaci per (OUTCOME). verso l'assenza di interventi psicoeducativi? Se si. quali?	8.18
3	TREATMENT	Negli adulti con ASD i servizi con équipe multidisciplinari specializzate nell'ASD. verso servizi senza équipe multidisciplinare specializzata nell'ASD. sono efficaci per (OUTCOME)?	8.06
4	DIAGNOSIS	Per la diagnosi di ASD negli adulti è utile l'utilizzo di strumenti strutturati di supporto alla diagnosi. in aggiunta all'osservazione e al colloquio clinico (OCC). verso il solo OCC? Se si. quale?	7.94
5	DIAGNOSIS	In adulti con ASD è utile l'utilizzo di strumenti strutturati di valutazione del profilo adattivo in aggiunta all'OCC. verso il solo OCC? Se si. quale?	7.71
6	TREATMENT	Negli adulti con ASD gli interventi di informazione/supporto per familiari sono efficaci per (OUTCOME degli adulti con ASD). rispetto all'assenza di tali interventi?	7.71
7	DIAGNOSIS	In adulti con ASD quali test o esami diagnostici dovrebbero essere effettuati per identificare l'eventuale presenza di co-occorrenze (sia psichiatriche che mediche)?	7.65
8	TREATMENT	Negli adulti con ASD la pianificazione di progetti individualizzati di vita. basati su una procedura standardizzata di assessment delle preferenze. verso una 9pianificazione non basata su una procedura standardizzata di assessment delle preferenze. è efficace per (OUTCOME)? Se si. quali sono le migliori procedure standardizzate di assessment delle preferenze?	7.65
9	TREATMENT	Negli adulti con ASD gli interventi di supporto alle attività occupazionali sono efficaci per (OUTCOME). verso l'assenza di attività occupazionali? Se si. quali?	7.65
10	TREATMENT	Negli adulti con ASD gli interventi di informazione/supporto per caregiver sono efficaci per (OUTCOME degli adulti con ASD). rispetto all'assenza di tali interventi?	7.59
11	TREATMENT	Negli adulti con ASD gli interventi di supporto alla vita autonoma sono efficaci per (OUTCOME). verso l'assenza di interventi di supporto alla vita autonoma? Se si. quali?	7.53
12	DIAGNOSIS	In adulti con ASD è utile l'utilizzo di test o esami diagnostici. in aggiunta all'OCC. per identificare l'eventuale presenza di co-occorrenze (sia psichiatriche che mediche). verso il solo OCC?	7.41

13	TREATMENT	Negli adulti con ASD i servizi con équipe dedicate alla transizione dall'età evolutiva all'età adulta. verso servizi senza équipe dedicate alla transizione dall'età evolutiva all'età adulta. sono efficaci per (OUTCOME)?	7.35
14	TREATMENT	Negli adulti con ASD le psicoterapie sono efficaci per (OUTCOME). verso l'assenza di psicoterapia? Se si. quali?	7.35
15	DIAGNOSIS	In adulti con sospetto di ASD quali test o esami diagnostici dovrebbero essere effettuati. in aggiunta all'OCC. per una corretta diagnosi differenziale con eventuali condizioni mediche o psichiatriche. verso il solo OCC?	7.35
16	TREATMENT	Negli adulti con ASD gli interventi di informazione/supporto per "altre figure" sono efficaci per (OUTCOME degli adulti con ASD). rispetto all'assenza di tali interventi?	7.29
17	DIAGNOSIS	In adulti con ASD è utile l'utilizzo di strumenti strutturati di valutazione del profilo cognitivo in aggiunta all'OCC. verso il solo OCC? Se si. quale?	7.24
18	DIAGNOSIS	In adulti con ASD è utile l'utilizzo di strumenti strutturati di valutazione del profilo neuropsicologico in aggiunta all'OCC. verso il solo OCC? Se si. quale?	7.24
19	TREATMENT	Negli adulti con ASD l'assunzione di farmaci antipsicotici è efficace per (OUTCOME). verso placebo / no treatment? Se si. quali?	7.13
20	TREATMENT	Negli adulti con ASD intraprendere un percorso abitativo. verso non intraprenderlo. è efficace per (OUTCOME)?	7.00
21	TREATMENT	Negli adulti con ASD l'assunzione di farmaci antiepilettici è efficace per (OUTCOME). verso placebo / no treatment? Se si. quali?	7.00
22	TREATMENT	Negli adulti con ASD l'assunzione di farmaci antidepressivi è efficace per (OUTCOME). verso placebo / no treatment? Se si. quali?	6.94
23	TREATMENT	Negli adulti con ASD l'assunzione di farmaci stimolanti è efficace per (OUTCOME). verso placebo / no treatment? Se si. quali?	6.53
24	TREATMENT	Negli adulti con ASD l'assunzione di benzodiazepine è efficace per (OUTCOME). verso placebo / no treatment? Se si. quali?	6.47
25	TREATMENT	Negli adulti con ASD l'assunzione di farmaci "affecting cognition" è efficace per (OUTCOME). verso placebo / no treatment? Se si. quali?	6.29
26	TREATMENT	Negli adulti con ASD intraprendere un percorso residenziale. verso non intraprenderlo. è efficace per (OUTCOME)?	6.18
27	TREATMENT	Negli adulti con ASD intraprendere un percorso semi-residenziale. verso non intraprenderlo. è efficace per (OUTCOME)?	6.12
28	TREATMENT	Negli adulti con ASD l'assunzione di terapie ormonali è efficace per (OUTCOME). verso placebo / no treatment? Se si. quali?	5.87
29	TREATMENT	Negli adulti con ASD gli interventi biomedical sono efficaci per (OUTCOME). verso l'assenza di tali interventi?	5.50

30	TREATMENT	Negli adulti con ASD gli interventi sul sonno sono efficaci per (OUTCOME). verso l'assenza di tali interventi?	5.47
31	TREATMENT	Negli adulti con ASD gli interventi nutrizionali sono efficaci per (OUTCOME). verso l'assenza di tali interventi?	4.94

Appendix 3: Ratings of research questions for the ISS guideline on management of autism spectrum disorder

Appendix 4

Si dovrebbero usare gli acidi grassi poli-insaturi vs. non usare gli acidi grassi poli-insaturi for il trattamento di disturbi dello spettro autistico in bambini e adolescenti??	
POPOLAZIONE:	bambini e adolescenti con disturbi dello spettro autistico
INTERVENTO:	acidi grassi poli-insaturi
CONFRONTO:	placebo
ESITI PRINCIPALI:	Discontinuation due to any cause; Iperattività; Qualità del sonno; Autolesionismo; Aggressività; Irritabilità; Ansia; Attenzione; Funzionamento adattivo; Interazione sociale; Interessi e comportamenti ristretti e ripetitivi; Comunicazione; Iperattività e comportamenti dirompenti coesistenti con i sintomi coreith core symptoms; Numero di eventi avversi;
SETTING:	pazienti ambulatoriali
PROSPETTIVA:	Sistema Sanitario Nazionale Italiano
BACKGROUND:	
CONFLITTI DI INTERESSE	Nessuno

VALUTAZIONE

Problema Il problema è una priorità?		
GIUDIZI	RICERCA DELLE PROVE DI EVIDENZA	CONSIDERAZIONI AGGIUNTIVE
○ No ○ Probabilmente no ○ Probabilmente si ● Si ○ Varia ○ Non so	Il disturbo dello spettro autistico è caratterizzato da un neurosviluppo anomalo, con alterazioni persistenti dell'interazione sociale, della comunicazione e con interessi e comportamenti ristretti e ripetitivi che causano un funzionamento ridotto, indipendentemente dalla abilità intellettiva (4). La prevalenza del disturbo dello spettro autistico tra i bambini in Italia è circa 1.35% (dati ISS, unpublished), mentre nel resto dell'Europa varia da 0.63% in Danimarca e Svezia, a 1.16% nel Regno Unito. Negli Stati Uniti la prevalenza è 1.69% attualmente ed è cresciuta molto negli ultimi 20 anni, passando da 0.67% nel 2000 a 1.14% nel 2008 a 1.69% nel 2014 (www.cdc.gov/ncbddd/autism/documents/ASDPrevalenceDataTable2016.pdf). La prevalenza media nel mondo, da studi europei, asiatici ed americani, si attesta tra 1% e 2% (www.cdc.gov/ncbddd/autism/documents/ASDPrevalenceDataTable2016.pdf). Un recente studio italiano, effettuato su 7927 bambini ed adolescenti con diagnosi dello spettro autistico, ha trovato che il rapporto maschi: femmine è di circa 4:1 (5) e che un bambino su due (47.6%) ha anche una disabilità intellettiva, in accordo con la letteratura internazionale (6). I costi del disturbo dello spettro autistico sono enormi sia per le famiglie che per la società. Una revisione recente negli Stati Uniti e nel Regno Unito ha considerato un costo totale per tutta la vita di circa 2 milioni di euro per supportare un bambino con disturbo dello spettro autistico e disabilità intellettiva e di circa 1.2 milioni di euro per supportare un bambino con disturbo dello spettro autistico in assenza di disabilità intellettiva (7), con costi omogenei tra le due nazioni e ripartiti soprattutto in educazione ed in perdita di lavoro genitoriale. L'autismo nel Regno Unito è la patologia con maggiori costi socio-sanitari, maggiore rispetto alle demenze e maggiore rispetto a patologie tumorali, patologie cardiovascolari ed ictus messe insieme (7). Gli individui con disturbo dello spettro autistico hanno bisogno di sostegno da parte di servizi sanitari e assistenziali, medici, farmacie e ospedali, per tutta la vita. Tuttavia, troppi individui con disturbo dello spettro autistico aspettano anni prima di ottenere una diagnosi attraverso il servizio sanitario nazionale e di ottenere la cura e il supporto di cui hanno bisogno (8). Inoltre, molti individui con disturbo dello spettro autistico continuano ad avere una salute fisica e mentale significativamente peggiore rispetto al pubblico in generale e possono anche essere maggiormente a rischio di morire presto (9). Il disturbo dello spettro autistico è stato dichiarato come una parte fondamentale del sistema sanitario Italiano, la cui cura può essere modellata intorno ai bisogni delle persone (10).	

	Gli acidi grassi poli-insaturi (PUFA) sono grassi che contengono almeno due doppi legami carbonio-carbonio nella loro catena carbossilica. I PUFA si dividono, secondo la distanza del primo doppio legame dal gruppo metilico posto al termine della molecola, in omega-3, omega-6 e omega-9 (questi ultimi non sono essenziali nell'uomo in quanto possono essere sintetizzati dai carboidrati o da altri acidi grassi). Gli oli di pesce sono ricchi di omega-3, quelli delle piante di omega-6 e due PUFA, l'acido alfalinoleico (omega-3) e l'acido linoleico (omega-6) sono nutrienti essenziali nell'uomo (11). Il ruolo di EPA e DHA nei disturbi del sistema nervoso centrale è stato ampiamente indagato nelle ultime due decadi (2). EPA e DHA sono fattori importanti nello sviluppo dei sistemi nervoso e immune fetali. EPA e DHA sono importanti componenti dei fosfolipidi e degli esteri di colesterolo delle membrane neuronali, specialmente dei dendriti e delle sinapsi; pertanto il razionale nell'uso di questi agenti nei disordini di natura psichiatrica sarebbe proprio la loro azione primaria nel produrre modificazioni della membrana sinaptica, con implicazioni nella trasmissione e trasduzione del segnale (2). Ad esempio, studi di risonanza magnetica hanno suggerito che una ridotta connessione funzionale di aree cerebrali a lunga distanza è correlata alle difficoltà nelle interazioni sociali nel disturbo dello spettro autistico (1). In ambito psichiatrico, EPA e DHA sono stati sperimentati nella terapia di ADHD, autismo, disturbo bipolare, unipolare, disturbi d'ansia, disturbo ossessivo-compulsivo, aggressività, ostilità, impulsività, disturbo di personalità borderline, uso di sostanze e anoressia nervosa (2). L'ipotesi è che i PUFA possano essere efficaci sui sintomi core dell'autismo e che abbiano un buon profilo di sicurezza. Ultimamente c'è stato un aumento sul mercato della disponibilità di molti farmaci e integratori alimentari ed è aumentato anche il numero di PUFA ad uso pediatrico. Un'altra ipotesi è che l'assunzione indiscriminata ed eccessiva di PUFA potrebbe dare dei problemi di sicurezza ed è fondamentale raccomandarne un uso appropriato (3).	

Effetti desiderabili

Quanto considerevoli sono gli effetti desiderabili attesi?

GIUDIZI	RICERCA DELLE PROVE DI EVIDENZA	CONSIDERAZIONI AGGIUNTIVE
○ Irrilevanti ●Piccoli ○ Moderati ○ Grandi ○ Variano ○ Non lo so	È stata effettuata una ricerca sistematica della letteratura sulle banche dati CENTRAL, PubMed/Medline, Embase, PsycINFO, Web Of Science, dalla data della creazione delle rispettive banche dati fino al 30 Ottobre 2018, senza limitazioni di lingua. La strategia di ricerca è disponibile su richiesta. La selezione degli studi, l'estrazione dei dati, la valutazione del rischio di bias e della certezza dell'evidenza secondo il metodo GRADE è stata effettuata da due revisori in modo indipendente. I risultati degli studi che effettuavano confronti diretti sono stati combinati attraverso meta-analisi *pairwise* usando come misure di risultato il risk ratio (RR) per gli esiti dicotomici e la differenza standardizzata tra medie (SMD) per gli esiti continui, utilizzando per entrambe le misure un modello ad effetti casuali. **Risultati della ricerca per singole Banche dati** CENTRAL= 69 Embase= 209 MEDLINE= 153 PsycINFO= 90 Web Of Science= 265 Numero di documenti trovati dopo ricerca sistematica= 786 Numero di documenti trovati dopo la rimozione dei duplicati= 558 La strategia di ricerca utilizzata ha permesso il ritrovamento di 786 documenti, dei quali 228 sono stati rimossi, essendo dei duplicati. Dei 558 documenti da valutare, 24 documenti sono stati valutati in full-text. Di questi, 12 sono stati esclusi. Tra gli esclusi, 4 studi includevano bambini nati pretermine tra i 18 ed i 36 mesi con alto rischio di ASD (21, 22); (23) (24), 3 studi non avevano un gruppo di controllo (25); (26); (27), 1 studio non era un RCT (28), 1 studio era un case report (29). Abbiamo trovato uno studio clinico in corso (ACTRN1265000144516) e 2 studi clinici completati ma di cui non abbiamo ancora i risultati (clinicaltrials.gov identifiers: NCT00577447; NCT02059577). Infine, 12 documenti, corrispondenti ad un totale di 9 studi (351 partecipanti) sono stati inclusi (20); (30); (31); (12); (13); (14, 15); (16, 17); (18) (19, 1). Gli studi clinici che comprendevano partecipanti in età prescolare erano 6 (66.6%), mentre 3 studi clinici (33.3%) includevano anche individui adolescenti. La maggior parte degli individui inclusi era di	Queste regole empiriche si basano sulle analisi di Cohen a riguarda della proporzione dell'effetto: - 0.2 è un piccolo effetto (ad esempio SMD 0.45 corrisponde ad un piccolo effetto); - 0.5 è un effetto moderato (ad esempio, SMD 0.7 corrisponde ad un effetto moderato); - 0,8 è un grande effetto (ad esempio, SMD 0,95

	sesso maschile (86.6%). In 7 casi la diagnosi è stata effettuata utilizzando i criteri del DSM-IV, in un caso i criteri del DSM-5, in un altro caso la diagnosi era riportata dai genitori. In 5 casi è stato riportato l'utilizzo di scale per supporto alla diagnosi, tra cui ADI-R, ADOS, CARS, SCQ. Tra i 9 studi inclusi, 8 studi hanno confrontato i PUFA verso placebo, mentre 1 solo studio ha confrontato i PUFA verso un intervento in cui veniva proposto di seguire una dieta sana. Degli studi, 2 sono stati condotti in Europa, 5 in Nord America, 1 in Asia, 1 in Oceania. Le scale usate per la misurazione degli esiti includevano la Aberrant Behavior Checklist (ABC), la Behavior Assessment System for Children (BASC), la Expressive Vocabulary Test (EVT) , la Mullen Scales of Early Learning, la Peabody Picture Vocabulary Test (PPVT), la Social Responsiveness Scale (SRS), la Vineland Adaptive Behavior Scale (VABS). Per quanto riguarda la composizione dei PUFA, in 5 studi clinici era presente una combinazione di acido eicosapentaenoico (EPA) ed acido docosaesaenoico (DHA), mentre in 4 studi era presente solo DHA. Le dosi di EPA variavano tra 693mg e 840mg/die, mentre le dosi di DHA variavano dai 200mg ai 722mg/die. La modalità di assunzione di PUFA variava grandemente negli studi clinici randomizzati analizzati, con una dose mediana di 1155 mg/die, e dosaggi da un minimo di 200 mg/die (18) ad un massimo di 1540 mg/die (20). La mediana della durata degli studi clinici era di 12 settimane (range: 6-52). Per quanto riguarda l'assunzione raccomandata di acidi grassi omega-3 per neonati, l'OMS suggerisce 400 mg per 10 kg di peso corporeo (WHO/FAO Expert Consultation on Diet, undefined) , (Lee, 2013)), mentre l' International Scientific Society of Fatty Acids and Lipids (ISSFAL) suggerisce 350-750 mg ogni 10 kg di peso corporeo (http://www.issfal.org/news-links/resources/publications/PUFAIntakeReccomdFinalReport.pdf, undefined). Riguardo invece la massima dose tollerabile di omega-3, la Food and Drug Administration (FDA) raccomanda di non assumere più di 3 g/die di EPA e DHA, dei quali fino a 2d/die attraverso i supplementi (National Institutes of Health, undefined) La limitazione giornaliera è importante al fine di limitare l'assunzione di vitamine liposolubili, quali Vitamina A e Vitamina D (Bays, 2007) (Lee, 2013). L'Institute Of Medicine (IOM) non ha stabilito un tolerable Upper Intake Level (UL) per l'assunzione di omega-3, ma ha evidenziato che dosi elevate (più di 900mg/die di EPA più 600 mg/die di DHA) potrebbero ridurre la risposta immunitaria, mentre dosi tra i 2 e i 15 grammi di EPA e/o DHA potrebbero avere effetti negativi sulla coagulazione, favorendo i sanguinamenti (Institute of Medicine et al., undefined). Secondo la European Food Safety Authority (EFSA), invece, la supplementazione con dosi fino a 5g/die di EPA e/o DHA sarebbe sicura, non essendo stati riscontrati effetti collaterali riguardo il sanguinamento e risposta immune (EFSA Panel on Dietetic Products NaA. Scientific opinion on the tolerable upper intake level of eicosapentaenoic acid (EPA) et al., undefined). Una revisione sistematica recente sottolinea la differenza in materia di sicurezza tra Omega-3 prodotti come nutraceutici rispetto ai farmacologici, sottolineando come i prodotti farmacologici prescritti sono supportati da robusti programmi di sviluppo clinico e di monitoraggio della sicurezza, mentre i prodotti nutraceutici non sono tenuti a dimostrare sicurezza o efficacia prima del marketing (Hilleman D, 2016). I nutraceutici possono anche contenere componenti potenzialmente dannosi, tra cui altri lipidi, colesterolo e tossine e non sono prodotti in Good Manifacturing Practice (GMP). I prodotti farmacologici omega-3 possono contenere DHA ed EPA o EPA ad elevata purezza (Hilleman D, 2016) (Santini A, 2018). Nonostante nei prodotti ittici sia presete metil-mercurio in varie quantità, questo non si dovrebbe ritrovare abitualmente nei supplementi a base di omega-3, in quanto rimosso nel processo di produzione (ConsumerLab.com. Product review: fish oil and omega-3 fatty acid supplements review (including krill, undefined); (National Institutes of Health, undefined) Oltre alla presa in esame degli studi inclusi, dalla precedente revisione sistematica Cochrane (James S, 2011) abbiamo anche ripreso gli studi osservazionali esclusi, per valutare la presenza di eventuali evidenze aggiuntive sull'accettabilità e la sicurezza della supplementazione con omega-3 negli individui autistici in età pediatrica. Tra gli studi osservazionali, alcuni andavano a valutare la sicurezza dell'assunzione dei PUFA per individui con disturbo dello spettro autistico. In uno studio la supplementazione con PUFA è stata associata ad un aumento dell' iperattività e problemi comportamentali, riferiti dai genitori (Bell JG, 2004).	corrisponde ad un grande effetto). Trivial: Insignificante - 2 Small: Basso - 13 Moderate: Moderato- 2 Large: Grande- 0 Astenuto- 1

Esiti	Effetto assoluto anticipato* (95% CI)		Effetto relativo (95% CI)	№ dei partecipanti (studi)	Certainty of the evidence (GRADE)	Commenti
	Rischio con placebo	Rischio con acidi grassi poli-insaturi				
Iperattività	La media iperattività eran-a	SMD 0.27 inferiore (0.6 inferiore a 0.06 maggiore)	-	146 (5 RCT)	⊕⊕◯◯ BASSA[a]	
Aggressività	La media aggressività eran-a	SMD 0.29 inferiore (1.08 inferiore a 0.49 maggiore)	-	25 (1 RCT)	⊕⊕◯◯ BASSA[a]	
Irritabilità	La media irritabilità eran-a	SMD 0.02 inferiore (0.42 inferiore a 0.38 maggiore)	-	146 (5 RCT)	⊕⊕◯◯ BASSA[a]	
Ansia	La media ansia eran-a	SMD 1.01 inferiore (1.86 inferiore a 0.17 inferiore)	-	25 (1 RCT)	⊕◯◯◯ MOLTO BASSA[a,b]	
Funzionamento adattivo	La media funzionamento adattivo eran-a	SMD 0.49 inferiore (1.2 inferiore a 0.22 maggiore)	-	59 (2 RCT)	⊕◯◯◯ MOLTO BASSA[a,c,d]	
Interazione sociale	La media interazione sociale eran-a	SMD 0.27 maggiore (0.03 inferiore a 0.57 maggiore)	-	172 (4 RCT)	⊕◯◯◯ MOLTO BASSA[a,e]	

Interessi e comportamenti ristretti e ripetitivi	La media interessi e comportamenti ristretti e ripetitivi eran-a	SMD 0.01 maggiore (0.36 inferiore a 0.39 maggiore)	-	223 (6 RCT)	⨁⨁◯◯ BASSA[a]	
Comunicazione	La media comunicazione era n-a SD	SMD 0.05 SD inferiore (0.5 inferiore a 0.4 maggiore)	-	223 (6 RCT)	⨁⨁◯◯ BASSA[a]	
Numero di eventi avversi	Popolazione in studio		**RR 1.54** (0.79 a 2.97)	157 (5 RCT)	⨁⨁◯◯ BASSA[f]	
	132 per 1,000	203 per 1,000 (104 a 391)				

Downgraded of two levels because population size <400 and there is a wide 95%CI, which includes no effect
Downgraded of one level because the measure used was the internalizing subscale of the BASC, which only indirectly measures anxiety
Downgraded of one level because one study is at high risk for incomplete outcome data and unclear risk for blinding and selective reporting
Downgraded of one level, because in one study the "social skills, parents assessed" of the subscale "adaptive skills" of the BASC was extracted
Downgraded of one level because in two studies Social interaction was analysed by the "inappropriate speech" subscale of the ABC, which relates more to behaviour and indirectly to social interaction
Downgraded of two levels because optimal information size (OIS) not met and there is a wide 95%CI, which includes no effect

Effetti indesiderabili

Quanto considerevoli sono gli effetti indesiderabili attesi?

