RESEARCH ARTICLE

A deeper analysis in thyroid research: A meta-epidemiological study of the American Thyroid Association clinical guidelines

Dalia A. Castillo-Gonzalez[1,2☯], **Edgar G. Dorsey-Trevino**[1,2☯], **Jose G. Gonzalez-Gonzalez**[1,2,3], **Mariana Garcia-Leal**[1], **Karen G. Bautista-Orduño**[1,2], **Karina Raygoza**[2], **Michael R. Gionfriddo**[4], **Naykky M. Singh Ospina**[5,6], **Rene Rodriguez-Gutierrez**[1,2,5,7]*

1 Endocrinology Division, Department of Internal Medicine, University Hospital "Dr. José E. González", Universidad Autónoma de Nuevo León, Monterrey, México, **2** Plataforma INVEST Medicina UANL–KER Unit Mayo Clinic (KER Unit México), Universidad Autónoma de Nuevo León, Monterrey, México, **3** Research Unit, University Hospital "Dr. José E. González", Universidad Autónoma de Nuevo León, Monterrey, México, **4** Center for Pharmacy Innovation and Outcomes, Geisinger Health System, Forty Fort, PA, United States of America, **5** Knowledge and Evaluation Research Unit in Endocrinology, Mayo Clinic, Rochester, MN, United States of America, **6** Division of Endocrinology, Department of Medicine, University of Florida, Gainesville, FL, United States of America, **7** Division of Endocrinology, Diabetes, Metabolism, and Nutrition, Department of Medicine, Mayo Clinic, Rochester, MN, United States of America

☯ These authors contributed equally to this work.
* rodriguezgutierrez.rene@mayo.edu

OPEN ACCESS

Citation: Castillo-Gonzalez DA, Dorsey-Trevino EG, Gonzalez-Gonzalez JG, Garcia-Leal M, Bautista-Orduño KG, Raygoza K, et al. (2020) A deeper analysis in thyroid research: A meta-epidemiological study of the American Thyroid Association clinical guidelines. PLoS ONE 15(6): e0234297. https://doi.org/10.1371/journal.pone.0234297

Editor: Honey V. Reddi, The Jackson Laboratory for Genomic Medicine, UNITED STATES

Received: December 23, 2019

Accepted: May 24, 2020

Published: June 10, 2020

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pone.0234297

Abstract

Background

The American Thyroid Association (ATA) uses the GRADE or the American College of Physicians (ACP) system to develop recommendations. Recommendations based on low-quality evidence should spur for the conduction of clinical studies, if feasible. The extent to which recommendations by the ATA based on low-quality of evidence are being actively researched remains unknown.

Methods

Clinical guidelines produced by the ATA using the GRADE or the ACP system to classify evidence were deemed eligible. Reviewers, in duplicate and independently, extracted therapeutic recommendations based on low-quality evidence, whereas recommendations with higher quality of evidence, aimed at diagnosis, or best practice statements were excluded. Eligible recommendations based on low-quality evidence were deconstructed to their components using the PICO format. We then searched on clinicaltrials.gov to identify ongoing research. Trials were deemed eligible if they addressed the PICO question with at least one of the intended outcomes.

Results

A total of 543 recommendations were retrieved, of which 305 (56%) were based on low-quality of evidence and only 90 were deemed eligible. Of these, we found that 33 (37%) recommendations were actively being researched in 53 clinical trials. Most of the trials were

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The authors received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

randomized and funded by non-profit organizations. Many clinical trials studied thyroid nodules and differentiated thyroid cancer (26/53; 49%), whereas few studied were aimed at anaplastic thyroid cancer (2/53; 4%).

Conclusion

One out of three of gaps in evidence, identified as low quality during the development of ATA guidelines, are currently actively researched. This finding calls for the need to develop a better research infrastructure and funding to support thyroid research.

Introduction

When creating guideline recommendations, guideline panelists estimate the benefits and harms of relevant options and rate the confidence they have in those estimates. [1, 2] The Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach facilitates this process by offering guidance about how to determine the quality of evidence and strength of each recommendation. [3, 4] Quality of the evidence is determined to be low or very-low when the source of the estimates is susceptible to bias or methodological limitations (e.g., observational studies), [5] when several studies provide widely different estimates of effect (i.e., inconsistency), [6] or the estimates are not directly applicable to the population of interest (i.e., indirectness) [7]. As a result, recommendations derived from low or very-low quality of evidence lead to weak (or conditional) recommendations that are often seen as gaps in knowledge. [8, 9]

These knowledge gaps present areas of opportunities for research that will likely have an impact on the quality of evidence, strength of future recommendations, and patient care. However, recent evidence suggests that these knowledge gaps often remain so due to lack of active research. [10] Recently, the American Thyroid Association (ATA)—a professional organization that develops clinical practice guidelines for the care of patients with thyroid conditions—has crafted some clinical guidelines based on the GRADE system and the American College of Physicians (ACP) Guideline Grading System, which is a modified version of the GRADE system that adopted most of its rationale and judgmental criteria when developing clinical recommendations [11]. Since then, five clinical guidelines have been developed based on these systems—hyperthyroidism and thyrotoxicosis, [12] differentiated thyroid cancer and thyroid nodules, [13] thyroid dysfunction in pregnancy and postpartum, [14] hypothyroidism, [15] and anaplastic thyroid cancer. [16] There is, therefore, an opportunity to assess current knowledge gaps and the current research environment within conditions affecting the thyroid.

To that end, we performed a meta-epidemiological study on ATA guidelines aiming to describe recommendations supported by low-quality evidence and determine the extent to which these potential knowledge gaps are being actively researched.