GIUDIZI	RICERCA DELLE PROVE DI EVIDENZA	CONSIDERAZIONI AGGIUNTIVE

○ Grandi ○ Moderati ●Piccoli ○ Irrilevanti ○ Variano ○ Non lo so	È stata effettuata una ricerca sistematica della letteratura sulle banche dati CENTRAL, PubMed/Medline, Embase, PsycINFO, Web Of Science, dalla data della creazione delle rispettive banche dati fino al 30 Ottobre 2018, senza limitazioni di lingua. La strategia di ricerca è disponibile su richiesta. La selezione degli studi, l'estrazione dei dati, la valutazione del rischio di bias e della certezza dell'evidenza secondo il metodo GRADE è stata effettuata da due revisori in modo indipendente. I risultati degli studi che effettuavano confronti diretti sono stati combinati attraverso meta-analisi *pairwise* usando come misure di risultato il risk ratio (RR) per gli esiti dicotomici e la differenza standardizzata tra medie (SMD) per gli esiti continui, utilizzando per entrambe le misure un modello ad effetti casuali. **Risultati della ricerca per singole Banche dati** CENTRAL= 69 Embase= 209 MEDLINE= 153 PsycINFO= 90 Web Of Science= 265 Numero di documenti trovati dopo ricerca sistematica= 786 Numero di documenti trovati dopo la rimozione dei duplicati= 558 La strategia di ricerca utilizzata ha permesso il ritrovamento di 786 documenti, dei quali 228 sono stati rimossi, essendo dei duplicati. Dei 558 documenti da valutare, 24 documenti sono stati valutati in full-text. Di questi, 12 sono stati esclusi. Tra gli esclusi, 4 studi includevano bambini nati pretermine tra i 18 ed i 36 mesi con alto rischio di ASD (21, 22); (23) (24), 3 studi non avevano un gruppo di controllo (25); (26); (27), 1 studio non era un RCT (28), 1 studio era un case report (29). Abbiamo trovato uno studio clinico in corso (ACTRN1265000144516) e 2 studi clinici completati ma di cui non abbiamo ancora i risultati (clinicaltrials.gov identifiers: NCT00577447; NCT02059577). Infine, 12 documenti, corrispondenti ad un totale di 9 studi (351 partecipanti) sono stati inclusi (20); (30); (31); (12); (13); (14, 15); (16, 17); (18) (19, 1). Gli studi clinici che comprendevano partecipanti in età prescolare erano 6 (66.6%), mentre 3 studi clinici (33.3%) includevano anche individui adolescenti. La maggior parte degli individui inclusi era di sesso maschile (86.6%). In 7 casi la diagnosi è stata effettuata utilizzando i criteri del DSM-IV, in un caso i criteri del DSM-5, in un altro caso la diagnosi era riportata dai genitori. In 5 casi è stato riportato l'utilizzo di scale per supporto alla diagnosi, tra cui ADI-R, ADOS, CARS, SCQ. Tra i 9 studi inclusi, 8 studi hanno confrontato i PUFA verso placebo, mentre 1 solo studio ha confrontato i PUFA verso un intervento in cui veniva proposto di seguire una dieta sana. Degli studi, 2 sono stati condotti in Europa, 5 in Nord America, 1 in Asia, 1 in Oceania. Le scale usate per la misurazione degli esiti includevano la Aberrant Behavior Checklist (ABC), la Behavior Assessment System for Children (BASC), la Expressive Vocabulary Test (EVT) , la Mullen Scales of Early Learning, la Peabody Picture Vocabulary Test (PPVT), la Social Responsiveness Scale (SRS), la Vineland Adaptive Behavior Scale (VABS). Per quanto riguarda la composizione dei PUFA, in 5 studi clinici era presente una combinazione di acido eicosapentaenoico (EPA) ed acido docosaesaenoico (DHA), mentre in 4 studi era presente solo DHA. Le dosi di EPA variavano tra 693mg e 840mg/die, mentre le dosi di DHA variavano dai 200mg ai 722mg/die. La modalità di assunzione di PUFA variava grandemente negli studi clinici randomizzati analizzati, con una dose mediana di 1155 mg/die, e dosaggi da un minimo di 200 mg/die (18) ad un massimo di 1540 mg/die (20). La mediana della durata degli studi clinici era di 12 settimane (range: 6-52). Per quanto riguarda l'assunzione raccomandata di acidi grassi omega-3 per neonati, l'OMS suggerisce 400 mg per 10 kg di peso corporeo (WHO/FAO Expert Consultation on Diet, undefined) , (Lee, 2013)), mentre l' International Scientific Society of Fatty Acids and Lipids (ISSFAL) suggerisce 350-750 mg ogni 10 kg di peso corporeo (http://www.issfal.org/news-links/resources/publications/PUFAIntakeReccomdFinalReport.pdf, undefined). Riguardo invece la massima dose tollerabile di omega-3, la Food and Drug Administration (FDA) raccomanda di non assumere più di 3 g/die di EPA e DHA, dei quali fino a 2d/die attraverso i supplementi (National Institutes of Health, undefined) La limitazione giornaliera è importante al fine di limitare l'assunzione di vitamine liposolubili, quali Vitamina A e Vitamina D (Bays, 2007) (Lee, 2013). L'Institute Of Medicine (IOM) non ha stabilito un tolerable Upper Intake Level (UL) per l'assunzione di omega-3, ma ha evidenziato che dosi elevate (più di 900mg/die di EPA più 600 mg/die di DHA)	

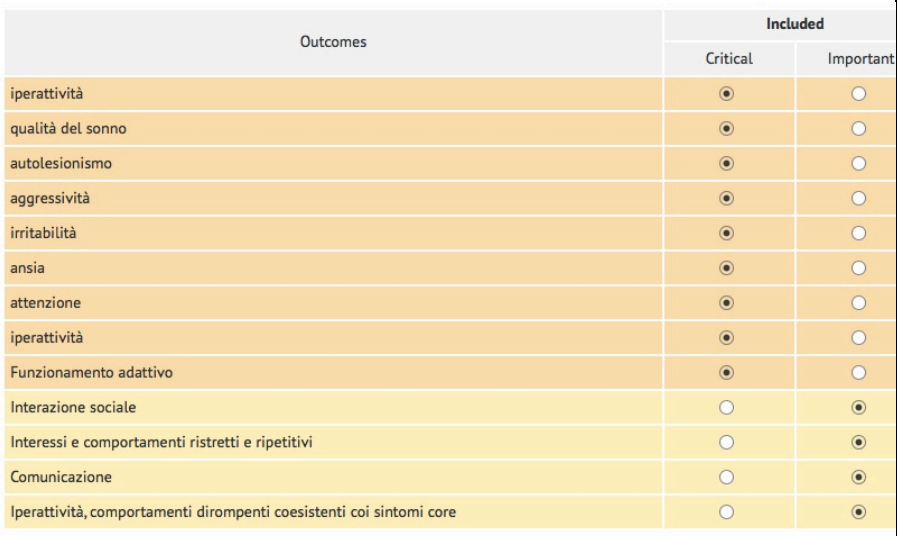
potrebbero ridurre la risposta immunitaria, mentre dosi tra i 2 e i 15 grammi di EPA e/o DHA potrebbero avere effetti negativi sulla coagulazione, favorendo i sanguinamenti (Institute of Medicine et al., undefined). Secondo la European Food Safety Authority (EFSA), invece, la supplementazione con dosi fino a 5g/die di EPA e/o DHA sarebbe sicura, non essendo stati riscontrati effetti collaterali riguardo il sanguinamento e risposta immune (EFSA Panel on Dietetic Products NaA. Scientific opinion on the tolerable upper intake level of eicosapentaenoic acid (EPA) et al., undefined).
Una revisione sistematica recente sottolinea la differenza in materia di sicurezza tra Omega-3 prodotti come nutraceutici rispetto ai farmacologici, sottolineando come i prodotti farmacologici prescritti sono supportati da robusti programmi di sviluppo clinico e di monitoraggio della sicurezza, mentre i prodotti nutraceutici non sono tenuti a dimostrare sicurezza o efficacia prima del marketing (Hilleman D, 2016). I nutraceutici possono anche contenere componenti potenzialmente dannosi, tra cui altri lipidi, colesterolo e tossine e non sono prodotti in Good Manifacturing Practice (GMP). I prodotti farmacologici omega-3 possono contenere DHA ed EPA o EPA ad elevata purezza (Hilleman D, 2016) (Santini A, 2018). Nonostante nei prodotti ittici sia presete metil-mercurio in varie quantità, questo non si dovrebbe ritrovare abitualmente nei supplementi a base di omega-3, in quanto rimosso nel processo di produzione (ConsumerLab.com. Product review: fish oil and omega-3 fatty acid supplements review (including krill, undefined); (National Institutes of Health, undefined)
Oltre alla presa in esame degli studi inclusi, dalla precedente revisione sistematica Cochrane (James S, 2011) abbiamo anche ripreso gli studi osservazionali esclusi, per valutare la presenza di eventuali evidenze aggiuntive sull'accettabilità e la sicurezza della supplementazione con omega-3 negli individui autistici in età pediatrica. Tra gli studi osservazionali, alcuni andavano a valutare la sicurezza dell'assunzione dei PUFA per individui con disturbo dello spettro autistico. In uno studio la supplementazione con PUFA è stata associata ad un aumento dell' iperattività e problemi comportamentali, riferiti dai genitori (Bell JG, 2004).

Esiti	Effetto assoluto anticipato* (95% CI)		Effetto relativo (95% CI)	№ dei partecipanti (studi)	Certainty of the evidence (GRADE)	Commenti
	Rischio con placebo	Rischio con acidi grassi poli-insaturi				
Iperattività	La media iperattività eran-a	SMD 0.27 inferiore (0.6 inferiore a 0.06 maggiore)	-	146 (5 RCT)	⊕⊕◯◯ BASSA[a]	
Aggressività	La media aggressività eran-a	SMD 0.29 inferiore (1.08 inferiore a 0.49 maggiore)	-	25 (1 RCT)	⊕⊕◯◯ BASSA[a]	
Irritabilità	La media irritabilità eran-a	SMD 0.02 inferiore (0.42	-	146 (5 RCT)	⊕⊕◯◯ BASSA[a]	

		inferiore a 0.38 maggiore)				
Ansia	La media ansia eran-a	SMD 1.01 inferiore (1.86 inferiore a 0.17 inferiore)	-	25 (1 RCT)	⊕◯◯◯ MOLTO BASSA[a,b]	
Funzionamento adattivo	La media funzionamento adattivo eran-a	SMD 0.49 inferiore (1.2 inferiore a 0.22 maggiore)	-	59 (2 RCT)	⊕◯◯◯ MOLTO BASSA[a,c,d]	
Interazione sociale	La media interazione sociale eran-a	SMD 0.27 maggiore (0.03 inferiore a 0.57 maggiore)	-	172 (4 RCT)	⊕◯◯◯ MOLTO BASSA[a,e]	
Interessi e comportamenti ristretti e ripetitivi	La media interessi e comportamenti ristretti e ripetitivi eran-a	SMD 0.01 maggiore (0.36 inferiore a 0.39 maggiore)	-	223 (6 RCT)	⊕⊕◯◯ BASSA[a]	
Comunicazione	La media comunicazione era n-a SD	SMD 0.05 SD inferiore (0.5 inferiore a 0.4 maggiore)	-	223 (6 RCT)	⊕⊕◯◯ BASSA[a]	
Numero di eventi avversi	Popolazione in studio		**RR 1.54** (0.79 a 2.97)	157 (5 RCT)	⊕⊕◯◯ BASSA[f]	
	132 per 1,000	203 per 1,000 (104 a 391)				

Downgraded of two levels because population size <400 and there is a wide 95%CI, which includes no effect
Downgraded of one level because the measure used was the internalizing subscale of the BASC, which only indirectly measures anxiety
Downgraded of one level because one study is at high risk for incomplete outcome data and unclear risk for blinding and selective reporting
Downgraded of one level, because in one study the "social skills, parents assessed" of the subscale "adaptive skills" of the BASC was extracted

	Downgraded of one level because in two studies Social interaction was analysed by the "inappropriate speech" subscale of the ABC, which relates more to behaviour and indirectly to social interaction Downgraded of two levels because optimal information size (OIS) not met and there is a wide 95%CI, which includes no effect	
Certezza delle prove Qual'è la certezza complessiva delle prove di efficacia e sicurezza?		
GIUDIZI	RICERCA DELLE PROVE DI EVIDENZA	CONSIDERAZIONI AGGIUNTIVE
● Molto bassa ○ Bassa ○ Moderata ○ Alta ○ Nessuno studio incluso	La certezza delle prove è stata abbassata per imprecisione, rischio di distorsione sistematica, valutazioni indirette della misura di esito. Complessivamente, la certezza delle prove è molto bassa (i due esiti critici "ansia" e "funzionamento adattivo" hanno una certezza delle prove molto bassa).	
Valori C'è incertezza o variabilità su quanto le persone possano considerare importanti gli esiti principali?		
GIUDIZI	RICERCA DELLE PROVE DI EVIDENZA	CONSIDERAZIONI AGGIUNTIVE

○ Importante incertezza o variabilità ● Probabilmente importante e incertezza o variabilità ○ Probabilmente non importante e incertezza o variabilità ○ Nessuna importante e incertezza o variabilità	 Recentemente, la Fondazione di pazienti americana "Autism Speaks" ha supportato gruppi di lavoro di esperti per valutare gli strumenti di misurazione di tre esiti importanti: 1) interessi ristretti e comportamenti ripetitivi (32); 2) Ansia (33); 3) Comportamenti di comunicazione sociale (34). Non ci sono lavori sui Patient Reported Outcomes (PROMS) specifici per il disturbo dello spettro autistico (Digital Education Resource Archive, Oxford Patient-Reported Outcomes Measurement database), mentre è presente una revisione sistematica sulle misure di esito per bambini con disturbo dello spettro autistico, denominata progetto MeASURe (Measurement in Autism Spectrum disorder Under Review) (35). Il progetto MeASURe ha cercato di identificare valori specifici del bambino e / o della famiglia che i genitori dei bambini con ASD percepiscono come importanti, in tre modi diversi: 1) intraprendendo una revisione della letteratura qualitativa, usando le banche dati MEDLINE, CINAHL e PsycINFO; 2) conducendo una consultazione - tramite gruppi e via e-mail - per bambini ed adolescenti con disturbo dello spettro autistico; 3) facendo un sondaggio attraverso reti di professionisti della salute e dell'istruzione per esplorare quali esiti sono più spesso misurati dai professionisti della prima infanzia nel monitorare i progressi dei bambini. Il progetto MeASUrE (35) ha riscontrato una notevole differenza tra gli esiti valutati importanti dai genitori e gli esiti più frequentemente misurati da clinici ed educatori. L'esperienza dei genitori con i loro figli li portano a enfatizzare risultati come il benessere emotivo dei bambini ed il loro funzionamento all'interno del nucleo famigliare. I clinici e gli educatori hanno riconosciuto di essere portati a misurare ciò per cui hanno gli strumenti. Inoltre, hanno riconosciuto che la loro pratica è influenzata da un'enfasi sulle caratteristiche core che definiscono lo spettro del disturbo autistico e sui disturbi del comportamento, quindi sono più portati a pensare come il bambino agisce sull'ambiente piuttosto che vedere il quadro più ampio e misurare come il bambino è influenzato dall' ambiente (35). Il progetto MeASUrE, dopo aver condotto interviste a clinici, a genitori ed a pazienti sui valori e le preferenze degli esiti, ha prodotto una lista di strumenti usati, raggruppati per dominio concettuale primario: 1) Severità dei sintomi autistici: Autism Behavior Checklist; Autism Diagnostic Interview-Revised (ADI-R); Autism Diagnostic Observation Schedule (ADOS, including Toddler Module and Calibrated Severity Score); Autism Observation Scale for Infants; The Baby and Infant Screen for Children with aUtIsm Traits-Part 1 (BISCUIT); Behavioral Summarized Evaluation (BSE-R; including Revised and Infant); Childhood Autism Rating Scale; Gilliam Autism Rating Scale (GARS and GARS-2); Modified Checklist for Autism in Toddlers; Parent Observation of Early Markers Scale; Pervasive Developmental Disorders Rating Scale; Social Communication Questionnaire; Social Responsiveness Scale (SRS). 2) Misure globali di esiti: Autism Treatment and Evaluation Checklist; Pervasive Developmental Disorders Behavior Inventory (PDDBI).	

	3) Coscienza sociale: Imitation Battery; Preschool Imitation and Praxis Scale (PIPS). 4) Interessi ristretti e ripetitivi: Repetitive Behavior Scale-Revised. 5) Sensory processing: Sense and Self-Regulation Checklist; Sensory Profile including Short Sensory Profile. 6) Linguaggio: MacArthur–Bates Communicative Development Inventories (MCDI); Preschool Language Scale-Fourth Edition. 7) Abilità cognitive: Leiter International Performance Scale-Revised; Mullen Scales of Early Learning; Stanford–Binet Intelligence Scales-Fifth Edition. 8) Regolazione emotiva: Baby and Infant Screen for Children with aUtIsm Traits-Part 2 (BISCUIT-Part 2); Children's Global Assessment Scale; Infant–Toddler Social–Emotional Assessment (including Brief form). 9) Gioco: Test of Pretend Play. 10) Problemi comportamentali: Child Behavior Checklist (CBCL 1.5–5 and CBCL 6–18); Aberrant Behavior Checklist; BISCUIT-Part 3; Home Situations Questionnaire-Pervasive Developmental Disorders (HSQ-PDD) version; Nisonger Child Behavior Rating Form. 11) Misure globali di funzionamento: Global measure of functioning Behavior Assessment System for Children-Second Edition; Psychoeducational Profile-Revised (and Third Edition); Scales of Independent Behavior-Revised; Vineland Adaptive Behavior Scales (VABS; including Classroom and Screener versions). 12) Stress Genitoriale: Autism Parenting Stress Index; Parenting Stress Index-Short Form (PSI-SF); Questionnaire on Resources and Stress-Friedrich Short Form.	
Bilanciamento degli effetti Il bilanciamento tra effetti desiderabili ed indesiderabili favorisce l'intervento o il confronto?		
GIUDIZI	RICERCA DELLE PROVE DI EVIDENZA	CONSIDERAZIONI AGGIUNTIVE
○ E' in favore del confronto ○ Probabilmente è in favore del confronto ● Non è in favore né dell'intervento né del confronto ○ Probabilmente è in favore dell'intervento ○ E' in favore dell'intervento		

○ Varia ○ Non lo so		

Risorse necessarie

Quanto grandi sono le risorse necessarie (costi)?

GIUDIZI	RICERCA DELLE PROVE DI EVIDENZA	CONSIDERAZIONI AGGIUNTIVE
○ Costi molto elevati ● Costi moderati ○ Costi e risparmi irrilevanti ○ Risparmi moderati ○ Grandi risparmi ○ Varia ○ Non so	**Il costo della supplementazione con PUFA nel mondo** Uno studio Statunitense ha identificato nei supplementi il mezzo più economico, subito dopo l'olio di fegato di pesce, per assumere un dosaggio elevato (≥500 mg/die) di PUFA quali DHA ed EPA. Il costo dell'equivalente di una compressa contenente 1000 mg di EPA più DHA risultava infatti pari a 0.88$ ± 0.16$ (36). Analogamente, secondo un altro studio americano, che ha preso in considerazione un numero di supplementi presenti sul mercato più elevato, il costo dell'equivalente di una compressa contenente 1000 mg di EPA più DHA era di 0.70$ ± 1.11$ (37). **Il costo della supplementazione con PUFA in Italia** Prendendo in considerazione gli RCT inclusi per il calcolo dell'efficacia, si ottiene che la dose giornaliera mediana di PUFA somministrati per la terapia dei sintomi dell'autismo è pari a 1155 mg/die, e i dosaggi vanno da un minimo di 200 mg/die a un massimo di 1540 mg/die. Il prezzo del farmaco in Italia invece è stato ricavato selezionando tutti i farmaci attualmente in commercio e prescrivibili in classe A per patologia secondo le note AIFA 13 e 94. I dati relativi al numero di capsule per confezione, il dosaggio della singola capsula e il prezzo della confezione sono stati ricavati da Farmadati Italia Srl (38). Dalle schede tecniche dei farmaci così selezionati si evince che su 1000 mg di prodotto si ritrovano all'incirca 850 mg di principio attivo. Dall'analisi di questi dati si ricava che il prezzo mediano per 1000 mg di prodotto netto è di 0.68 euro, e va da un minimo di 0.65 euro a un massimo di 0.83 euro. Tenendo in considerazione i dati suddetti riguardanti le posologie utilizzate negli RCT e i prezzi dei farmaci, il costo di una giornata di terapia potrebbe dunque variare da 0.13 euro a 1.28 euro, con un valore mediano di 0.78 euro. Negli RCT vengono effettuati dei cicli di trattamento di durata molto variabile, dalle 6 alle 52 settimane, con una mediana di 12 settimane. Considerando questa variabilità, assieme alla variabilità di prezzo e di dosaggio, si può calcolare che, in un presunto scenario tipico in cui venisse effettuato un ciclo di 3 mesi a una posologia di 1155 mg/die di PUFA al costo di 0.68 euro/1000mg, il costo di un ciclo di terapia sarebbe pari a 65.51 euro. L'analisi di sensitività completa, con gli scenari di spesa minimo e massimo, è presentata in tabella. Non sono disponibili dati sul costo specifico per bambini ed adolescenti con disturbo dello spettro autistico.	Costi moderati- 11 Costi o risparmi trascurabili- 6 Astenuto-1

Certezza delle prove relativamente alle risorse necessarie

Qual'è la certezza delle prove relativamente alle risorse necessarie (costi)?		
GIUDIZI	RICERCA DELLE PROVE DI EVIDENZA	CONSIDERAZIONI AGGIUNTIVE
● Molto bassa ○ Bassa ○ Moderata ○ Alta ○ Nessuno studio incluso	Ci sono incertezze riguardanti il costo di compresse con bassi dosaggi di EPA più DHA. Infatti, tra i farmaci in classe A per la prevenzione secondaria delle malattie cardiovascolari, non vi era alcuna compressa di dosaggio inferiore a 500mg di PUFA, mentre per lo scenario di spesa minimo è stato preso in considerazione un dosaggio di 200mg/die, formulazione per la quale non erano disponibili dati di costo. I costi degli scenari di spesa minima potrebbero essere quindi leggermente superiori, visto il peso relativo dei costi fissi di confezionamento e distribuzione.	
Costo efficacia L'analisi di costo efficacia favorisce l'intervento o il confronto?		
GIUDIZI	RICERCA DELLE PROVE DI EVIDENZA	CONSIDERAZIONI AGGIUNTIVE
○ È in favore del confronto ○ Probabilmente è in favore del confronto ○ Non è in favore né del confronto né dell'intervento ○ Probabilmente è in favore dell'intervento ○ È in favore dell'intervento ○ Varia ● Nessuno studio incluso	Non ci sono dati di costo-efficacia in letteratura per l'utilizzo di Omega-3 in bambini ed adolescenti con disturbo dello spettro autistico.	
Equità Quale sarebbe l'impatto in termini di equità?		
GIUDIZI	RICERCA DELLE PROVE DI EVIDENZA	CONSIDERAZIONI AGGIUNTIVE

○ Riduce l'equità ● Probabilmente riduce l'equità ○ Probabilmente nessun impatto ○ Probabilmente migliora l'equità ○ Miglior a l'equità ○ Varia ○ Non lo so	Il farmaco, non avendo come indicazione la terapia del disturbo dello spettro autistico (ASD), non è rimborsato dal servizio sanitario nazionale (SSN) ed è completamente a carico della famiglia del paziente. I costi di un ciclo di trattamento non sono ben definiti ma non dovrebbero superare i 466,54 euro/anno (vedi sezione Risorse richieste). Secondo diversi studi, il titolo di studio dei genitori influenza la scelta di intraprendere una terapia alternativa o complementare, quale quella con PUFA (39), così come il carico percepito della terapia in termini di tempo, denaro ed energie influenzerebbe l'aderenza al trattamento. L'assunzione di farmaci o supplementi dovrebbe comunque avere un basso impatto sull'impegno da parte dei caregiver familiari, essendo un compito relativamente concreto e circoscritto per i genitori (40).	

Accettabilità

L'intervento è accettabile per i principali stakeholders?

GIUDIZI	RICERCA DELLE PROVE DI EVIDENZA	CONSIDERAZIONI AGGIUNTIVE
○ No ○ Probabilmente no ○ Probabilmente si ○ Si ● Varia ○ Non lo so	L'uso degli omega-3 negli individui affetti da disturbo dello spettro autistico sembra piuttosto diffuso: In uno studio osservazionale americano viene mostrato come la totalità dei bambini ed adolescenti con disturbo dello spettro autistico presi in considerazione effettuavano qualche forma di terapia, a cui il 55% affiancava l'utilizzo di almeno un supplemento nutrizionale; la supplementazione nutrizionale veniva considerata utile dal 50% dei genitori di individui con disturbo dello spettro autistico (ASD). Gli omega-3 erano utilizzati dal 18% dei pazienti considerati (42). Secondo altri studi, invece, a far uso di PUFA negli USA sarebbe il 51% dei bambini con ASD (43). Secondo uno studio volto a valutare l'aderenza riportata dai genitori ai trattamenti per l'ASD, a seconda del tipo di trattamento seguito, l'aderenza media al trattamento con approccio alternativo, quale l'utilizzo di supplementi come gli omega-3 o la dieta sana, risultava essere significativamente inferiore rispetto all'aderenza alla terapia farmacologica o alla terapia evolutiva, mentre era sovrapponibile alla terapia comportamentale. Sempre secondo questo studio, un importante fattore predittivo di aderenza è risultato essere il peso percepito della terapia sulla famiglia in termini di tempo, energie, denaro (40). Un'altra possibile criticità nell'assunzione dei supplementi di PUFA da parte dei bambini con disturbo dello spettro autistico potrebbe essere dovuta alla difficoltà nel deglutire le capsule contenenti il medicinale (44). Una possibile alternativa, utilizzata in alcuni studi, sarebbe quella di fornire i PUFA in una formulazione liquida, più adatta ai bambini, soprattutto se di età pre-scolare. Tuttavia, la stessa formulazione liquida, se accompagnata dalle caratteristiche sensoriali che tipicamente caratterizzano i supplementi contenenti olio di pesce, potrebbe potenzialmente ridurre la compliance al trattamento (44), (45), (23). In uno degli studi clinici randomizzati, per favorire l'aderenza al trattamento, si è deciso di optare per una capsula contenente un dosaggio inferiore di PUFA (200mg/die) e si afferma che, nonostante la supplementazione con dosaggi superiori sia stata presa in considerazione, questa è stata scartata proprio per le difficoltà nel somministrare anche solo una capsula al giorno nella popolazione autistica (18). Riguardo l'accettabilità, quella dei PUFA non si discosta da quella del Placebo negli studi clinici, RR 1.01 (95%CI 0.66, 1.54) ed è stata studiata in 7 studi con una popolazione complessiva di 315 bambini ed adolescenti con disturbo dello spettro autistico (vedi tabella in basso). Studi osservazionali sembrano confermare la buona accettabilità dei PUFA (45) (25) (28) (27). Diversi studi clinici randomizzati avevano inoltre tra gli obiettivi la valutazione dell'aderenza al trattamento, sia con metodiche soggettive che oggettive. In questi studi, l'aderenza è stata considerata da buona a eccellente (15), (12), (30), (18), (17), (23). Nel dettaglio, Mazahery riporta, nei bracci di pazienti randomizzati ad omega-3 e ad omega-3 più vitamina D, un aumento dell'indice di omega-3 rispettivamente del 4,4% e del 4% rispetto al baseline (14), mentre Voigt et al. (18) riportano un	

aumento significativo dei livelli di DHA circolante in tutti i soggetti randomizzati a PUFA, con un aumento mediano del 430% dei livelli di DHA plasmatico; Johnson et al (12) riportano che un solo partecipante allo studio (10%) non ha assunto regolarmente la medicazione, mentre nello studio di Bent et al (30), la compliance al trattamento era giudicata perfetta o quasi perfetta nel 69% dei pazienti randomizzati agli omega-3, contro il 75% dei pazienti randomizzati al placebo (75%). Percentuali molto elevate di aderenza (97%) sono riportate in entrambi i bracci da Parellada et al (17).
Infine, tra le evidenze riportate dagli studi osservazionali, una clinica psichiatrica prescolastica per il disturbo dello spettro autistico in Israele ha suggerito l'assunzione di Omega-3 a 250 bambini ed adolescenti con disturbo dello spettro autistico (41). Di questi, circa due terzi hanno accettato di prendere Omega-3. La metà di quelli che hanno accettato di prendere Omega-3 ha poi smesso l'assunzione perchè non ha visto alcun miglioramento o per il cattivo sapore. Non sono stati riscontrati altre barriere diverse dal cattivo sapore (41).

Esiti	Effetto assoluto anticipato* (95% CI)		Effetto relativo (95% CI)	№ dei partecipanti (studi)	Certainty of the evidence (GRADE)	Comment i
	Rischio con placebo	Rischio con acidi grassi poli-insaturi				
Discontinuazione dovuta a qualsiasi causa	Popolazione in studio		**RR 1.01** (0.66 a 1.54)	315 (7 RCT)	⊕⊕◯◯ BASSA[a]	
	213 per 1.000	215 per 1.000 (141 a 328)				

Downgraded of two levels because optimal information size (OIS) not met and there is a wide 95%CI, which includes no effect

Fattibilità
E' fattibile l'implementazione dell'intervento?

GIUDIZI	RICERCA DELLE PROVE DI EVIDENZA	CONSIDERAZIONI AGGIUNTIVE
○ No ○ Probabi	Abbiamo effettuato una ricerca bibliografica per identificare barriere e facilitatori che dessero delle indicazioni sulla fattibilità del trattamento con PUFA in bambini ed adolescenti con disturbo dello	

lmente no ● Probabi lmente si ○ Si ○ Varia ○ Non lo so	spettro autistico. Al fine di trovare studi sulla fattibilità, abbiamo valutato le revisioni sistematiche e gli studi clinici provenienti dalla ricerca sull'efficacia e la sicurezza dei PUFA e le revisioni sistematiche e gli studi clinici provenienti dalla ricerca sui valori e preferenze. Alcuni studi hanno mostrato come i medici non siano percepiti dai genitori di soggetti con disturbo dello spettro autistico (ASD) come sufficientemente ben informati riguardo le terapie alternative per l'ASD, tra le quali figura quella con i PUFA (42). La maggior parte dei genitori sceglierebbe di intraprendere una terapia alternativa o complementare da medici alternativi, infermieri e nutrizionisti. I medici dovrebbero essere in grado di intraprendere, con i genitori dei soggetti con ASD, una discussione sull'efficacia e i possibili rischi dei trattamenti alternativi o complementari (39), tra cui quello con PUFA. L'assunzione dei supplementi di PUFA da parte del paziente autistico pediatrico presenta tra nelle possibili difficoltà nell'applicazione anche l'aderenza al trattamento, sia nel convincere il paziente ad assumere il supplemento (difficoltà nella deglutizione delle compressa, caratteristiche sensoriali sgradevoli del prodotto) (44), (45), (23) (18), sia per la propensione dei genitori a reputare utile o efficace il trattamento (40). Tuttavia, negli studi considerati, l'aderenza alla terapia è generalmente buona o eccellente (14), (12), (30), (18), (17), (23), e così anche l'accettabilità, sia negli RCT inclusi che negli studi osservazionali (45) (25) (28) (27). L'utilizzo di medicinali alternativi e di supplementi nutrizionali è in crescita nei bambini ed adolescenti con disturbo dello sviluppo. La crescita di questi agenti è alta soprattutto per quelle patologie nelle quali esiste incertezza circa il trattamento più efficace, o quando questo è gravato da importanti effetti collaterali. La crescita della prescrizione off-label e la vendita di molti PUFA come agenti nutraceutici, non come farmaci, ha aumentato l'accessibilità a questi prodotti. Tuttavia, il controllo sulla qualità nei prodotti nutraceutici è minore rispetto a quello sul farmaco, con potenziali rischi per la sicurezza (46).	

RIASSUNTO DEI GIUDIZI

	GIUDIZI						
PROBLEMA	No	Probabilmente no	Probabilmente si	Si		Varia	Non so
EFFETTI DESIDERABILI	Irrilevanti	Piccoli	Moderati	Grandi		Varia	Non so
EFFETTI INDESIDERABILI	Grandi	Moderati	Piccoli	Irrilevanti		Varia	Non so
CERTEZZA DELLE PROVE	Molto bassa	Bassa	Moderata	Alta			Nessuno studio incluso
VALORI	Importante incertezza o variabilità	Probabilmente importante incertezza o variabilità	Probabilmente nessuna importante incertezza o variabilità	Nessuna importante incertezza o variabilità			
BILANCIAMENTO DEGLI EFFETTI	A favore del confronto	Probabilmente a favore del confronto	Non è favorevole né al confronto né all'intervento	Probabilmente a favore dell'intervento	A favore dell'intervento	Varia	Non so
RISORSE RICHIESTE	Costi elevati	Costi moderati	Costi e risparmi irrilevanti	Risparmi moderati	Grandi risparmi	Varia	Non so
CERTEZZA DELLE PROVE RELATIVAMENTE	Molto bassa	Bassa	Moderata	Alta			Nessuno studio incluso

	GIUDIZI						
ALLE RISORSE NECESSARIE							
COSTO EFFICACIA	A favore del confronto	Probabilmente a favore del confronto	Non è favorevole né al confronto né all'intervento	Probabilmente a favore dell'intervento	A favore dell'intervento	Varia	Nessuno studio incluso
EQUITA'	Riduce l'equità	Probabilmente riduce l'equità	Probabilmente nessun impatto sull'equità	Probabilmente aumenta l'equità	Aumenta l'equità	Varia	Non so
ACCETTABILITÀ	No	Probabilmente no	Probabilmente si	Si		Varia	Non so
FATTIBILITÀ	No	Probabilmente no	Probabilmente si	Si		Varia	Non so

TIPO DI RACCOMANDAZIONE

Forte raccomandazione contro l'intervento	Raccomandazione condizionale contro l'intervento	Raccomandazione condizionale a favore sia dell'intervento che del confronto	Raccomandazione condizionale a favore dell'intervento	Forte raccomandazione a favore dell'intervento
○	✓	○	○	○

CONCLUSIONI

Raccomandazione

Il gruppo della linea guida del ISS, suggerisce di non usare acidi grassi poli-insaturi in bambini e adolescenti con disturbi dello spettro autistico (raccomandazione condizionata, certezza della prove molto bassa).

Giustificazione

Nessuna.

Considerazioni relative ai sottogruppi

Nessuna.

Considerazioni per l'implementazione

Nessuna.

Monitoraggio e valutazione

Nessuna.

Priorità della ricerca

Studi randomizzati controllati, con una definizione chiara degli esiti.

References for Appendix 4

1. Yui K, Koshiba M,Nakamura S,Kobayashi Y. Effects of large doses of arachidonic acid added to docosahexaenoic acid on social impairment in individuals with autism spectrum disorders: a double-blind, placebo-controlled, randomized trial. Journal of clinical psychopharmacology; 2012.

2. Bozzatello P, Brignolo E,De Grandi E,Bellino S. Supplementation with Omega-3 Fatty Acids in Psychiatric Disorders: A Review of Literature Data. Journal of Clinical Medicine; 2016.

3. Bays, HE. Safety considerations with omega-3 fatty acid therapy. American Journal of Cardiology; 2007.

4. Lord C, Elsabbagh M,Baird G,Veenstra-Vanderweele J. Autism spectrum disorder. Lancet; 2018.

5. Postorino V, Fatta LM,Sanges V,Giovagnoli G,De Peppo L,Vicari S,Mazzone L. Intellectual disability in Autism Spectrum Disorder: Investigation of prevalence in an Italian sample of children and adolescents. Research in Developmental Disabilities; 2016.

6. Lai MC, Lombardo MV,Baron-Cohen S. Autism. Lancet; 2014.

7. Buescher AV, Cidav Z Knapp M Mandell DS. Costs of autism spectrum disorders in the United Kingdom and the United States. JAMA Pediatrics; 2014.

8. Operto FF, Martino F,Rinaldi A,Cerracchio A,Salvati G,Orza M,Lembo C,Panzarino G,Di Paolantonio C,Verrotti A,Farello G,Coppola G. Long-term outcome of autistic spectrum disorder: a retrospective case study in a southern italian region. Italian Journal of Pediatrics; 2017.

9. Domellöf E, Hedlund L,Ödman P. Health-related quality of life of children and adolescents with functional disabilities in a northern Swedish county. Quality of Life Research; 2014.

10. 134, Legge. Disposizioni in materia di diagnosi, cura e abilitazione delle persone con disturbi dello spettro autistico e di assistenza alle famiglie. Gazzetta Ufficiale; 2015.

11. Abdelhamid AS, Martin N,Bridges C,Brainard JS,Wang X,Brown TJ,Hanson S,Jimoh OF,Ajabnoor SM,Deane KH,Song F,Hooper L. Polyunsaturated fatty acids for the primary and secondary prevention of cardiovascular disease. Cochrane Database of Systematic Review; 2018.

12. Johnson CR, Handen BL,Zimmer M,Sacco K. Polyunsaturated fatty acid supplementation in young children with autism. Journal of Developmental and Physical Disabilities; 2010.

13. Mankad D, Dupuis A,Smile S,Roberts W,Brian J,Lui T,Genore L,Zaghloul D,Iaboni A,Marcon PM,Anagnostou E. A randomized, placebo controlled trial of omega-3 fatty acids in the treatment of young children with autism. Mol Autism; 2015.

14. Mazahery H, Conlon CA,Beck KL,Mugridge O,Kruger MC,Stonehouse W,Camargo CA,Meyer BJ,Jones B,von Hurst PR. A randomised controlled trial of vitamin D and omega-3 long chain polyunsaturated fatty acids in the treatment of irritability and hyperactivity among children with Autism Spectrum Disorder. Journal of Steroid Biochemistry and Molecular Biology; 2018.

15. Mazahery H, Conlon C,Beck KL,Kruger MC,Stonehouse W,Camargo CA Jr,Meyer BJ,Tsang B,Mugridge O,von Hurst PR. Vitamin D and omega-3 fatty acid supplements in children with autism spectrum disorder: a study protocol for a factorial randomised, double-blind, placebo-controlled trial. Trials; 2016.

16. Moreno C, Calvo-Escalona R,Gutierrez S,Graell M,Romo J,Dorado ML,Giraldez ML,Llorente C,Arango C,Parellada M. Effect of omega-3 polyunsaturated fatty acids on oxidative stress in children and adolescents with autism spectrum disorders. European neuropsychopharmacology; 2014.

17. Parellada M, Llorente C,Calvo R,Gutierrez S,Lázaro L,Graell M,Guisasola M,Dorado ML,Boada L,Romo J,Dulin E,Sanz I,Arango C,Moreno C. Randomized trial of omega-3 for autism spectrum disorders: Effect on cell membrane composition and behavior. Eur Neuropsychopharmacol; 2017.

18. Voigt RG, Mellon MW,Katusic SK,Weaver AL,Matern D,Mellon B,Jensen CL,Barbaresi WJ. Dietary docosahexaenoic acid supplementation in children with autism. Journal of Pediatric Gastroenterology and Nutrition; 2014.

19. Yui K, Koshiba L,Nakamura S. Effects of adding large doses of arachidonic acid to docosahexaenoic acid on social impairment in individuals with autism spectrum disorders. Current psychopharmacology; 2013.

20. Amminger GP, Berger GE,Schäfer MR,Klier C,Friedrich MH,Feucht M. Omega-3 fatty acids supplementation in children with autism: A double-blind randomized, placebo-controlled pilot study. Biological Psychiatry; 2007.

21. Boone KM, Gracious B,Klebanoff MA,Rogers LK,Rausch J,Coury DL,Keim SA. Omega-3 and -6 fatty acid supplementation and sensory processing in toddlers with ASD symptomology born preterm: a randomized controlled trial. Early Human Development; 2017.

22. Boone KM, Gracious B,Klebanoff MA,Rogers LK,Rausch J,Coury DL,Keim SA. Corrigendum to "Omega-3 and -6 fatty acid supplementation and sensory processing in toddlers with ASD symptomology born preterm: a randomized controlled trial".. Early Human Development; 2018.

23. Keim SA, Gracious B,Boone KM,Klebanoff MA,Rogers LK,Rausch J,Coury DL,Sheppard KW,Husk J,Rhoda DA. ω-3 and ω-6 Fatty Acid Supplementation May Reduce Autism Symptoms Based on Parent Report in Preterm Toddlers. Journal of Nutrition; 2018.

24. Sheppard KW, Boone KM,Gracious B,Klebanoff MA,Rogers LK,Rausch J,BartlettC,Coury DL,Keim SA. Effect of Omega-3 and -6 Supplementation on Language in Preterm Toddlers Exhibiting Autism Spectrum Disorder Symptoms. Journal of Autism Developmental Disorder; 2017.

25. Meiri G, Bichovsky Y,Belmaker RH. Omega 3 fatty acid treatment in autism. Journal of Child and Adolescent Psychopharmacology; 2009.

26. Patrick L, Salik R. The effect of essential fatty acid supplementation on language development and learning skills in autism and Asperger's syndrome. Autism Asperger's Digest; 2005.

27. Politi P, Cena H,Comelli M,Marrone G,Allegri C,Emanuele E,Ucelli di Nemi S. Behavioral effects of omega-3 fatty acid supplementation in young adults with severe autism: an open label study. Archives of Medical Research; 2008.

28. Meguid NA, Atta HM,Gouda AS,Khalil RO. Role of polyunsaturated fatty acids in the management of Egyptian children with autism. Clinical Biochemistry; 2008.

29. Johnson SM, Hollander E. Evidence that eicosapentaenoic acid is effective in treating autism. Journal of Clinical Psychiatry; 2003.

30. Bent S, Bertoglio K,Ashwood P,Bostrom A,Hendren RL. A pilot randomized controlled trial of omega-3 fatty acids for autism spectrum disorder. Journal of Autism and Developmental Disorders; 2011.

31. Bent S, Hendren RL,Zandi T,Law K,Choi JE,Widjaja F,Kalb L,Nestle J,Law P. Internet-based, randomized, controlled trial of omega-3 fatty acids for hyperactivity in autism. J Am Acad Child Adolesc Psychiatry; 2014.

32. Scahill L, Aman MG,Lecavalier L,Halladay AK,Bishop SL,Bodfish JW,Grondhuis S,Jones N,Horrigan JP,Cook EH,Handen BL,King BH,Pearson DA,McCracken JT,Sullivan KA,Dawson G. Measuring repetitive behaviors as a treatment endpoint in youth with autism spectrum disorder. Autism; 2015.

33. Lecavalier L, Wood JJ,Halladay AK,Jones NE,Aman MG,Cook EH,Handen BL,King BH,Pearson DA,Hallett V,Sullivan KA,Grondhuis S,Bishop SL,Horrigan JP,Dawson G,Scahill L. Measuring anxiety as a treatment endpoint in youth with autism spectrum disorder. Journal of Autism Developmental Disorder; 2014.

34. Anagnostou E, Jones N,Huerta M,Halladay AK,Wang P,Scahill L,Horrigan JP, Kasari C,Lord C,Choi D,Sullivan K,Dawson G. Measuring social communication behaviors as a treatment endpoint in individuals with autism spectrum disorder. Autism; 2014.

35. McConachie H, Parr JR,Glod M,Hanratty J,Livingstone N,Oono IP,Robalino S,Baird G,Beresford B,Charman T,Garland D,Green J,Gringras P,Jones G,Law J, Le Couteur AS,Macdonald G,McColl EM,Morris C,Rodgers J,Simonoff E,Terwee CB,Williams K. Systematic review of tools to measure outcomes for young children with autism spectrum disorder. Health Technology Assessment; 2015.

36. Watters CA, Edmonds CM,Rosner LS,Sloss KP,Leung PS. A Cost Analysis of EPA and DHA in Fish, Supplements, and Foods. Journal of Nutrition and Food Sciences; 2012.

37. Press, R. The Omega-3 Fatty Acid Composition and Cost Analysis of Fish Oil Supplements: Fishing for the Best Deals. 2011.

38. Farmadati, Italia. .

39. Salomone E, Charman T,McConachie H,Warreyn P, Working Group 4, COST Action Enhancing the Scientific Study of Early Autism'. Prevalence and correlates of use of complementary and alternative medicine in children with autism spectrum disorder in Europe.. European Journal of Pediatrics; 2015.

40. Hock R, Kinsman A,Ortaglia A. Examining treatment adherence among parents of children with autism spectrum disorder. Disability and Health Journal; 2015.

41. Belmaker RH, Meiri G. Omega-3 Fatty Acids in the Treatment of Autism Spectrum Disorder. Comprehensive Guide to Autism; 2013.

42. Huang A, Seshadri K,Matthews TA,Ostfeld BM. Parental perspectives on use, benefits, and physician knowledge of complementary and alternative medicine in children with autistic disorder and attention-deficit/hyperactivity disorder. Journal of Alternative and Complementary Medicine; 2013.

43. Hopf KP, Madren E,Santianni KA. Use and Perceived Effectiveness of Complementary and Alternative Medicine to Treat and Manage the Symptoms of Autism in Children: A Survey of Parents in a Community Population. Journal of Alternative Complementary Medicine; 2016.

44. Nadon G, Ehrmann Feldman D,Dunn W,Gisel E. Association of Sensory Processing and Eating Problems in Children with Autism Spectrum Disorders. Autism Research and Treatment; 2011.

45. Ooi YP, Weng SJ,Jang LY,Low L,Seah J,Teo S,Ang RP,Lim CG,Liew A,Fung DS,Sung M. Omega-3 fatty acids in the management of autism spectrum disorders:

findings from an open-label pilot study in Singapore. European Journal of Clinical Nutrition; 2015.

46. Santini A, Cammarata SM,Capone G,Ianaro A,Tenore GC,Pani L,Novellino E. Nutraceuticals: opening the debate for a regulatory framework. British Journal of Clinical Pharmacology; 2018.

Appendix 4: Evidence to Decision framework for the research question used to pilot in-person the process

CHAPTER 3. A MODELING APPROACH TO DERIVE BASELINE RISK ESTIMATES FOR GRADE RECOMMENDATIONS: CONCEPTS, DEVELOPMENT, AND RESULTS OF ITS APPLICATION TO THE AMERICAN SOCIETY OF HEMATOLOGY 2019 GUIDELINES ON PREVENTION OF VENOUS THROMBOEMBOLISM IN SURGICAL HOSPITALIZED PATIENTS

Authors:

Gian Paolo Morgano,[1] Wojtek Wiercioch,[1] David R. Anderson[2], Jan L. Brożek,[1] Nancy Santesso,[1] Feng Xie,[1] Adam Cuker,[3] Robby Nieuwlaat,[1] Elie A. Akl,[4] Andrea Darzi,[1] Juan José Yepes-Nuñez,[1,5] Itziar Exteandia-Ikobaltzeta,[1] Maryam Rahman,[6] Anita Rajasekhar,[7] Frederick Rogers,[8] Kari A. O. Tikkinen,[9] Adolph J. Yates,[10] Philipp Dahm,[11,12] and Holger J. Schünemann[1,13]

[1]Department of Health Research Methods, Evidence, and Impact, McMaster University, Canada

[2]Department of Medicine, Dalhousie University, Canada

[3]Department of Medicine and Department of Pathology & Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, USA

[4]Department of Internal Medicine, American University of Beirut, Lebanon

[5]School of Medicine, Universidad de los Andes, Colombia

[6]Lillian S. Wells Department of Neurosurgery, University of Florida, USA

[7]Division of Hematology and Oncology, Department of Medicine, University of Florida, USA

[8]Trauma and Acute Care Surgery, Penn Medicine Lancaster General Health, USA

[9]Departments of Urology and Public Health, University of Helsinki and Helsinki University Hospital, Finland [10]Department of Orthopedic Surgery, University of Pittsburgh Medical Center, USA

[11]Urology Section, Minneapolis VA Health Care System, USA

[12]Department of Urology, University of Minnesota, USA

[13]Department of Medicine, McMaster University, Canada

Corresponding author:

Holger J. Schünemann,

Department of Health Research Methods, Evidence and Impact, McMaster University,

1280 Main St West, Hamilton, ON;

L8N 3Z5, Canada;

e-mail: schuneh@mcmaster.ca

Abstract

Objective: Systematic reviews of prognostic observational studies are scarce and the available estimates of associations are often not directly applicable to patient-important outcomes. The goal of this study was to develop an approach to model baseline risks for patient-important outcomes for guideline recommendations when only baseline risks for surrogate outcomes are available.

Study design: The McMaster University GRADE Centre and the ASH guideline panel for the prevention of VTE in surgical patients developed a modeling approach based on explicit assumptions about the distribution of symptoms, anatomical location, and severity of VTE events.

Results: We applied the approach to derive modeled estimates of baseline risk. These estimates were used to calculated absolute measures of anticipated effects that informed the discussion of the evidence and the formulation of 30 guideline recommendations. The approach increased transparency and reduced potential error in the decision-making process.

Conclusions: Our approach can assist guideline developers facing a lack of information about baseline risk estimates that directly apply to outcomes of interest. It also addresses potential bias of over- or underestimating absolute anticipated effects of interventions that can result from the use of surrogate data.

Keywords:

GRADE; clinical practice guidelines; baseline risk; modeling approach; surrogate estimates; absolute effects

Running title: A modeling approach to derive baseline risk estimates for GRADE recommendations

Word count (abstract): 187

Word count (main text): 3122 words

Number of images: 1

Number of tables: 3

Number of references: 32

1. Introduction

Guideline developers should transparently present the findings of their systematic reviews and meta-analysis using using structured tables.[1] These tables allow summarizing the anticipated effects of interventions using both relative and absolute measures. Absolute measures, such as the risk difference (RD) or number needed to treat (NNT), as opposed to relative risk reductions, are thought to be more easily interpreted by clinicians and required to estimate the trade-off between the health benefits and harms of interventions.[2] To generate an absolute effect, systematic reviewers multiply the pooled relative risk (generated by the meta-analysis) by the baseline risk (or control group risk) of the outcome of interest. The baseline risk that is used to calculate the RD[3], should be as specific as possible for the population of interest. However, the choice of the baseline risk can have a major influence on the resulting absolute effect size estimate and, thus, the interpretation of the clinical impact of the intervention. Using a hypothetical example of a 50% relative risk reduction of developing venous thromboembolism (VTE) with an intervention, Table 1 demonstrates the impact of varying baseline risks on the RD and its 95% confidence interval, which could lead to different guideline recommendations.

Estimates of baseline risk should be derived from well-conducted systematic reviews of directly applicable observational studies providing realistic prognostic data on the outcomes of interest.[4] Such reviews, however, are seldom available for clinical settings and their conduct presents methodological challenges.[5] Furthermore, available data often does not provide details on patient important outcome features (e.g. symptomology, specific location, and the severity of symptoms) that may be of interest to guideline panels making judgments to develop recommendations. Alternatively, estimates of baseline risk for outcome events can be drawn from event rates in the control arm of randomized controlled trials (RCTs) included in the meta-analyses. Populations enrolled in RCTs, however, typically have fewer comorbidities and better outcomes than patients encountered in clinical practice.[6] Therefore, the use of baseline risk rates derived from RCT data may underestimate both the absolute benefits and harms associated with the intervention in clinical practice.[7]

In the absence of direct prognostic evidence, i.e. baseline risks for the outcomes of interest, baseline risk estimates for surrogate outcomes could be used to derive RD for benefits and harms. For example, the composite outcome of any VTE, which includes symptomatic and asymptomatic events of VTE detected by sensitive screening tests, could be used as a surrogate for the outcome of symptomatic VTE. However, the uncritical use of surrogates poses the risk of introducing bias in the estimates of absolute treatment effects.[8] In this case, the use of any VTE could substantially overestimate the

baseline risk for symptomatic VTE, and thereby overestimate the absolute effect of an intervention.

As part of the methodology used to develop the American Society of Hematology (ASH) guidelines for the management of VTE, the McMaster University GRADE Centre and ASH developed an approach to derive modeled baseline risks for patient-important outcomes prioritized for recommendations, from baseline risk estimates for surrogate outcomes reported in observational studies. This modeling approach addresses situations of absence of direct baseline risk estimates for patient-important outcomes. In this article, we first present the concept and development process, and then describe its application in the ASH Guideline on Prevention of VTE in Surgical Hospitalized Patients.[9]

2. Methods

The guideline process was supervised by the McMaster University GRADE Centre in collaboration with ASH and followed the methods of the Guideline International Network (G-I-N)-McMaster checklist. The process included the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach to assess the certainty in the evidence and to formulate recommendations.[10-12] The lists of guideline questions and corresponding recommendations are available in the guideline publication.[9]

2.1 Prioritization of outcomes and use of health outcome descriptors

Following a structured approach,[13] a multidisciplinary panel of experts prioritized the following outcomes as critical for decision-making: *mortality, symptomatic PE of moderate severity, symptomatic proximal DVT of moderate severity, severe symptomatic distal DVT, major bleeding, and reoperation due to bleeding*.

When prioritizing patient-important outcomes, we used written outcome descriptions, known as *"marker states" or* "*health outcomes descriptors*".[14] For each outcome, a marker state offered a structured description of its symptoms, time horizon, testing and treatment, and consequences. The marker states facilitated the explicit differentiation of outcomes and reduced the risk that different panel members have different understanding of the same outcomes. These marker states are available online (https://ms.gradepro.org) while the underpinning concepts, development process, and use are described elsewhere (manuscript in preparation).

2.2 Systematic review of the evidence for baseline risks

We conducted a systematic review of baseline risk searching for published reviews of longitudinal observational studies or, if we did not identify any relevant systematic review, searching for primary studies. Although directly applicable to the surgical populations of interest, the identified baseline risk estimates often did not match the patient-important outcomes that were prioritized by the panel. In fact, most of the reported estimates pooled symptomatic and screening-detected events (e.g. 'any DVT' or

'any VTE'), and study authors rarely provided information about the anatomical distribution or severity of VTE events. We considered that the selection of the available data as surrogate baseline risks posed the risk of introducing bias in the estimates of the anticipated absolute effects. For example, a retrospective cohort of 172,320 patients undergoing abdominal surgery in the U.S.[15] based on administrative records fulfilled our selection criteria. However, study authors reported rates of any symptomatic DVT without providing information on the anatomical location and on the severity of the symptoms. The use of the available data when considering the impact of prophylaxis on symptomatic proximal DVT of moderate severity would have introduced a bias in overestimating the anticipated absolute effect as the baseline risk would be higher when combining proximal and distal DVT events of any severity.

2.3 Development of an approach to derive modeled baseline risk estimates

With the goal to reduce bias in the estimates of the anticipated absolute effects associated with the absence of direct prognosis data from the literature, we developed a modeling approach based on *explicit assumptions* about the distribution of events and consisting of quantitative adjustments on baseline risk estimates for surrogate outcomes.

2.4 Panel interaction and definitions of assumptions

We introduced to the panel the rationale and concepts of the approach. The panel endorsed the approach and contributed in the formulation of assumptions about the

distribution of VTE events based on their clinical experience as well as available literature. Assumptions were related to the anatomical location of the blood clot, symptoms, and severity. The panel chairs and methods team then prepared a diagram with could clearly outline the proposed approach and serve as a summary of the assumptions made. The use of the approach was approved by the panel and the diagram reviewed before the in-person meeting to inform the discussion of the evidence and formulation of recommendations.

2.5 Presentation of anticipated absolute effects based on modeled baseline risk estimates

We derived an Excel spreadsheet (Microsoft Excel, version 16.23) to calculate modeled baseline risk from surrogate baseline risk estimates. We then used GRADE's app GRADEpro (www.gradepro.org) to calculate the anticipated RD and its 95% C.I. boundaries from the baseline risk and the relative risk (RR) using the formula RD = baseline risk × (RR − 1). We presented the findings of our synthesis of the evidence using GRADE Evidence Profiles and Summary of Findings (SoF) tables.[4] We reported the estimates of anticipated absolute effects and their 95% C.I. boundaries using natural frequencies, in the form of estimated number of people experiencing the event per 1,000 people if receiving the intervention. When using modeled baseline risk estimates in the SoF tables, we provided the following details in explanatory footnotes[16]: the study design and sample size, the surrogate baseline risk estimate used to inform the calculation, and the assumptions applied to derive the modeled estimate.

3. Results

Figure 1 shows the summary of our approach including the results of the syntheses of the evidence, the modeled estimates, and an example of the use of modeled baseline risk estimates.

3.1 The Diagram

The diagram (Figure 1) includes nodes that each represent a possible outcome and lines that indicate how outcomes are related. The diagram can be used to guide the modeling of baseline risk for the patient-important outcomes by multiplying the reported baseline risk for surrogate outcomes by the proportions listed on the lines connecting those outcomes and the outcome of interest.

Assumptions regarding distribution of symptoms and anatomical location were defined as follows: 20% of postoperative thromboembolic events (VTE, PE, DVT) present with symptoms, 90% of VTE events are DVTs, of which 75% are proximal. To address the panel's input regarding the severity of symptomatic PEs and symptomatic proximal DVTs, we assumed that these events always present with moderate severity. We also assumed that only 25% of symptomatic distal DVTs present with severe symptoms and are critical for decision-making. In order to test the robustness of the model, we conducted a sensitivity analysis. The sensitivity analysis was performed by varying the

assumption around the ratio of proximal/distal DVTs from the base-case of 75%/25%, to 50%/50%, and 25%/75% and evaluating to what extent the results (i.e. anticipated absolute effects) would differ from the primary analysis.

3.2 Synthesis of the evidence

We conducted twenty-six new systematic reviews and updated four published systematic reviews of RCTs on effects of interventions. In order to deal with the paucity of data from RCTs on the effectiveness of IVC filters and of VTE prophylaxis in cardiac, vascular, and neurosurgical patients, we also synthetized evidence from non-randomized studies to address these questions.

We identified surgery-specific baseline risk data for all the populations of interest. Systematic reviews of risk of thrombosis were identified for urological[17,18], cardiac[19] and neurosurgical[20] procedures, whereas single cohort studies, including administrative records, were used to obtain baseline risk information for orthopedic, major general, laparoscopic cholecystectomy, trauma and gynecological surgical procedures.[15,21-27] Where deemed appropriate by the panel, we differentiated between sub-populations at lower or higher baseline risk for thrombosis. We obtained rates of symptomatic PE for the majority of surgical procedures, while available DVT rates were not directly representative of outcomes of interest.

3.3 Modeled estimates of baseline risk

Table 2 presents the estimates of baseline risks used in the guideline and calculated in the spreadsheet.

3.4 Example of the use of modeled baseline risk estimates in generating absolute effects

Using the data on DVT from the meta-analysis comparing the effects of pharmacological prophylaxis versus mechanical prophylaxis methods in patients undergoing general surgery, we show the impact of using observed baseline risk for a surrogate outcome versus modeled baseline risk estimates for a patient-important outcome on the magnitude of the anticipated absolute effects. The baseline risk estimate of 2.2% (observed risk of any symptomatic DVT reported in Spyropolous 2009[15]) was the closest surrogate available for both outcomes of symptomatic proximal and distal DVT. We applied the assumptions presented in the diagram to derive from the surrogate estimate of 2.2% the modeled estimates of 1.65% (0.022 x 0.75 = 0.0165) for risk of moderate symptomatic proximal DVT and of 0.1375% (0.022 x 0.25 x 0.25 = 0.001375) for risk of severe symptomatic distal. As shown in the Summary of Finding table (Table 3), the use of modeled estimates, as compared to using surrogate estimates, resulted in lower estimates of anticipated absolute effects. In fact, the use of estimates that are more representative of outcomes of interest mitigates the bias introduced by using surrogates that account also for events other than the outcome of interest. It is noticeable that, while the impact is small for symptomatic proximal DVTs, it is substantial for the outcome of symptomatic

distal DVT with RDs changing from 18 fewer events per 1,000 patients (95% CI: from 21 fewer to 9 fewer) to 1 fewer event per 1,000 patients (95% CI: from 1 fewer to 1 fewer). This large variation is due to the fact that modeled estimates were calculated considering the assumption that only 25% of observed distal DVTs were deemed patient important by the guideline panel. The Summary of Findings Table (Table 3) also shows the results of the sensitivity analysis demonstrating that the assumptions around relative proportions of proximal and distal DVT produced stable estimates and had little impact our overall findings.

4. Discussion

We developed and applied a modeling approach that will allow guideline developers to deal with a lack of observed baseline risk estimates for prioritized patient-important outcomes.

4.1 Strengths and limitations

There are several strengths to our work. First, modeled estimates of baseline risk can allow for better estimates of anticipated absolute effects for the populations of interest. As described through the case example of VTE, the use of surrogate baseline risk estimates could introduce bias in the anticipated absolute effects of interventions and impact judgments about the magnitude of the effects of interventions. While we cannot be certain

that modeled baseline risk estimates are fully representative of the true population baseline risk, we hypothesize that adjustments based on the clinical experience of a multidisciplinary panel informed by the literature when available can reduce error stemming from use of indirect, surrogate data without any adjustments. Second, the proposed approach increases the transparency of the guideline process and the panel's decision-making. Panel members can agree on explicit assumptions for modeling to account for indirect surrogate baseline risk estimates, as opposed to basing decision-making on implicit assumptions before making judgments about the benefits and harms of an intervention. In the absence of a transparent approach, implicit assumptions could make it difficult for users of the guideline to understand the panel's decisions. Furthermore, implicit assumptions could differ among panel members introducing unwanted variability that could further impact the decision-making process. Lastly, the proposed approach is simple to apply, without requiring substantial additional work from the guideline developers. Assumptions for modeling the distribution or severity of outcomes may be informed by data from the literature, which would be captured in a systematic review of prognosis and baseline risks, and in the absence of such data can be informed through expert panel consensus.

The approach also has limitations. Assumptions to inform the modeling of baseline risk estimates need to be accurate to closely reflect true baseline risks; they otherwise introduce bias themselves. In the absence of research evidence to inform these assumptions, the model relies on the collective expertise and input of the panel. The panel

must, therefore, include the necessary breadth of expertise and experience in the relevant clinical fields to reach reliable assumptions to inform the model. Ideally, developing these specific assumptions would be accomplished by using expert evidence.[28] The variability that will arise from using expert evidence can then be addressed through sensitivity analyses in our approach. That way, a guideline panel can determine if different assumptions about baseline risks and distribution of outcomes lead to changes in the balance of benefits and harms. We believe this transparency achieves at least two goals. First, users understand recommendations better. Second, appropriate research questions result from laying out where expert evidence is used rather than proper research evidence.

4.2 Relation to other work

The importance of basing decision-making on absolute measures of intervention effects in the context of recommendations has been established. However, guidance regarding the selection of the most appropriate baseline risk estimates is limited. In previously published VTE guidelines, Guyatt at al. outlined four alternative strategies for estimating the absolute difference in the frequency of VTE for alternative approaches to antithrombotic management.[8] The authors carefully described merit and limitations of using evidence from RCT or observational studies and of focusing on symptomatic or asymptomatic events for both effectiveness and baseline risk data. Without identifying a preferable solution, the authors suggest adopting the approach that would lead to highest-quality evidence. Tikkinen et al. conducted two systematic reviews on procedure-specific

risk of thrombosis and bleeding in patients undergoing urological surgery.[17,18] In accordance with the study protocol[5], they selected estimates of baseline risk from observational studies, preferably from those at low risk of bias, and used the median value of estimates from the studies with the lowest risk of bias. In addition, they provided an approach to adjust estimates of baseline risk for use of prophylaxis and to model risk of VTE over time. To date, the GRADE working group has not provided specific guidance on the assessment of the confidence in estimates of baseline risk and resulting estimates of absolute treatment benefits or harms, but suggests that the domains currently used to assess the certainty in the evidence of effects (risk of bias, inconsistency, indirectness, imprecision, and publication bias) can help understand issues of certainty in estimates of baseline risks.[7,29] Our work adds important conceptual thinking to these other approaches by introducing transparency and reducing error through the notion of marker states, explicit assumptions based on literature data and expert evidence, and a visual representation of the problem.[28]

4.3 Implication for research and clinical practice

Systematic reviews of prognostic studies are seldom available but are required for decision-making in the context of guidelines. Literature on baseline risk data for our example have shown poor consistency in outcome reporting. Such variability, which may be due to disagreement around the clinical relevance of asymptomatic VTE events[30,31], represents a barrier to the production of informative, large bodies of evidence focused on

patient-important outcomes. Consequently, study investigators as well as clinicians should consider the adoption and promotion of standardized core outcome sets as suggested, for example, by the COMET initiative,[32] and use descriptions such as of marker states to operationalize outcomes. Laying out the results of the sensitivity analyses we conducted will make guideline panels and clinicians more certain in their decision even if estimates of effect differ. This is because they can acknowledge both certainty and uncertainty when sensitivity analysis produces alternative decisions.

4.4 Next steps and future research

While we have applied this approach successfully in the ASH VTE guidelines, application in other clinical areas will help to further refine the approach based on panel experience and feedback. Opportunities for future research also include the assessment of the impact of modeled estimates on panel judgments, for example through a trial measuring the agreement between two groups within the same panel randomized to make judgments about estimates of anticipated absolute effects based on either surrogate or modeled baseline risk estimates.

5. Conclusions

We believe that our approach to integrating knowledge about surrogate outcomes can facilitate the production of high-quality guidelines. The use of modeled estimates will

better represent the baseline risk for the outcomes of interest and can address potential bias of over- or underestimating absolute anticipated effects of interventions. The approach also increases transparency in the process and makes the baseline risk used by guideline experts explicit during their decision-making.

Acknowledgments

The authors thank Rob Kunkle and Rob Plovnick for oversight and management of the American Society of Hematology 2019 guidelines on prevention of venous thromboembolism in surgical hospitalized patients. We also thank Rob Kunkle for critical review and comments provided for the manuscript.

Authorship Contribution

G.P.M. wrote the first draft of the manuscript and revised the manuscript based on authors' suggestions; H.J.S. conceptualized the modeling approach and guided its development; W.W. and H.J.S contributed to drafting and critical revisions of the manuscript; D.R.A. and P.D. were the chair and vice-chair of the panel, led the panel meetings, facilitated the development of assumptions and the application of the model; J.L.B., N.S., and F.X. critically reviewed the model providing methodological input; G.P.M. and H.J.S. performed initial selection of the sources of baseline risk for all guideline questions for review by the chair and vice-chair of the panel; G.P.M., W.W., I.E.I, J.J.Y.N., lead the review of the evidence used to inform the guideline and calculated modeled estimates of baseline risk; A.D. and E.A. conducted the systematic searches for prognosis data and defined its selection criteria; D.A., P.D., M.R., A.R., K.A.O.T, F.R., and A.J.Y. were members of the panel and provided expert evidence for the development of assumptions. H.J.S., A.C., R.N., and W.W. verified that the methods

described in this manuscript are consistent with what was used in the ASH 2019 guidelines on prevention of venous thromboembolism in surgical hospitalized patients. All of the authors revised the manuscript critically for important intellectual content and approved the final version submitted for publication.

Ethical approval

Not required. This study does not involve de novo patient data collection. No patient informed consent and Institutional Review Board approval have been sought.

Declarations of interest

A.C. has served as a consultant for Synergy and his institution has received research support on his behalf from Alexion, Bayer, Novo Nordisk, Pfizer, Sanofi, Spark, and Takeda; all other authors declare no competing financial interests.

Data sharing

The Excel spreadsheet used to calculate modeled baseline risk from surrogate baseline risk estimates is available upon request.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

1. Guyatt G, Oxman AD, Akl EA, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. Journal of clinical epidemiology 2011;64:383-94.

2. Deeks JJ, Higgins JP, Altman DG, Group CSM. Analysing data and undertaking meta-analyses. Cochrane handbook for systematic reviews of interventions 2019:241-84.

3. Schünemann HJ, Vist GEV, Higgins JP, et al. Interpreting results and drawing conclusions. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors) Cochrane Handbook for Systematic Reviews of Interventions version 6 (updated July 2019): Cochrane; 2019.

4. Guyatt GH, Oxman AD, Santesso N, et al. GRADE guidelines: 12. Preparing summary of findings tables-binary outcomes. Journal of clinical epidemiology 2013;66:158-72.

5. Tikkinen KA, Agarwal A, Craigie S, et al. Systematic reviews of observational studies of risk of thrombosis and bleeding in urological surgery (ROTBUS): introduction and methodology. Syst Rev 2014;3:150.

6. Kennedy-Martin T, Curtis S, Faries D, Robinson S, Johnston J. A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. Trials 2015;16:495.

7. Spencer FA, Iorio A, You J, et al. Uncertainties in baseline risk estimates and confidence in treatment effects. BMJ 2012;345:e7401.

8. Guyatt GH, Eikelboom JW, Gould MK, et al. Approach to outcome measurement in the prevention of thrombosis in surgical and medical patients: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. Chest 2012;141:e185S-e94S. doi: 10.1378/chest.11-2289.

9. Anderson DR, Morgano GP, Bennett C, et al. American Society of Hematology 2019 guidelines for management of venous thromboembolism: prevention of venous thromboembolism in surgical hospitalized patients. Blood Adv 2019;3:3898-944.

10. Atkins D, Eccles M, Flottorp S, et al. Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches The GRADE Working Group. BMC Health Serv Res 2004;4:38.

11. Balshem H, Helfand M, Schunemann HJ, et al. GRADE guidelines: 3. Rating the quality of evidence. Journal of clinical epidemiology 2011;64:401-6.

12. Schunemann HJ, Wiercioch W, Etxeandia I, et al. Guidelines 2.0: systematic development of a comprehensive checklist for a successful guideline enterprise. CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne 2014;186:E123-42.

13. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. Journal of clinical epidemiology 2011;64:395-400.

14. Schünemann HJ, Armstrong D, Degl'innocenti A, et al. A randomized multicenter trial to evaluate simple utility elicitation techniques in patients with gastroesophageal reflux disease. Med Care 2004;42:1132-42.

15. Spyropoulos AC, Hussein M, Lin J, Battleman D. Rates of symptomatic venous thromboembolism in US surgical patients: a retrospective administrative database study. Journal of thrombosis and thrombolysis 2009;28:458-64.

16. Santesso N, Carrasco-Labra A, Langendam M, et al. Improving GRADE evidence tables part 3: detailed guidance for explanatory footnotes supports creating and understanding GRADE certainty in the evidence judgments. J Clin Epidemiol 2016;74:28-39.:10.1016/j.jclinepi.2015.12.006. Epub 6 Jan 12.

17. Tikkinen KAOC, S.; Agarwal, A.; Violette, P. D.; Novara, G.; Cartwright, R.; Naspro, R.; Siemieniuk, R. A. C.; Ali, B.; Eryuzlu, L.; Geraci, J.; Winkup, J.; Yoo, D.; Gould, M. K.; Sandset, P. M.; Guyatt, G. H. Procedure-specific Risks of Thrombosis and Bleeding in Urological Cancer Surgery: Systematic Review and Meta-analysis. European urology 2018;73:242-51.

18. Tikkinen KAOC, S.; Agarwal, A.; Siemieniuk, R. A. C.; Cartwright, R.; Violette, P. D.; Novara, G.; Naspro, R.; Agbassi, C.; Ali, B.; Imam, M.; Ismaila, N.; Kam, D.; Gould, M. K.; Sandset, P. M.; Guyatt, G. H. Procedure-specific Risks of Thrombosis and Bleeding in Urological Non-cancer Surgery: Systematic Review and Meta-analysis. European urology 2018;73:236-41.

19. Ho KMB, E.; Pavey, W. Incidence of Venous Thromboembolism and Benefits and Risks of Thromboprophylaxis After Cardiac Surgery: A Systematic Review and Meta-Analysis. Journal of the American Heart Association 2015;4:e002652.

20. Glotzbecker MPB, C. M.; Wood, K. B.; Harris, M. B. Thromboembolic disease in spinal surgery: a systematic review. Spine 2009;34:291-303.

21. Gao YL, A.; Xie, Z.; Meng, Y.; Tan, J.; Lv, H.; Zhang, L.; Zhang, L.; Tang, P. The compliance of thromboprophylaxis affects the risk of venous thromboembolism in patients undergoing hip fracture surgery. SpringerPlus 2016;5:1362.

22. Jameson SS, Baker PN, Charman SC, et al. The effect of aspirin and low-molecular-weight heparin on venous thromboembolism after knee replacement: a non-randomised comparison using National Joint Registry Data. J Bone Joint Surg Br 2012;94:914-8. doi: 10.1302/0301-620X.94B7.29129.

23. Jameson SS, Charman SC, Gregg PJ, Reed MR, van der Meulen JH. The effect of aspirin and low-molecular-weight heparin on venous thromboembolism after hip replacement: a non-randomised comparison from information in the National Joint Registry. J Bone Joint Surg Br 2011;93:1465-70. doi: 10.302/0301-620X.93B11.27622.

24. Paffrath TW, A.; Lefering, R.; Simanski, C.; Bouillon, B.; Spanholtz, T.; Wutzler, S.; Maegele, M. Venous thromboembolism after severe trauma: incidence, risk factors and outcome. Injury 2010;41:97-101.

25. Peedicayil A, Weaver A, Li X, Carey E, Cliby W, Mariani A. Incidence and timing of venous thromboembolism after surgery for gynecological cancer. Gynecol Oncol 2011;121:64-9. doi: 10.1016/j.ygyno.2010.11.038. Epub Dec 22.

26. Ritch JM, Kim JH, Lewin SN, et al. Venous thromboembolism and use of prophylaxis among women undergoing laparoscopic hysterectomy. Obstet Gynecol 2011;117:1367-74. doi: 10.097/AOG.0b013e31821bdd16.

27. Stromberg JS-A, O.; Videhult, P.; Hammarqvist, F.; Sandblom, G. Incidence and risk factors for symptomatic venous thromboembolism following cholecystectomy. Langenbeck's archives of surgery 2015;400:463-9.

28. Schunemann HJ, Zhang Y, Oxman AD, Expert Evidence in Guidelines G. Distinguishing opinion from evidence in guidelines. BMJ 2019;366:l4606.

29. Schunemann HJ. Interpreting GRADE's levels of certainty or quality of the evidence: GRADE for statisticians, considering review information size or less emphasis on imprecision? J Clin Epidemiol 2016;75:6-15.

30. Bounameaux H, Agnelli G. Symptoms and clinical relevance: a dilemma for clinical trials on prevention of venous thromboembolism. Thromb Haemost 2013;109:585-8. doi: 10.1160/TH12-08-0627. Epub 2013 Jan 24.

31. Llinas A, Cardenas L. Are asymptomatic DVTs relevant?: one measure for all. Chest 2013;143:874. doi: 10.1378/chest.12-2510.

32. Gargon E. The COMET (Core Outcome Measures in Effectiveness Trials) Initiative. Maturitas 2016;91:91-2.:10.1016/j.maturitas.2016.06.007. Epub Jun 9.

Figures

Figure 1

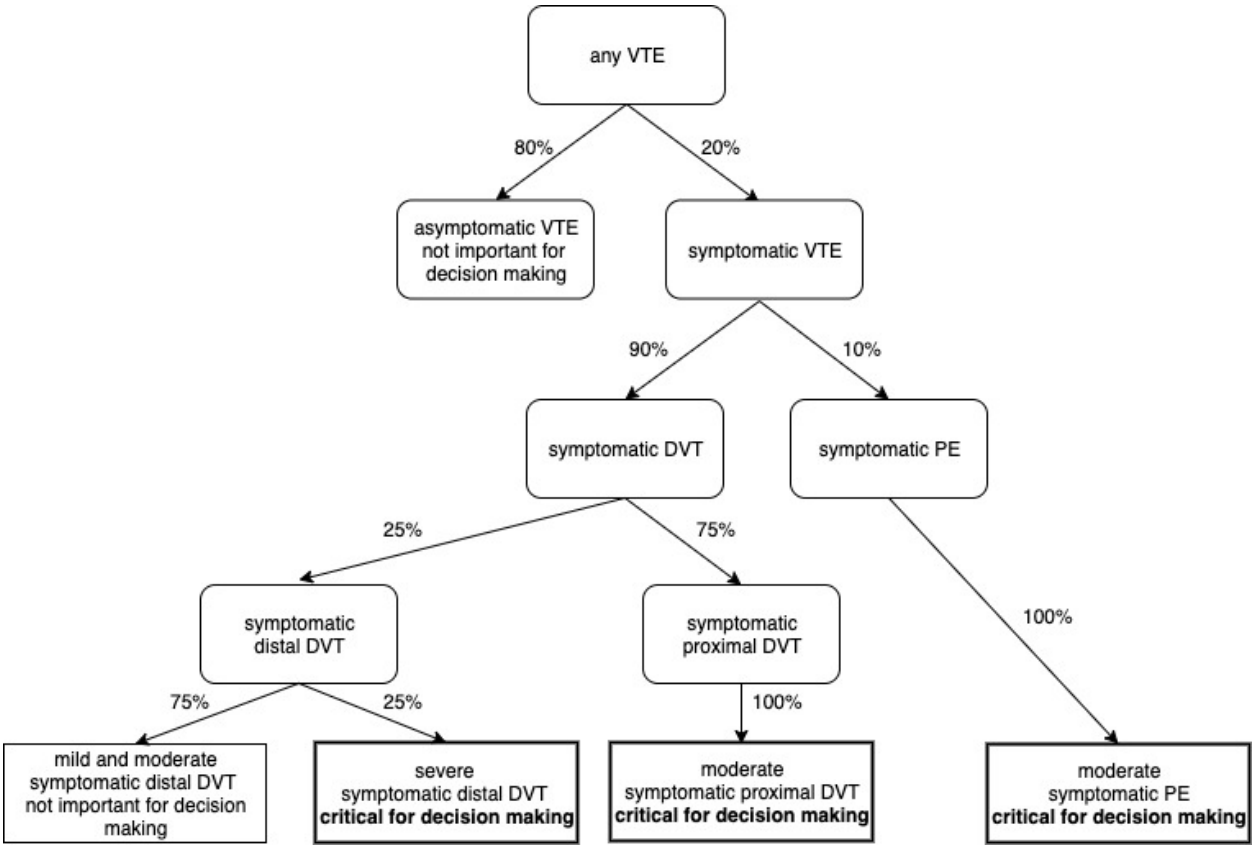


Figure 1: Diagram with the assumptions developed by the panel

Tables

Table 1.

Relative Risk Reduction for VTE using the intervention	Baseline Risk of developing VTE	Anticipated Absolute Effects or Risk Difference
50%	4.0%	20 fewer VTE events per 1,000 patients (95% CI: from 29 fewer to 3 fewer)
50%	0.8%	4 fewer VTE events per 1,000 patients (95% CI: from 6 fewer to 1 fewer)
50%	0.08%	0 fewer VTE events per 1,000 patients (95% CI: from 1 fewer to 0 fewer)

Table 1: Impact of baseline risk on anticipated absolute treatment effects

Table 2

Question group		Baseline Risk estimate							
		Observed[a]					Modeled[b]		
		VTE	PE	any DVT	proximal DVT	distal DVT	PE	proximal DVT	distal DVT
Major surgery in general	low	-	0.8%	2.2%	-	-	-	1.65%	0.13%
	high	-	1.1%	3.5%	-	-	-	2.62%	0.21%
Orthopedic surgery total hip and knee arthroplasty		-	0.56%	0.785%	-	-	-	0.588%	0.04%
Orthopedic surgery hip fracture repair		-	0.3%	-	2.5%	7.1%	-	-	1.77%
Major general surgery		-	0.8%	2.2%	-	-	-	1.65%	0.13%
Laparoscopic cholecystectomy		-	0.023 %	0.034%	-	-	-	0.0255%	0.00212%
Major neurosurgical procedures		-	0.2%	1.6%	-	-	-	1.2%	0.1%
Transurethral resection of the prostate (TURP)	low	0.2%	-	-	-	-	0.02%	0.14%	0.01%
	high	0.8%	-	-	-	-	0.08%	0.54%	0.05%
Radical prostatectomy	low	0.2%	-	-	-	-	0.02%	0.14%	0.01%
	high	0.9%	-	-	-	-	0.09%	0.61%	0.05%
Cardiac or major vascular surgery		3.62%	-	-	-	-	0.36%	2.44%	0.2%
Major trauma		-	0.68%	0.9%	-	-	-	0.67%	0.05%
	low	-	0.07%	0.91%	-	-	-	0.68%	0.05%

Major gynecological surgery	high	4.0%	-	-	-	-	0.4%	2.7%	0.22%

a. All observed events are symptomatic. When the observed estimate was directly applicable to the outcome, modeling was not necessary.

b. Modeled estimates for the outcomes deemed critical for decision making by the guideline panel.

Table 2: Modeled estimates of baseline risk used in the guideline

Table 3

Outcomes	Relative effect (95% CI)	Anticipated absolute effects	
		Risk with mechanical prophylaxis	**Risk difference with pharmacological prophylaxis**
1 - Symptomatic Proximal DVT *no modeling*	**RR 0.75** (0.11 to 5.32)	*baseline risk from control group of meta-analysis*	
		0.2%	**1 fewer per 1,000** (2 fewer to 9 more)
		surrogate baseline risk from Spyropoulos	
		2.2%	**6 fewer per 1,000** (20 fewer to 95 more)
2- Symptomatic Proximal DVT *modeling base case*	**RR 0.75** (0.11 to 5.32)	*75% of DVTs are proximal and 25% distal (of which 25% are of importance)*	
		1.7%	**4 fewer per 1,000** (15 fewer to 71 more)

Outcomes	Relative effect (95% CI)	Anticipated absolute effects	
		Risk with mechanical prophylaxis	Risk difference with pharmacological prophylaxis
3 - Symptomatic Proximal DVT *modeling sensitivity analysis*	**RR 0.75** (0.05 to 5.32)	*50% of DVTs are proximal and 50% distal (of which 25% are of importance)*	
		1.1%	**3 fewer per 1,000** (10 fewer to 48 more)
4- Symptomatic Proximal DVT *modeling sensitivity analysis*	**RR 0.75** (0.05 to 5.32)	*25% of DVTs are proximal and 75% distal (of which 25% are of importance)*	
		0.6%	**1 fewer per 1,000** (5 fewer to 24 more)
1 - Symptomatic Distal DVT *no modeling*	**RR 0.16** (0.05 to 0.58)	*baseline risk from control group of meta-analysis*	
		1.5%	**13 fewer per 1,000** (15 fewer to 6 fewer)
		surrogate baseline risk from Spyropoulos	

Outcomes	Relative effect (95% CI)	Anticipated absolute effects	
		Risk with mechanical prophylaxis	**Risk difference with pharmacological prophylaxis**
		2.2%	**18 fewer per 1,000** (21 fewer to 9 fewer)
2- Symptomatic Distal DVT *modeling base case*	**RR 0.16** (0.05 to 0.58)	*75% of DVTs are proximal and 25% distal (of which 25% are of importance)*	
		0.1%	**1 fewer per 1,000** (1 fewer to 1 fewer)
3- Symptomatic Distal DVT *modeling sensitivity analysis*	**RR 0.16** (0.05 to 0.58)	*50% of DVTs are proximal and 50% distal (of which 25% are of importance)*	
		0.3%	**2 fewer per 1,000** (3 fewer to 1 fewer)
4 -Symptomatic Distal DVT		*25% of DVTs are proximal and 75% distal (of which 25% are of importance)*	

Outcomes	Relative effect (95% CI)	Anticipated absolute effects	
		Risk with mechanical prophylaxis	Risk difference with pharmacological prophylaxis
modeling sensitivity analysis	**RR 0.16** (0.05 to 0.58)	0.4%	**3 fewer per 1,000** (4 fewer to 2 fewer)

CI: Confidence interval; **RR:** Risk ratio

Table 3: Use of modeled baseline estimates and sensitivity analysis

CHAPTER 4. DEFINING DECISION THRESHOLDS FOR JUDGMENTS ON HEALTH BENEFITS AND HARMS USING THE EVIDENCE TO DECISION FRAMEWORK: CONCEPTS, METHODS, AND PRELIMINARY RESULTS

Authors:

Gian Paolo Morgano[1,2], Lawrence Mbuagbaw[1], Nancy Santesso[1,2], Feng Xie[1], Jan L. Brozek[1,2], Antonio Bognanni[1,2], Wojtek Wiercioch[1,2], Thomas Piggott[1,2], Andrea Darzi[1,2], Miranda Langendam[3], Ilse Verstijnen[4], Elena Parmelli[5], Zuleika Saz-Parkinson[5], Pablo Alonso-Coello[6], Holger J. Schünemann[1,2,7]

Author Affiliations:

1. Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada
2. Michael G. DeGroote Cochrane Canada & McMaster GRADE Centres; Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada
3. Department of Clinical Epidemiology, Biostatistics, and Bioinformatics, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherland
4. Dutch National Health Care Institute (ZIN) (formerly named CVZ), Diemen, The Netherlands
5. Iberoamerican Cochrane Centre, CIBERESP-IIB Sant Pau, Barcelona, Spain

6. European Commission, Join Research Centre (JRC), Ispra, VA, Italy
7. Department of Medicine, McMaster University, Hamilton, ON, Canada

Corresponding author:

Holger J. Schünemann

Department of Health Research Methods, Evidence and Impact, McMaster University,

1280 Main St West, Hamilton,

ON; L8N 3Z5, Canada

Abstract

Introduction

Guideline panels using the GRADE Evidence to Decision (EtD) framework are asked to judge how substantial are the effects of interventions on desirable and undesirable patient-important health outcomes. To date, guiding principles on how to select across the available judgments of trivial, small, moderate, and large effects include considerations around the relative and absolute magnitude of the anticipated effects and the outcomes' importance. However, decision thresholds (DTs) that could help differentiating across judgments and serve as references for interpretation of findings are not yet available to EtD users.

Methods

Our objective was to design an approach that allows to derive DTs for the EtD criteria that relate to health benefits and harms and to support judgments on any single, dichotomous outcome, regardless of the underlying disease or problem. Using an iterative process and multiple stages of pilot testing, we carefully and rigorously developed a conceptual approach and survey for the calculation of DTs based on judgments by stakeholders based on ten health scenarios with different values or utilities attached to them and a variety of effect estimates. The resulting DTs are the mean of the product of the value of the health scenario and the assigned effect estimate that would make the respondent transition from one effect size to another, e.g. moderate to large. We investigated the validity of our findings by assessing the agreement between judgments made by guideline panels and the

judgments that would be suggested if applying our DTs on the same guideline data using examples from multiple guidelines.

Results

We successfully implemented the conceptual approach and obtained responses from 75 stakeholders and present the results of the preliminary analysis based on all participants recruited until July 21st, 2020. The findings support our a priori hypothesis of a difference in the DTs for trivial, small, moderate and large effects and are suggestive of a relation between raters' judgments and the joint measure of absolute effects and outcome values. Our subgroup analyses provide evidence in favor of a single set of DTs that are applicable to any health scenario, regardless of the direction of the interventions' effect and the value of the outcome.

Conclusions

The DTs for judgments on desirable and undesirable effects could be used to initiate and inform discussion, to ensure consistency in judgments across different research questions, and to promote transparency in judgments. A limitation of our work is that we are not yet able to draw conclusions from the comparison of the DTs with judgments made by guideline panel due to the limited number of eligible judgments until now, but the approach is feasible. The full survey participants and further validation of the DTs are needed to further confirm our findings.

1. Introduction

As advocated by the National Academy of Medicine of the United States (NAM, formerly known as the Institute of Medicine), the assessment of the benefits and harms of alternative care options (i.e. interventions, actions) is an essential component of any decision-making process underlying guideline recommendations.[1] This assessment should be explicit and include consideration around the magnitude and importance of health benefits and harms, and other desirable and undesirable consequences of the recommendation or decision.[2] The Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group has developed the Evidence to Decision (EtD) frameworks to help guideline developers use the evidence in a structured and transparent way and to ensure that they consider all the criteria relevant to their decisions.[3,4] To date, more than 100 organizations globally, including the World Health Organization and the National Institute for Health Care and Excellence (NICE), have adopted the GRADE approach and use EtD frameworks to develop clinical practice guidelines (www.gradeworkinggroup.org). The EtD frameworks require decision-makers to evaluate explicitly the benefits and harms of alternative care options through separate judgments based on the two following questions: *"How substantial are the desirable anticipated effects (health benefits)?"*, *"How substantial are the undesirable anticipated effects (health harms)?"*. The former judgment relies on the body of evidence for outcomes for which the effect is desirable and the latter on the evidence for outcomes for which the effect is undesirable.

To facilitate communication, the GRADE Working Group suggested expressing these judgments by assigning the health benefits or health harms of some intervention under evaluation to one of the following four categories: 'Trivial or None', 'Small', 'Moderate' and 'Large'. If it is deemed not possible to make a judgment using any of the previous options, judgments of 'Don't know' and 'Varies' are also available. In Figure 1, we illustrate the association between a hypothetical health benefit A and the four possible categories for judgments. In this example, a given health benefit A was assigned to the category of 'Trivial or None'. To be useful, however, this simplification requires that EtD users have a similar understanding of what magnitude of health benefits or health harms belong into which category and are consistent in their judgments. A similar common understanding is also important between those assigning a category and those interpreting the meaning of a category that is communicated to them (i.e. "imagining" how substantial is an effect based on the category). This can be achieved only when people make similar judgments. To direct EtD users on how to make these judgments appropriately, the GRADE Working Group has produced guidance articles that include the description of the underpinning concepts and examples of judgments based on clinical scenarios.[4,5] For continuous outcomes, EtD users are advised to revert to statistical notions such as Cohen's standardized effect sizes or the Minimal Important Difference (MID) to interpret the magnitude of effects.[6,7] However, empirical data supporting judgments on health benefits and harms for dichotomous outcomes are not yet available for the EtDs.

1.1 Concept of Decision Thresholds for EtD judgments on health benefits and harms

Despite the popular use of decision-thresholds to support decision-making in various fields of healthcare research[8-10] and its adoption in some context of the GRADE approach[11,12], their use for EtD judgments about health benefits and harms is not yet established. Explicit Decision-Thresholds (DTs), providing indication which could be the appropriate judgment for a given scenario, might have the potential to support panels of decision-makers in their work, facilitate common understanding, and promote consistency and transparency in judgments. The results will inform suggestions about the most appropriate judgment for a given health benefit or health harm in Appendix 1 by a panel of decision-makers (Figure 1).

1.2 Objectives of the study

Our objective was to derive DTs for EtD judgments on health benefits and harms to support judgments on any single, dichotomous outcome, regardless of the underlying disease or problem. We will base the calculation of the DTs on judgments on health benefits and health harms that we collected through a survey to stakeholders. To verify the validity of our findings, we used data from existing guidelines to compare our DTs to judgments made by guideline panels.[13-17] In this article, we first illustrate the rigorous development of the methods proposed to derive the DTs and then discuss preliminary findings based on the pre-planned interim analysis of data.

2. Methods

2.1 Approach to compare health benefits/harms to the DTs

To estimate the DTs and compare these to the health benefits and harms being evaluated, we sought to determine a quantitative link between health benefits and harms and the four categories for EtD judgments: trivial to none, small, moderate and large. In developing our approach, we considered that judgments on how substantial are health benefits and harms should be influenced by the size of intervention's effects on each outcome (e.g. the number of people who would benefit or be harmed) as well as the value assigned to those outcomes by the people being affected.[5] Under this assumption, we used the product of these two factors to quantify the magnitude of a given health benefits and harms. We provide an example of how the DTs would allow assigning a given magnitude of health benefits or harms to one of the EtD categories in Appendix 1 (Figure 2).

2.2 Survey

Layout

We developed a conceptual approach and a survey to collect empirical data about the association between magnitudes of health benefits or harms and EtD judgment (www.surveymonkey.com). We separated the survey in three sections: introduction and example, iterative ratings of magnitudes of health benefits/harms using case-scenarios, questions about respondent demographics.

Target population

The target population of the survey included clinicians, epidemiologists, health research methodologists, experts in Health Technology Assessment (HTA), and members of guideline working groups. Prior knowledge of the GRADE approach and experience with the EtD framework was not required for participation.

Development of the conceptual approach and survey piloting

Using an iterative process and multiple stages of pilot testing, we carefully and rigorously developed a conceptual approach and survey for the calculation of DTs based on judgments by stakeholders. To ensure usability and clarity of the survey across respondents having different background or expertise, study co-authors as well as complementary representatives of the target population (n=15) participated to the pilot-test. Comments on three iterations of the survey were collected either electronically through SurveyMonkey or voice recordings and discussed by study authors during meetings.

Development of the conceptual approach, case-scenario and selection of outcomes

To elicit judgments by survey respondents, we developed case-scenarios (see below) providing information about the potential health effects (benefits or harms) of an intervention on an outcome. Since we sought to derive DTs applicable to judgments on any outcome, we aimed to collect data applicable to outcomes having different values.

We deemed this essential because the same person should judge the same magnitude of anticipated effects differently depending on the value of the outcome under consideration. For example, a person could consider the health benefit of 15 deaths fewer per 1000 'Large' but consider the same magnitude of benefit (15 fewer per 1000) as 'Trivial or None' if related to the reduction of a mild, short-term, outcome. We selected the following outcome for the development of case-scenarios: death, major ischemic stroke, pulmonary embolism of moderate severity, moderate diarrhea, and mild nausea/vomiting. We developed two case-scenarios for each outcome, one descriptive of desirable and another one descriptive of undesirable health effects of an intervention based on the direction of the effect on the outcome.

Structure of case-scenarios

We drafted detailed case-scenarios (Appendix 1, Figure 3) to reduce the risk of introducing error that could have resulted if survey respondents had a different understanding of interventions' effect, and outcomes' value or key characteristics. Each case-scenario included a GRADE Summary of Finding (SoF) table, the value of the outcome, and a description of the key attributes (health outcome descriptor, HOD) of the outcome under consideration.[18,19] The SoF table provided information about the PICO (population, intervention, comparator, outcome), the relative and absolute anticipated effects of the intervention, and the certainty in the evidence.[20,21] To avoid interpretation of uncertainty around the point estimate, we rated the certainty of evidence (CoE) as 'HIGH'

and did not provide any 95% confidence interval inherent the effects' estimates and the value of the outcome. For each outcome, we provided HODs presenting the symptoms, testing and treatment, and the long-term health consequences associated with that outcome together with their duration (i.e. the time horizon).[19] An example HOD is available in Appendix 1 (Box 1). The value of the outcome, also referred to as 'health utility' in health economics, was expressed through a measure on a scale from 0 (being dead) and 1 (perfect health) which meant that outcomes with higher value were valued closer to perfect health as compared to outcomes with a lower value.[22] Estimates of effects were drawn from existing systematic reviews [17,23-30], while, where available, outcomes' values and HODs were obtained from existing guidelines.[19] To also include outcomes entailing a small reduction in health that were not available from existing guidelines, we developed new HODs for the outcomes of moderate diarrhea, and mild nausea/vomiting. What differed between the benefit and harm outcome was the effect (increase versus decrease) but not the health outcome descriptor.

Sample size calculation

We based our sample size calculation on the data collected during pilot-testing (n=15 participants). Based on this data, we computed the mean thresholds T1, T2, T3 for each outcome separately and estimated that we need to recruit 1406 survey respondents to demonstrate a difference of 15% of the mean with non-overlapping 95% confidence

intervals. Further information about the sample size calculation is available in Appendix 1 (Table 3).

Collection of judgments of health benefits and health harms

We employed an iterative approach for the collection of judgments, randomizing survey participants to 4 different case-scenarios. For each case-scenario, we first asked survey participants to consider interventions' effects and the value of the outcome to judge how substantial the described health benefits or health harms are (Appendix 1, Figure 4). Then, we asked survey participants to indicate the lower and upper bound for the ranges of magnitudes of ARD that they would have associated to the judgments of 'Small' and 'Moderate' (Appendix 1, Figure 5). We did not inquire about ranges for judgments of 'Trivial or None' and of 'Large' since any estimate below the lower bound for 'Small' was considered as 'Trivial or None', and any estimate above the upper bound of 'Moderate' was considered as 'Large'.

Ethics

After review, the Hamilton Integrated Research Ethics Board determined that as a quality improvement project, this study was exempt from formal ethics review. We informed respondents of this decision and the anonymous nature of the study.

Distribution

We distributed the survey through colleagues, the research group's e-mail lists including that of the Cochrane Collaboration, Guidelines International Network (G-I-N), and of the Global Evidence Synthesis Initiative (GESI). Twitter, LinkedIn, and other social medial platforms were also used for broad distribution.

2.3 Statistical analysis

Calculation of thresholds from survey ratings

We used the ranges of ARD for judgments of 'Small' and 'Moderate' collected through the survey to calculate the thresholds associated with each rating. The thresholds were derived through the product between each range boundary and the reduction in value from perfect health (1 - outcome's value) for the outcome associated with that rating. The thresholds were calculated as follows: $\text{Threshold}_{\text{Trivial/Small}}$ equal to the product considering the lower bound for judgments of 'Small', $\text{Threshold}_{\text{Small/Moderate}}$ equal to the mean between the product considering the upper bound for judgments of 'Small' and the product considering the lower bound for judgments of 'Moderate', and $\text{Threshold}_{\text{Moderate/Large}}$ equal to the smallest number larger than the product considering the upper bound for judgments of 'Moderate'. We provide an example of the calculations of the threshold associated with a single rating in Appendix 1 (Tables 1 and 2).

Calculation of DTs

We calculated the DTs T1, T2, and T3, where T1=$DT_{Trivial/Small}$, T2=$DT_{Small/Moderate}$, and T3=$DT_{Moderate/Large}$, as the weighted mean of the corresponding thresholds derived from survey ratings. We used a weighted mean to account for multiple ratings from the same survey respondent.

Descriptive statistics

We used frequencies and percentages to describe characteristics of survey respondents. For each DT, we calculated mean, standard deviation (SD), and 95% confidence intervals (C.I.).

Primary analysis

Since each participant contributed data to each threshold, we employed a paired samples t-test to assess if the DTs were different (T1≠T2≠T3). Our a priori hypothesis was that there was a difference between the DTs and no difference between benefits and harms. All statistical tests were performed at the 0.05 level of significance.

Within-participant analyses

We used an independent samples t-test to assess whether, depending on the direction of intervention's effects, the same participant would have provided different thresholds.

Subgroup analyses

We defined subgroups of ratings depending on the direction of interventions' effects (case-scenarios descriptive of health benefits vs. case-scenarios descriptive of health harms), the value of the outcome, and the characteristics of participants (training in epidemiology, familiarity with the EtD framework, previous participation in guideline development groups). Our a priori hypotheses were that there would be differences between the DTs within subgroups ($T1 \neq T2 \neq T3$) and no differences if comparing each DT between subgroups (given subgroups a and b, $T1_a = T1_b$; $T2_a = T2_b$; $T3_a = T3_b$). We employed a paired samples t-test to assess if the DTs were different within subgroups and an ANOVA to examine whether each DT was different between subgroups.

Sensitivity analyses

We expected that, given the complexity of the topic, some responses might not be internally incoherent or outliers. We defined a threshold as incoherent if T1>T2 OR T2>T3. We defined thresholds as outliers if they fell more than 3 interquartile ranges below the first quartile or above the third quartile. We verified if the primary analysis would differ if excluding incoherent thresholds or data outliers. The a priori hypothesis for the sensitivity analyses was the same as for the primary.

Assessment of order effects

We conducted an ANOVA analysis to examine whether participants randomized to a case-scenario descriptive of a low-value outcome (outcome value <0.5) in the first iteration provided different thresholds as compared to participants who were randomized to a high-value outcome first. Similarly, we examined whether participants who provided a judgment of 'Small' in the first iteration provided different thresholds as compared to participants who provided a judgment of 'Large' in the first iteration. Our a priori hypothesis was of no differences if comparing each DT between these groups.

2.4 Assessment of validity of DTs through comparison with judgments from ASH VTE guidelines

We investigated the validity of our findings by assessing the agreement between judgments made by guideline panels and the judgments that would be suggested if applying our DTs on the guideline data. We purposively selected judgments from five guidelines on the management of Venous Thromboembolism (VTE) that were developed using the GRADE approach in a collaboration between the MacGRADE Centre and the American Society of Hematology (ASH).[13-17] We used frequencies and percentages to describe the agreement. The criteria used to select judgments are available in Appendix 1 (Box 2). We employed SPSS v26 (IBM Corp., Armonk, NY) to conduct all statistical analyses.

3. Results

We successfully designed a survey to measure DTs using ten health care scenarios. The adoption of structured presentation formats, such as the GRADE SoF tables and HOD, allowed us to lay out the data relevant for making judgments in a manner that was clear and accessible to stakeholders with different background. Through an iterative approach employing randomization, we were able to collect data based on different types of interventions and outcomes without exposing the survey participants to an overwhelming exercise.

3.1 Statistical analysis

Descriptive statistics

We planned a preliminary analysis based on survey data collected until July 21st, 2020. Our dissemination strategy allowed recruitment of 75 stakeholders who contributed a total of 295 ratings. Fifty-six survey participants had a background in research (74.6%) and 36 were healthcare professionals (50.6%). Thirty-four respondents (45.3%) were members of academia. Other major groups were participants from HTA organizations and professional societies (13.3% and 18.6%, respectively). Participants were equally randomized to case-scenarios descriptive of desirable and undesirable health effects (144/295, 49%; 151/295, 51%, respectively) and completed the entire exercise in the

majority of cases (68/75, 90.7%). Detailed descriptive characteristics of survey respondents and ratings are shown in Tables 1 and 2, respectively.

Characteristic [a]	**Respondents, n = 75**
Background [a]	n (%)
Clinical/Health Professional	38 (50.6)
Policymaking	6 (8.0)
Research	56 (74.6)
Teaching	18 (24.0)
Administrative	3 (4.0)
Patient representative	2 (2.6)
Other	3 (4.0)
Degree [a]	
Degree in Nursing (RN)	1 (1.3)
Medical School (MD)	30 (4.0)
Master of Sciences (MSc)	17 (22.6)
Master of Public Health (MPH)	9 (0.12)
Doctor of Philosophy (PhD)	25 (33.3)
None	2 (2.6)
Other	5 (6.6)

Formal Training in health research methodology/epidemiology/biostatistics	
Never completed	12 (16.0)
Completed some form of formal training but do not have a graduate degree	30 (40.0)
Earned a MSc degree	16 (21.3)
Earned a PhD degree	16 (21.3)
Not available	1 (1.4)
Organization [a]	
Cochrane collaboration	13 (17.3)
GRADE Working Group	16 (21.3)
World Health Organization	1 (1.4)
Guidelines International Network (G-I-N)	-
Health Technology Assessment (HTA) organization	10 (13.3)
Academia	34 (45.3)
Professional society	14 (18.6)
Familiarity with the Evidence to Decision framework	
Not at all familiar	5 (6.6)
Not so familiar	9 (12.0)
Somewhat familiar	16 (21.3)
Very familiar	30 (40.0)

Extremely familiar	8 (10.6)
Not available	7 (9.5)
Previous participation in guideline development groups	
Yes	52 (69.3)
No	18 (24.0)
Not available	5 (6.6)
Primary role in the guideline development group [a]	
Clinical Chair	5 (6.6)
Chair for methods	15 (19.8)
Guideline methodologist	29 (38.6)
Panel member	15 (19.8)
Topic or content expert	7 (9.5)
Patient representative	2 (2.6)
Systematic review author	26 (34.6)
Expert in Health Technology Assessment	3 (4.0)

Values represent the number and in parentheses the percentage.

[a] Percentages do not add up to 100 because respondents could choose more than one option.

Table 1: Characteristics of survey respondents

Characteristics of ratings collected through the survey	n (%)
Total number of ratings collected	295
Missing data (expected ratings - collected ratings/expected ratings)	17/312 (0.054)[a]
randomized to a scenario showing desirable effects	144/295 (49)
randomized to a scenario showing undesirable effects	151/295 (51)
randomized to the outcome of death	73/295 (25)
randomized to the outcome of major stroke	66/295 (22)
randomized to the outcome of pulmonary embolism	55/295 (19)
randomized to the outcome of moderate diarrhea	63/295 (21)
based on the outcome of mild nausea/vomiting	38/295 (13)

a. 73 participants were randomized to 4 case-scenarios, 2 were mistakenly randomized to 10.

Table 2: Descriptive statistics of survey ratings

Table 3 describes the estimates of DTs that were derived from survey ratings through the joint measure of absolute effects and outcome values. For example, an outcome valued as 0.8, these thresholds would indicate that the effect of an intervention preventing 30 events of that outcome per 1000 should be categorized as trivial (since 0.03*(1-0.8)) =0.006 is smaller than T1). More details about the calculation of the DTs are available in Appendix 1 (Table 1).

Decision Threshold			95% Confidence Interval	
	Estimate	Std. Deviation	Lower Bound	Upper Bound
T1: Trivial/Small	0.0165	0.0467	0.0059	0.0271
T2: Small/Moderate	0.0312	0.0601	0.0176	0.0448
T3: Moderate/Large	0.0577	0.0781	0.0400	0.0754

Table 3: Estimates of DTs

Primary analysis

Our analysis showed a difference in the estimates between T1 and T2 (mean difference [MD] -0.0147; 95% CI -0.0201 to -0.0093; $p<0.001$) and T2 and T3 (mean difference [MD] -0.0264; 95% CI -0.0544 to -0.0062; $p<0.001$).

Within-participant analyses

The analyses showed that at a respondent level there was no difference between DTs derived from judgments on benefits and from those on harms: $T1_{benefit}=T1_{harms}$ (mean difference [MD] -0.0040 ; 95% CI -0.0195 to 0.0116 ; $p=0.615$) ; $T2_{benefits}=T2_{harms}$ (mean difference [MD] -0.0124; 95% CI -0.0313 to 0.0064 ; $p=0.196$); $T3_{benefit}=T3_{harms}$ (mean difference [MD] -0.0209; 95% CI -0.0451 to 0.0033; $p=0.090$).

Subgroup analyses

Our subgroup analyses showed a difference in the estimates between T1 and T2, and T2 and T3 also in DTs derived from subgroup of ratings identified by outcome, direction of interventions' effects, and prior participation to guideline development groups. No difference was observed in the estimates between T1 and T2 in those with no experience with the EtD (mean difference [MD] -0.0046; 95% CI -0.0100 to 0.0006; p=0.810) and between T2 and T3 in those who had no training in epidemiology (mean difference [MD] -0.0056; 95% CI -0.0218 to 0.0106; p=0.483).

Sensitivity analyses

The findings of the sensitivity analyses conducted by excluding raters who provided incoherent thresholds (n=3; T1/T2 mean difference [MD] -0.0143; 95% CI -0.0192 to -0.0094; p<0.001; T2/T3 mean difference [MD] -0.0291; 95% CI -0.0417 to -0.0165; p<0.001) or who were presumed outliers (n=10; T1/T2 mean difference [MD] -0.0096 ; 95% CI -0.0113 to -0.0078; p<0.001; T2/T3 mean difference [MD] -0.0194; 95% CI -0.0240 to -0.0148; p<0.001) were similar to that of the primary analysis.

Assessment of order effects

The analyses suggest no difference between DTs derived from participants who evaluated a high-value outcome (i.e. moderate diarrhea) in the first iteration compared to those who evaluated a low-value outcome (i.e. death) first. Similarly, there was no difference in the DTs depending on whether the first judgment made was 'Small' or 'Large'.

3.6 Assessment of validity of DTs through comparison with judgments from ASH VTE guidelines

We analysed 208 EtD judgments on desirable and undesirable effects made by five ASH guideline panels.[13-17] We identified 53 judgments (53/208, 25%) eligible for inclusion because they inequivocally related to a single outcome and a single judgment (as opposed to a judgment across several outcomes). Of these, 38 were of 'Trivial or None' effects, 9 of 'Small' effects; 4 of 'Moderate' effects, and 2 of 'Large'. The analysis showed an overall agreement of 71.6% (38/53) between judgments made by guideline panels and the judgments that would be suggested by applying our DTs. A subgroup analysis (Figure 2) supports higher agreement on judgments of 'Trivial or None' (37/38, 97.3%) as compared to judgments of 'Small', 'Moderate', and 'Large' effects (0/9, 0%; 0/4, 0%; 1/2, 50%, respectively). More details of this analysis are reported in Appendix 1 (Table 5).

4. Discussion

In this study, we describe concepts and methods that aim to identify DTs for EtD judgments on health benefits and harms through rigorous research. We present findings from a pre-planned preliminary analysis based on ratings from the first 75 survey respondents. We conducted this analysis to fulfill the requirements of the thesis and to fully evaluate the feasibility of the approach. We believe we achieved the latter given the results indicated that participants provided distinct DTs and that they comprehended the task. Furthermore, our results suggest that this approach appears valid and that we will be able to explore the DTs fully with the full sample size.

4.1 Main findings and clinical interpretation

Participants were able to complete the exercise in the majority of cases. Only 7 out of 75 did not complete the survey. There were only 17 out of 312 expected ratings missing from these participants indicating that the approach to obtaining the DTs is feasible. This is true for people of varying backgrounds and educational level. We found that the DTs differed significantly and as expected between the scenarios that were provided. Although an important simplification of utility theory, we developed this survey approach to support guideline panels in their judgments in a way that would allow for the necessary pragmatism in many guideline development scenarios. Indeed, our preliminary findings suggest that the DTs can be used to support panels. For example, if considering an outcome valued as 0 (death), an increase/decrease in absolute risk of less than 1.65%

(95% C.I. 0.59 to 2.71) could be judged as a 'Trivial or None' effects, between 1.65% (95% C.I. 0.59 to 2.71) and 3.12% (95% C.I. 1.76 to 4.48) as 'Small' effects, between 3.12% (95% C.I. 1.76 to 4.48) and 5.77% (95% C.I. 4.00 to 7.54) as 'Moderate' effects, and more than 5.77% (95% C.I. 4.00 to 7.54) as 'Large' effects. If considering an outcome being valued 0.42 (e.g. pulmonary embolism of moderate severity), the suggested judgments according to the DTs would be: an increase/decrease in absolute risk is of less than 2.84% (95% C.I. 1.01 to 4.67) for 'Trivial or None', between 2.84% (95% C.I. 1.01 to 4.67) and 5.37% (95% C.I. 3.03 to 7.72) for 'Small', if between 5.37% (95% C.I. 3.03 to 7.72) and 9.94% (95% C.I. 6.89 to 13.00) for 'Moderate', and more than 9.94% (95% C.I. 6.89 to 13.00) for 'Large' effects, respectively. Guideline panels using the GRADE EtDs often ask what are trivial, small, moderate or large effects. Our preliminary analysis of the conceptual approach in the survey can provide guidance can be given for their judgments by suggesting possible ranges of effects for a given value of the outcome as presented in the two scenarios above. Once recruitment for our survey is completed and the estimates available, we will be able to utilize the DTs to give guidance to panels based on a simple calculator that requires input of the value of the outcome and our results (Figure 2).

4.2 Strengths and limitations

There are several strengths to our work. First, we based the calculation of the DTs on empirical data. Therefore, our DTs are informed by judgments that are similar to that of a guideline panel using the EtD framework. Second, we developed structured case-scenarios to present survey participants with the information relevant to make their judgments. The case-scenarios included effective presentation formats such as the GRADE SoF tables and HOD that can reduce the risk of error due to an inadequate presentation of data.[19,20] Third, we employed a randomization process that ensured that case-scenarios were equally distributed across survey participants and that ratings were collected through judgments on outcomes having different values. Lastly, the proposed approach is simple to apply, and does not require specific knowledge if not the ability to calculate the product between RD and the reduction in value associated with the outcome. Our work has also limitations. First, SurveyMonkey did not provide all the flexibility in developing the survey. As a consequence, we were not able to create a more appealing layout to directly elicit the thresholds and opted for collecting the ranges of estimates using the slider. Similarly, we were not able to prevent incoherent ratings. The implementation of these features would have probably simplified the survey and further reduced errors in ratings. However, the number of incoherent ratings was low which indirectly validates our approach as discussed above. Second, we acknowledge that the survey represents a quite challenging exercise and that this could impact test-retest reliability and applicability of the survey results. There are additional limitations to the

assumptions about the expected utility theory which are not met with our approach to determining values, but many guideline developers accept this limitation to not be the perfect be the enemy of the good.

4.3 Interpretation of statistical results

The results of our primary analysis support our a priori hypothesis of a difference in the DTs (T1≠T2≠T3). These findings are suggestive of a relation between raters' judgments and the joint measure of absolute effects and outcome values and they are consistent with our claim that DTs do not differ between health benefits and harms. Our subgroup analyses provided evidence in favor of a single set of DTs that are applicable to any scenario, regardless of the direction of interventions' effect and the value of the outcome. In fact, both the primary and the within-participant analyses showed how there is a no significant difference between the thresholds derived from ratings on case-scenarios descriptive of desirable effects and the thresholds derived from ratings on case-scenarios descriptive of undesirable effects. Similarly, our primary analysis showed a no difference between ratings based on outcomes having different value to patients. In building our dataset, we have accounted for these differences by collecting thresholds not only having very high or very low value, but also intermediate value (e.g. pulmonary embolism of moderate severity, value of 0.42). This allowed us to derive thresholds that might be applicable to any outcome regardless of the assigned value as suggested from the statistical results of our primary analysis.

Confirming our a priori hypothesis, the data analyses suggest that participants are able to provide thresholds for EtD judgments with a small number of participants providing incoherent thresholds (3/75, 4.0%). This strongly corroborates that survey respondents correctly interpreted the exercise and provided informative data. It also suggests that DTs could be used by guideline panels making recommendations. The only instance of no difference in thresholds was observed from data by participants without training in epidemiology or familiarity with the EtD. These results could be due to the low number of survey respondents with these characteristics (16% and 5%, respectively) or because participants who are less familiar with the concepts presented in the survey might have found the exercise challenging and provided inconsistent ratings. We acknowledge the high standard deviation around the mean DTs (coefficient of variation T1=2.83, T2=1.92, and T3=1.35, primary analysis). There are several possible explanations. First, it may simply be a result of our current analysis being an interim analysis (see sample size calculation). Second, despite not ideal and not suggested in the survey, respondents might have used their own values about the impact of the outcome on patient health (outcome value). This could have resulted in participants providing thresholds based on slightly different case-scenarios and introduced variability in their judgments. Third, a degree of variability in judgments is typically observed in guideline panels and it is due to several factors including clinical experience, personal beliefs, and decision-making style. Our investigation showed an overall agreement of 71.6% (38/53) between judgments made by guideline panels and the judgments that would be suggested if applying our DTs. We cannot draw firm conclusions from this data considering that the large majority of included judgments were

of 'Trivial or None' (38/53, 71.6%) and that the results of this analysis were mainly driven by agreement on this judgment (37/38, 97.3%) but these results are encouraging.

4.4 Relation to other work

Our endeavor expands the research on the use of decision thresholds within the GRADE methodology. Our work with Hultcrantz et al.[12] suggests using clinical decision thresholds to allow appropriate ratings of the certainty of the evidence but there is not empricial data. Furthermore, it focuses on the construct of CoE and targets different degrees of contextualization (systematic reviews, health technology assessment, clinical practice guidelines), while we address judgments on the magnitude of effects and made by users of the EtD framework. Another difference is that we provide decision-makers with estimates of thresholds derived from empirical data as opposed to asking them such estimates. However, we believe that our findings can benefit the ongoing activities of the GRADE Working Group on rating of the certainty, especially in partly contextualized settings where the certainty is rated in a specific magnitude of effects. The joint consideration of the estimate of effect and outcome's importance has been already adopted in another effort of the GRADE Working Group. In a concept paper, Alper et al. aim to define the certainty in the net benefit[31] and suggest calculating the net effect of an intervention by combining importance-adjusted effect estimates calculated from different outcomes. While this strategy is appealing and would allow us to apply our research to EtD judgments on the trade-off between benefits and harms, further research is needed to

establish if the estimates to be combined are independent and not correlated with each other. Other quantitative approaches to assess the benefits, harms, and net benefit associated with treatments are available in the literature[32], but none aims to characterize the magnitude of effects into categories (i.e 'Trivial or None', 'Small', 'Moderate', 'Large') as needed to make judgments using the EtD framework.

4.5 Implication for research

Our preliminary results support the feasibility of pursuing research based on complex conceptual thinking and for which the collection of empirical data might represent a barrier. The use of information technology can facilitate the development, testing, and social media dissemination of surveys that can successfully serve sophisticated data collection across heterogenous target populations.

4.6 Implication for clinical practice and decision-makers

Our approach emphasizes the importance of including the value assigned to outcomes by patients as an essential component of any clinical judgments or policy decisions in healthcare.

4.7 Next steps and future research

Next steps of this research include the recruitment of the required sample of survey participants, the assessment of the test-retest reliability of the survey, and the comparison

of our DTs with a larger number of selected judgments made by guideline panels. Opportunities for future research include the replication of the study using case-scenarios based on different settings or different outcomes as well as to investigate the use of DTs for EtD judgments on desirable and undesirable consequences of health interventions based on the effects on multiple outcomes.

5. Conclusions

We believe that DTs for judgments on desirable and undesirable health effects can be useful to decision-makers using the EtD framework. The DTs could be used to initiate and inform discussion and be integrated in GRADEpro, to ensure consistency in judgments across different research questions, and to promote transparency in judgments. The findings based on the preliminary analysis of the data support our hypothesis that the DTs can help discriminate between the judgments. However, a large sample of survey participants and further validations of the DTs are needed to draw informed conclusions.

References

1. Institute of Medicine Committee on Standards for Developing Trustworthy Clinical Practice G. Clinical Practice Guidelines We Can Trust. Washington (DC): National Academies Press (US); 2011.

2. Schünemann HJ. Guidelines 2.0: do no net harm-the future of practice guideline development in asthma and other diseases. Curr Allergy Asthma Rep 2011;11:261-8.

3. Atkins D, Eccles M, Flottorp S, et al. Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches The GRADE Working Group. BMC Health Serv Res 2004;4:38. doi: 10.1186/472-6963-4-38.

4. Alonso-Coello P, Schunemann HJ, Moberg J, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction. BMJ (Clinical research ed) 2016;353:i2016.

5. Alonso-Coello P, Oxman AD, Moberg J, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 2: Clinical practice guidelines. BMJ (Clinical research ed) 2016;353:i2089.

6. Faraone SV. Interpreting estimates of treatment effects: implications for managed care. P T 2008;33:700-11.

7. Johnston BC, Ebrahim S, Carrasco-Labra A, et al. Minimally important difference estimates and methods: a protocol. BMJ Open 2015;5:e007953.

8. Marcucci M, Sinclair JC. A generalised model for individualising a treatment recommendation based on group-level evidence from randomised clinical trials. BMJ Open 2013;3.

9. McCabe C, Claxton K, Culyer AJ. The NICE cost-effectiveness threshold: what it is and what that means. Pharmacoeconomics 2008;26:733-44.

10. Pauker SG, Kassirer JP. Therapeutic decision making: a cost-benefit analysis. N Engl J Med 1975;293:229-34.

11. Hultcrantz M, Mustafa RA, Leeflang MMG, et al. Defining ranges for certainty ratings of diagnostic accuracy: a GRADE concept paper. J Clin Epidemiol 2020;117:138-48.

12. Hultcrantz M, Rind D, Akl EA, et al. The GRADE Working Group clarifies the construct of certainty of evidence. J Clin Epidemiol 2017;87:4-13.

13. Cuker A, Arepally GM, Chong BH, et al. American Society of Hematology 2018 guidelines for management of venous thromboembolism: heparin-induced thrombocytopenia. Blood Adv 2018;2:3360-92.

14. Witt DM, Nieuwlaat R, Clark NP, et al. American Society of Hematology 2018 guidelines for management of venous thromboembolism: optimal management of anticoagulation therapy. Blood Adv 2018;2:3257-91.

15. Schünemann HJ, Cushman M, Burnett AE, et al. American Society of Hematology 2018 guidelines for management of venous thromboembolism: prophylaxis for hospitalized and nonhospitalized medical patients. Blood Adv 2018;2:3198-225.

16. Monagle P, Cuello CA, Augustine C, et al. American Society of Hematology 2018 Guidelines for management of venous thromboembolism: treatment of pediatric venous thromboembolism. Blood Adv 2018;2:3292-316.

17. Anderson DR, Morgano GP, Bennett C, et al. American Society of Hematology 2019 guidelines for management of venous thromboembolism: prevention of venous thromboembolism in surgical hospitalized patients. Blood Adv 2019;3:3898-944.

18. Baldeh T, Saz-Parkinson Z, Muti P, et al. Development and use of health outcome descriptors: a guideline development case study. Health Qual Life Outcomes 2020;18:167.

19. Wiercioch W, Nieuwlaat R, Akl EA, et al. Methodology for the American Society of Hematology VTE guidelines: current best practice, innovations, and experiences. Blood Adv 2020;4:2351-65.

20. Guyatt GH, Thorlund K, Oxman AD, et al. GRADE guidelines: 13. Preparing summary of findings tables and evidence profiles-continuous outcomes. J Clin Epidemiol 2013;66:173-83. doi: 10.1016/j.jclinepi.2012.08.001. Epub Oct 30.

21. O'Sullivan D, Wilk S, Michalowski W, Farion K. Using PICO to align medical evidence with MDs decision making models. Studies in health technology and informatics 2013;192:1057.

22. Spilker D. Quality of life and pharmacoeconomics in clinical trials

. 1996.

23. Chou R, Dana T, Blazina I, Daeges M, Jeanne TL. Statins for Prevention of Cardiovascular Disease in Adults: Evidence Report and Systematic Review for the US Preventive Services Task Force. JAMA 2016;316:2008-24.

24. Egger SJ, Willson ML, Morgan J, et al. Platinum-containing regimens for metastatic breast cancer. Cochrane Database Syst Rev 2017;6:CD003374-CD.

25. Gartlehner G, Patel SV, Feltner C, et al. Hormone Therapy for the Primary Prevention of Chronic Conditions in Postmenopausal Women: Evidence Report and Systematic Review for the US Preventive Services Task Force. JAMA 2017;318:2234-49.

26. Guirguis-Blake JM, Evans CV, Senger CA, O'Connor EA, Whitlock EP. Aspirin for the Primary Prevention of Cardiovascular Events: A Systematic Evidence Review for the U.S. Preventive Services Task Force. Ann Intern Med 2016;164:804-13.

27. Hempel S, Newberry SJ, Maher AR, et al. Probiotics for the prevention and treatment of antibiotic-associated diarrhea: a systematic review and meta-analysis. JAMA 2012;307:1959-69.

28. Miller J, Chan BKS, Nelson HD. Postmenopausal estrogen replacement and risk for venous thromboembolism: a systematic review and meta-analysis for the U.S. Preventive Services Task Force. Ann Intern Med 2002;136:680-90.

29. Sutherland A, Naessens K, Plugge E, et al. Olanzapine for the prevention and treatment of cancer-related nausea and vomiting in adults. Cochrane Database Syst Rev 2018;9:CD012555-CD.

30. Zareifopoulos N, Dylja I. Efficacy and tolerability of vilazodone for the acute treatment of generalized anxiety disorder: A meta-analysis. Asian J Psychiatr 2017;26:115-22.

31. Alper BS, Oettgen P, Kunnamo I, et al. Defining certainty of net benefit: a GRADE concept paper. BMJ Open 2019;9:e027445.

32. Puhan MA, Singh S, Weiss CO, Varadhan R, Boyd CM. A framework for organizing and selecting quantitative approaches for benefit-harm assessment. BMC Med Res Methodol 2012;12:173.

Figures

Figure 1

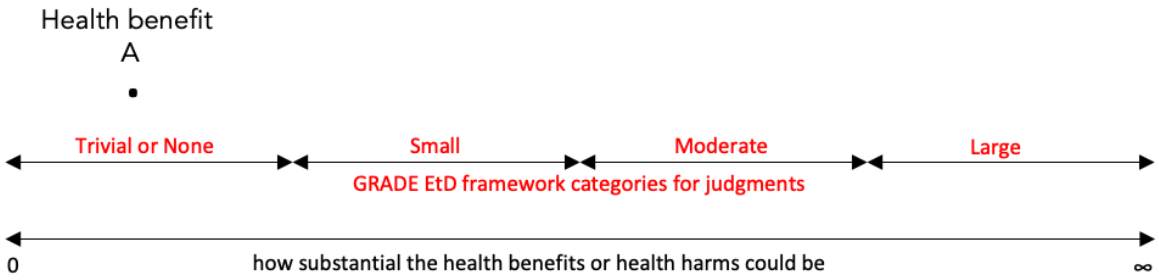


Figure 1: Example of association between health effects and categories of judgments

Figure 2

		Judgments from ASH VTE guideline panels			
		Trivial or None	Small	Moderate	Large
Judgments based on DTs	Trivial or None	37	9	2	0
	Small	1	0	1	0
	Moderate	0	0	0	1
	Large	0	0	1	1

Figure 2: Agreement between judgments from ASH VTE guideline panels and DTs

The green cells indicate the number of agreements for each category of judgment.

Appendices

Appendix 1

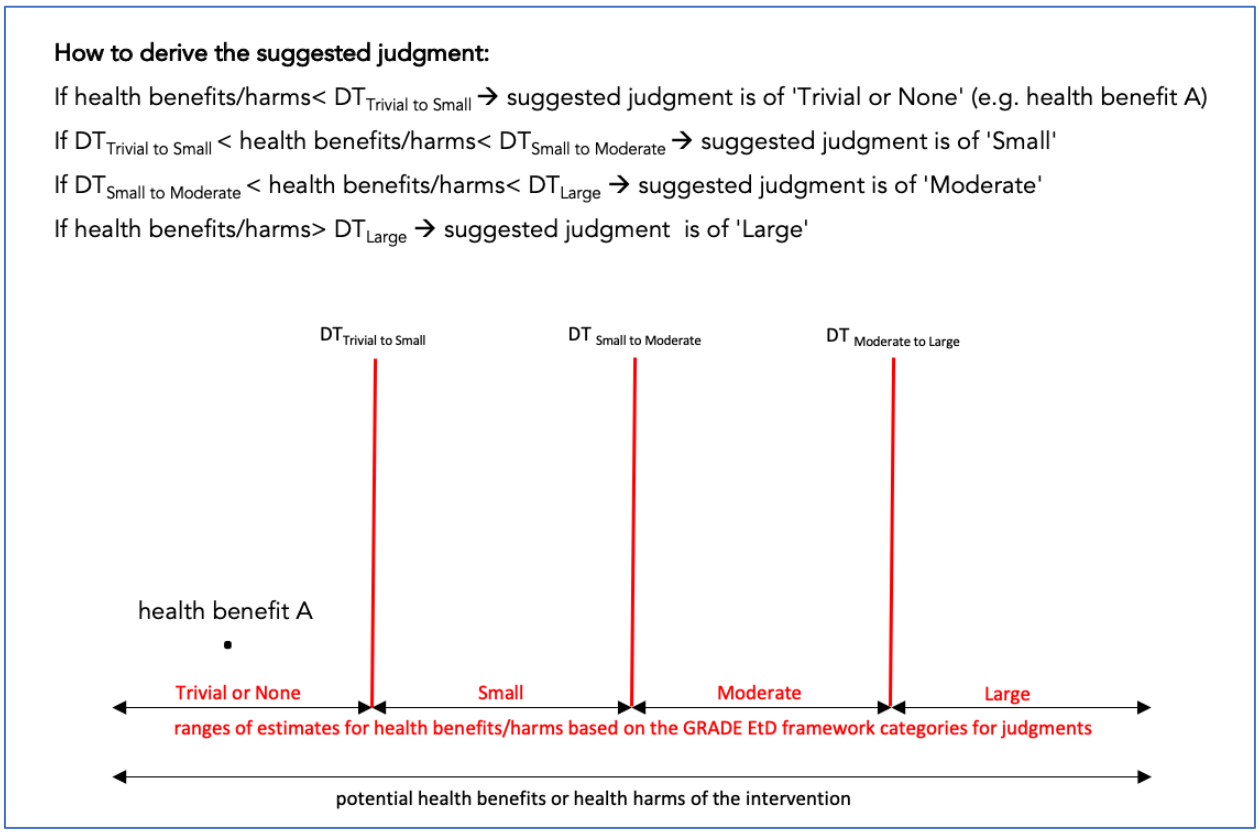


Figure 1

Figure 1 legend: The availability of three DTs ($DT_{Trivial/Small}$, $DT_{Small/Moderate}$, $DT_{Moderate/Large}$) would allow to discriminate between the four GRADE EtD framework categories for judgments. For a given health benefit/harm, the suggestion on the judgment would depend on how the estimate of health benefits/harms compares to the DTs. In this example, the health benefit A lies on the left (is smaller) of the $DT_{Trivial/Small}$ which would suggest that the judgment of 'Trivial or None' would be more appropriate than the others.

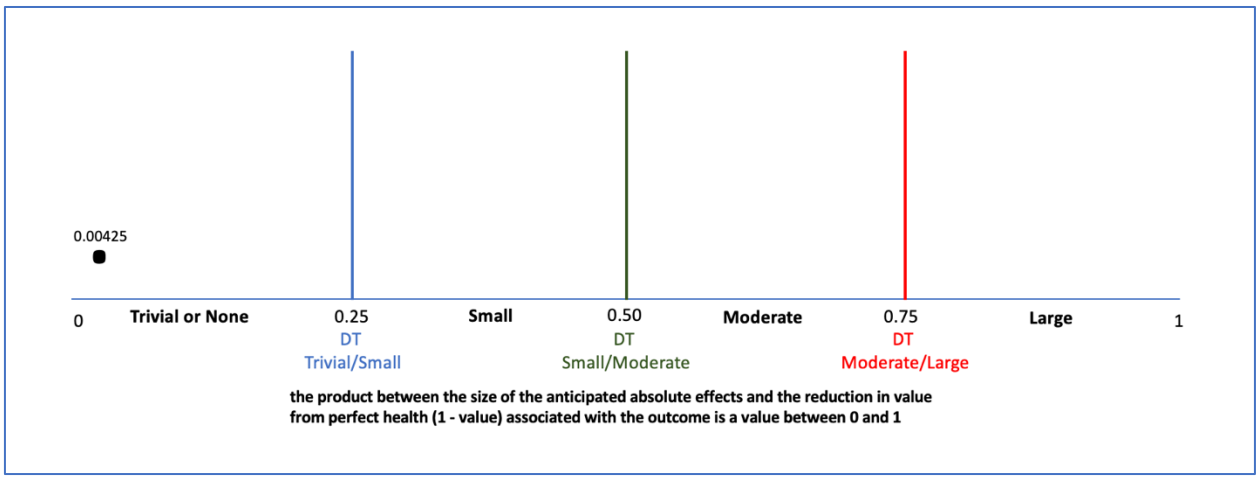


Figure 2

Figure 2 legend: We assumed to have known DTs ($DT_{Trivial/Small} = 0.25$, $DT_{Small/Moderate} = 0.50$, $DT_{Moderate/Large} = 0.75$) and wanted to assign to one of the 4 EtD categories the health benefit of an intervention showing an anticipated absolute effect of 17 fewer per 1000 on an outcome valued 0.75. Following the proposed approach, we calculated the result of the product of the size of anticipated effects (Absolute Risk Difference, ARD) and the reduction in value from perfect health (1 - outcome's value) associated with the outcome under evaluation. In this example, the following approach (ARD * (1 - outcome's value) = (17/1000)*(1 - 0.75)) resulted in the value of 0.00425. We then plotted this value and

obtained the suggested judgment according to the DTs approach that, in this case, would be of 'Trivial or None' considering that the calculated value is smaller than the $DT_{Trivial/Small}$.

Pulmonary Embolism of moderate severity

- **Symptoms:** You will experience shortness of breath, sometimes pain and tightness in your chest.
- **Time Horizon:** Moderate pulmonary embolism will impair you for weeks to months.
- **Testing and Treatment:** Testing includes x-rays and CT-scans. Treatment will be administered in the hospital for a few days or at home. It typically includes administration of blood thinners using a small tube inserted into your vein or injections, followed by pills for months to years. You may require oxygen administration to improve your symptoms. To identify the cause of your problem you may require additional testing such as blood work or other x-rays and similar tests.
- **Consequences:** You are at an increased risk of dying with a moderate pulmonary embolism. Consequences sometimes include persisting shortness of breath, particularly with exercise.

Box 1 - Example of Health Outcome Descriptor

Intervention A compared to no Intervenion A for primary prevention

Patient or population: healthy adults
Intervention: Intervention A
Comparison: no Intervention A

Outcome Follow-up	№ of participants (studies)	Certainty of the evidence (GRADE)	Relative effect	Anticipated absolute effects	
				Risk with no Intevention A	Risk difference with Intevention A
Outcome A follow up range: 2 to 4 weeks	2336 (8 RCTs)	HIGH	**RR 0.48**	85 per 1,000	**44 fewer per 1,000**

***The risk in the intervention group** is based on the assumed risk in the comparison group and the **relative effect** of the intervention.

RR: Risk ratio

As reported above, subjects who received Intervention A had **44 fewer cases of Outcome A per 1,000 people (4.4%)** compared to subjects who did not receive Intervention A.

This outcome has a value (utility) **of 0.8**. In other words, the reduction in value (disutility) from 1 (perfect health) is 0.2.

Figure 3 - Example of a case-scenario

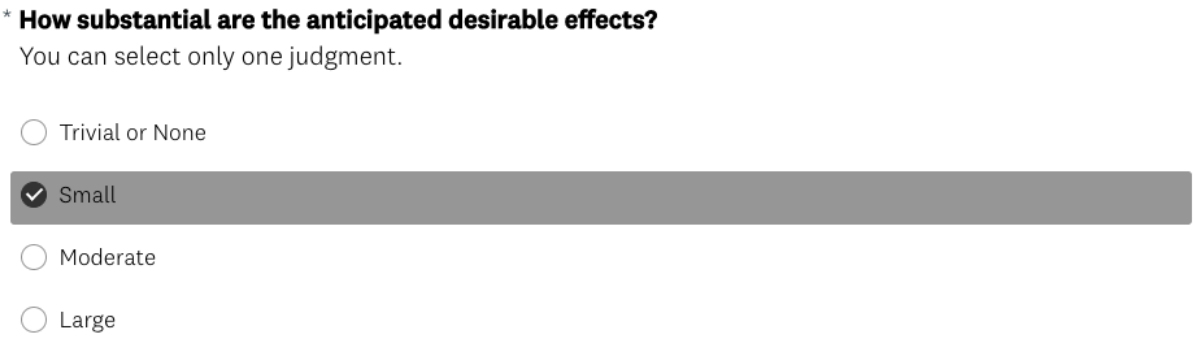


Figure 4 - Judgment on health benefits

Figure 5 - Selection of ranges for judgments of Small and Moderate

---

Example of calculation of DTs based on survey data

In the examples shown in Figures 4 and 6 we assumed that, after having evaluated a given case-scenario (ARD of 44 events fewer per 1000 on an outcome valued 0.8), a survey participant rated the hypothetical ranges of ARD for judgments of 'Small' and 'Moderate' of from 25 fewer per 1000 to 60 fewer per 1000, and of from 61 fewer per 1000 to 90 fewer per 1000, respectively. We used this data to derive the ranges of ARD for judgments of 'Trivial or None' and of 'Large' (table below).

boundaries of ranges described in Figure 6 value of the outcome = 0.8			
Trivial or None	Small	Moderate	Large

range of values						
lower bound	upper bound	lower bound	upper bound	lower bound	upper bound	range
0 per 1000	24 per 1000	25 per 1000	60 per 1000	61 per 1000	90 per 1000	more than 90 per 1000

Table 1 - Ranges of sizes of effects (ARD)

For each range of ARD, we calculated the product between range boundaries and the reduction in value from perfect health (1 - outcome's value) for the outcome associated. Then, we derived the DTs as follow: $DT_{Trivial/Small}$ equal to the the product calculated from the lower bound for the judgment of 'Small', $DT_{Small/Moderate}$ equal to average of the products calculated from the upper bound for the judgment of 'Small' and the lower bound for the judgment of 'Moderate', and $DT_{Moderate/Large}$ equal to the smallest number larger than the mean of the products calculated from the upper bound for the judgment of 'Moderate'.

product values = ARD * (1- outcome's value))						
Trivial or None		Small		Moderate		Large
range of values						
lower bound	upper bound	lower bound	upper bound	lower bound	upper bound	any value

(0/1000)*0.2	(24/1000) *0.2	(25 /1000) *0.2	(60/1000) *0.2	(61/1000)*0.2	(90/1000)*0.2	bigger than (90/1000)*0.2
0	0.0048	0.005	0.012	0.0122	0.018	>0.018

Table 2 - Ranges of product values

Using the data from Table 3, the DTs would result as follow: $DT_{Trivial/Small} = 0.005$, $DTs_{mall/Moderate} = 0.0121$, $DTModerate_{/Large} = 0.0180001$.

Outcome		iDTTrivial/Small	iDTSmall/Moderate	iDTModerate/Large
Death times it was randomizsd n=5	mean	0.00180000	0.01920000	0.04560000
	SD	0.00192354	0.01850126	0.03494710
	precision or margin of error 15%	0.00027	0.00288	0.00684
	required sample size	198	162	104
Major Ischemic Stroke n=8	mean	0.00316250	0.01203125	0.02626250
	SD	0.00565305	0.01788723	0.02737044
	precision or margin of error 15%	0.000474375	0.001804688	0.003939375
	required sample size	549	381	189
Moderate PE n=6	mean	0.00400000	0.01158333	0.02325000
	SD	0.00575326	0.01335648	0.02134772
	precision or margin of error 15%	0.0006	0.0017375	0.0034875
	required sample size	357	231	147
Moderate diarrhea n=6	mean	0.00735000	0.01600000	0.03908333
	SD	0.00878721	0.01923896	0.02814679
	precision or margin of error 15%	0.0011025	0.0024	0.0058625
	required sample size	248	250	92
Mild nausea/vomiting n=6	mean	0.00480000	0.01129000	0.02188000
	SD	0.00580517	0.00898342	0.01213036
	precision or margin of error 15%	0.00072	0.0016935	0.003282
	required sample size	253	112	56
Death n=7	mean	0.00185714	0.00750000	0.03371429
	SD	0.00167616	0.00525991	0.03542463
	precision or margin of error 15%	0.000278571	0.001125	0.005057143
	required sample size	143	87	192
Major Ischemic Stroke n=10	mean	0.00148500	0.00574750	0.01908500
	SD	0.00052178	0.00428135	0.01932415
	precision or margin of error 15%	0.00022275	0.000862125	0.00286275
	required sample size	25	98	179
Moderate PE n=3	mean	0.00141667	0.00675000	0.01575000
	SD	0.00101036	0.00388114	0.00584701
	precision or margin of error 15%	0.0002125	0.0010125	0.0023625
	required sample size	90	60	
Moderate diarrhea n=3	mean	0.00493333	0.00931667	0.02873333
	SD	0.00380044	0.00400385	0.01867655
	precision or margin of error 15%	0.00074	0.0013975	0.00431
	required sample size	105	35	76
Mild nausea/vomiting n=6	mean	0.00628333	0.01124583	0.01754167
	SD	0.00926616	0.01363886	0.01860807
	precision or margin of error 15%	0.0009425	0.001686875	0.00263125
	required sample size	375	255	196

Table 3 - Sample size calculation

Table 2 legend: We based our sample size calculation on the data collected during pilot-testing (n=15 participants). Based on this data, we computed the mean thresholds $DT_{Trivial/Small}$, $DTs_{mall/Moderate}$, $DT_{Moderate/Large}$ for each outcome separately. We estimated, for each of these thresholds, the required number of ratings to achieve the same mean for each

of the thresholds with an acceptable difference of 15% of the mean, and 95% confidence interval (Figure 7). We calculated that to achieve such level of precision in the estimate of each threshold, we need to recruit 1406 survey respondents. The required sample was calculated also accounting for the percentage of missing data that was observed analysis the responses from the first 50 participants (% of missing data = 0.025).

Selection criteria for inclusion of judgments from the ASH VTE guidelines in the comparative analysis:

We abstracted information (judgment, estimate of effects, value rating) about judgments on the EtD criteria about desirable and undesirable effects that were either based:

• on a single outcome

• on multiple outcomes, if only one outcome had an ARD different from 0 fewer/more per 1000

• of 'Trivial or none'; in this case, we assumed that the judgment was driven by the outcome with the smallest product between sizes of effects and value.

Box 2 - Selection criteria for inclusion of judgments

Judgments	**n, (%)**
available from the ASH VTE guidelines	208
met the inclusion criteria	53/208 (25.4)
from the VTE guideline on surgical patients	19/53 (36)
from the VTE guideline on medical patients	8/53 (15)
from the VTE guideline on pediatric patients	6/53 (11.4)
from the VTE guideline on optimal management of anticoagulation therapy	10/53 (18.8)
from the guideline on VTE treatment	10/53 (18.8)
on desirable effects	31/53 (58.5)
on undesirable effects	22/53 (41.5)
of 'Trivial or None' effects	38 (71.6)
of 'Small' effects	9 (17)
of 'Moderate' effects	4 (7.4)
of 'Large' effects	2 (4)

Table 4: Characteristics of judgments from the ASH VTE guidelines

CHAPTER 5. CONCLUSIONS

1. Summary of findings

This work presents three main pieces of research. Through these, the main findings can be summarized as follows:

a) The definition and application of explicit rigorous methods promote the development of transparent evidence-based guidelines and supports guideline panels in their activities.
b) The novel approach for guideline development at ISS seems feasible and acceptable to key stakeholders, including guideline panel members, those synthesizing the evidence, and the public.
c) The importance of basing decision-making on absolute measures of intervention effects in the context of recommendations has been established. However, guidance regarding the selection of the most appropriate baseline risk estimates is limited.
d) In the absence of direct prognostic evidence, the use of unadjusted surrogate data on baseline risk may bias the estimate of absolute treatment effects. The use of modeled estimates better represents the baseline risk for the outcomes of interest, addresses potential bias, and makes the decision-making process by guideline experts more explicit.

e) Structured approaches for making judgments on the GRADE EtD framework criteria on health benefits and harms are not yet available.

f) The analysis of judgments collected through the survey is suggestive of a relation between raters' judgments and the joint measure of absolute effects and outcome value. These results support the claim that decision-thresholds based on survey data might have the ability to discriminate between EtD judgments on health benefits and harms and inform panels.

2. Implications for guideline methodologists, panel members, and other stakeholders

The development of credible clinical practice guidelines entails many steps and requires various contributors. In the second Chapter of this dissertation, we described how systematic and transparent methods can support guideline groups in their activities and ensure that all the key components of a high-quality and trustworthy guideline are considered.[1,2] This implies that guideline groups should always apply rigorous pre-defined methods and adhere to them as much as possible. The absence of explicit criteria for guideline development poses the risk of introducing bias into the process and ultimately may undermine the credibility of the guideline recommendations.[3] In Chapter 3 and 4, we centered our efforts on conceptualizing novel approaches that may support guideline panels in their decision-making process and increase the overall transparency of the published guidelines. First, we showed how to derive a modelled estimate of the risk of having certain health conditions when this data is not directly available in the medical

literature. Second, we conceptualized and successfully tested an approach to support guideline panels in judging how substantial the desirable and undesirable effects of health interventions are. Other guideline methodologists may rely on our solutions to identify and address other areas in the guideline enterprise where further guidance is needed.[2] The strive for transparency that has fuelled this research should also stimulate organizations and guidelines' end-users toward the critical appraisal of existing guidelines and increased uptake of transparent, evidence-based, credible guidelines.

3. Strengths and challenges

This thesis dissertation has multiple strengths. In Chapter 2 we described, for the first time, the application of the new methodological standards for the development of practice guidelines at the Italian National Institute of Health (ISS). Apart from its novelty, the strength of this work stems from the fact that we strictly adhered to the ISS methodological manual for guideline development that we helped develop. The required process, which combines rigorous methods with systematic and transparent approaches, grants equal voices to panel members, includes an assessment of the certainty in the evidence according to the GRADE approach and endorses the use of the GRADE EtD framework.[4] While this process allowed the group on the management of autism spectrum disorders to produce a high-quality guideline, our work also aims to serve as a reference standard for future guideline development efforts in the Italian setting. The promotion of rigorous methods supports guideline developers in producing transparent evidence-based

guidelines that can improve the quality of healthcare and of health outcomes, and may lead to the reduction of unjustified variability in clinical practice in the national setting.[5] The approach to model baseline risk for patient-important outcomes when only baseline risk for surrogate outcomes are available described in Chapter 3 has the strength of simplicity, making its adoption appealing to guideline developers. The proposed approach requires the panel to agree on explicit assumptions and suggest their representation on a diagram allowing the panel to partake in a more intuitive process. The proposed approach can also be easily adapted to other scenarios where modeled estimates may be preferred over observed, surrogate estimates. Finally, this method to estimate baseline risk advocates for transparency in guideline development and, thus, makes it easier for guideline users to understand the rationale behind the panel's decisions. The conceptual development work described in Chapter 4 has the ambitious goal of identifying decision-thresholds for EtD judgments on desirable and undesirable effects of health interventions. The results of this early work suggestes that our study could open new avenues for increasing the transparency and consistency of judgments informing guideline recommendations. The main strength of that research is its reliance on empirical data that is derived from “guideline-like scenarios” that can inform judgments of other guideline panels. While our preliminary findings are encouraging, we acknowledge the potential challenge related to the number of participants required to reach adequate statistical power, and that our findings should be replicated in different settings and outcomes.

4. Final remarks

The guideline panel plays the key role in the development of trustworthy clinical practice guidelines. Despite the many investments in conceptualizing, refining, and disseminating structured methods that can aid panel members in the formulation of methodologically sound recommendations, there are many unresolved answers and we identified a few areas in guideline development that might benefit from practical guidance. Thus, this dissertation represents an effort to support guideline panels in making-decisions through the development and application of novel approaches and to serve as an example to guideline methodologists that might be interested in the development of original methodological research.

References

1. Qaseem A, Forland F, Macbeth F, et al. Guidelines International Network: toward international standards for clinical practice guidelines. Ann Intern Med 2012;156:525-31.

2. Schünemann HJ, Wiercioch W, Etxeandia I, et al. Guidelines 2.0: systematic development of a comprehensive checklist for a successful guideline enterprise. CMAJ 2014;186:E123-E42.

3. Grilli R, Magrini N, Penna A, Mura G, Liberati A. Practice guidelines developed by specialty societies: the need for a critical appraisal. Lancet 2000;355:103-6.

4. Morgano GP, Fulceri F, Nardocci F, et al. Introduction and methods of the evidence-based guidelines for the diagnosis and management of autism spectrum disorder by the Italian National Institute of Health. Health Qual Life Outcomes 2020;18:81.

5. Iannone P, Coclite D, Fauci AJ, Graziano G, Napoletano AM. [Italian guidelines in accordance with the new National Guidelines System: critical issues and perspectives.]. Recenti Prog Med 2017;108:360-2.