Methods

Study design

This study was conducted in alignment with an existing reporting guideline for meta-epidemiological studies (S1 Table). [17] A meta-epidemiological study is a synthesis of non-clinical evidence that usually uses non-patient outcomes, such as specific characteristics or settings of clinical studies, describing distribution of evidence, examining heterogeneity and exploring its causes, identifying and describing plausible biases, and providing empirical evidence for hypothesized association. [17]

Eligibility criteria and data collection

Clinical guidelines were downloaded from the ATA web page (https://www.thyroid.org/professionals/ata-professional-guidelines/) and only guidelines using the GRADE or the ACP system to provide recommendations were selected. [12–14, 16, 18] Using a standardized web-extraction form (Microsoft Excel 2016, Microsoft®, Redmont, WA, USA), reviewers, in duplicate and independently, extracted; guideline topic, year of publication, strength of recommendation, and quality of evidence. A pilot test using 20 recommendations was performed to standardized reviewers' criteria and avoid misclassification of recommendations as either diagnostic, prognostic, or intervention. Recommendations about treatment interventions based on low-quality of evidence were included, whereas those based on moderate or high-quality evidence were excluded. Recommendations focused on either diagnostic or prognostic evaluations, although commonly associated with low-quality evidence, were purposefully excluded as these studies are less likely to be answered with or conducted with an interventional study. [19] Also, there are some recommendations for which no sensible alternative exists, and thus, no one would consider doing a study to elucidate the answer to the implicit question. These recommendations were labeled as "best practice" and were also excluded from the study (e.g., *patients should be informed of side effects of ATDs and the necessity of informing the physician promptly if they should develop pruritic rash, jaundice, acholic stools or dark urine, arthralgias, abdominal pain, nausea, fatigue, fever, or pharyngitis. Preferably, this information should be in writing. Before starting ATD and at each subsequent visit, patient should be alerted to stop the medication immediately and call their physician if there are symptoms suggestive of agranulocytosis of hepatic injury*). [20] We have provided a table with examples of the aforementioned recommendations (S2 Table). Clinical guidelines using the ACP system lack de classification of "best practice, but as this system mimics the rationale and criteria of the GRADE system to craft recommendations, experts in the GRADE system (RR-G and MRG) analyzed all recommendations and determined which recommendations were "best-practice".

To elucidate the research gaps, recommendations having low-quality of evidence were deconstructed into its components using the PICO format—patient, intervention, comparison, and outcome—in order to create a research question for which a clinical trial could plausibly answer and to ensure the reproducibility of our methods. To calibrate reviewers' criteria for the latter, a pilot test using 15 recommendations were used to depict the research question, yielding an initial agreement of 80%. This process was repeated until percentage agreement was of 100%; this was achieved after 25 recommendations. Disagreements were initially resolved by consensus, and whenever this was not possible, an expert endocrinologist and/or methodologist (RR-G or MRG) made the decision.

A search on clinicaltrials.gov—considered the most complete registry site for clinical trials as it includes 208 countries—was performed. We used the research questions obtained from the recommendations based on low-quality of evidence for our search. [21] We specifically looked for clinical trials and excluded observational studies as they often provide low-quality evidence. A clinical trial was deemed eligible if it addressed the PICO question with at least one of the intended outcomes and if they were either completed or ongoing less than 5 years prior to the publication of the respective clinical guideline. Extracted information from each study included; funding, year of registry, location of study (country and continent), number of centers, type of allocation, recruitment status (i.e. not yet enrolling, ongoing, active), intervention model (e.g., single group, parallel, crossover, or factorial assignment), phase (i.e., I-IV), and masking (blinding).

To record the adherence from each clinical trial with the formulated research question, we labeled each trial into complete, partial, or incomplete alignment. Trials with complete alignment had to match all four dimensions of the PICO question, whereas trials with ≤2 dimensions were deemed to have incomplete alignment. All other trials were labeled as partially aligned. Reviewers working independently searched the clinical trial registry until a 100% agreement was achieved, a point reached after searching for 20 questions. A summarized diagram of the research methods is depicted in Fig 1.

Statistical analysis

We used descriptive statistics to report categorical variables with frequencies and percentages. We used SPSS version 22 (IBM Corp®, Armonk, NY, USA) for all statistical analysis. Graphics were designed using Microsoft Excel 2016 (Microsoft®, Redmont, WA, USA).

Results

We retrieved a total of seven clinical guidelines from the ATA published between 2012 to 2017, of which only five used the GRADE system to craft recommendations (Fig 2). [12–16] Interobserver agreement in the extraction of recommendations ranged from 0.9 to 1 and yielded a total of 543 recommendations. High-quality evidence supported 7% (n = 39) of the recommendations, moderate-quality 37% (n = 199), low-quality 51% (n = 278), and 5% (n = 27) of the recommendations were labeled as having insufficient evidence to make recommendations (Fig 2). A total of 238 (44%) recommendations were excluded due to moderate- or high-quality evidence. Afterward, recommendations considered as best practice or aimed at diagnostic endpoints were excluded, leaving a total of 90 recommendations supported by low-quality evidence (Fig 2).

Less than half of these recommendations (33/90; 37%) are being actively researched (Fig 3). Current active research is primarily focused on differentiated thyroid cancer (DTC) (14 out of 24 recommendations; 58%) and hypothyroidism (7 out of 14 recommendations; 50%); anaplastic thyroid cancer was the least active area of active research. (ATC) (1 out of 5; 16%) (Fig 3).

Most of the trials were funded by non-profit organizations (47/53; 89%), had a parallel design (30/53; 57%), were randomized (32/53; 60.4%), and were actively recruiting (21/53; 40%) (Table 1). Almost two-thirds of trials (33/53; 62%) were found to be in complete alignment with the PICO question, whereas only 3 of 53 trials (6%) had an incomplete alignment (Table 1).

Discussion

Summary of findings

Less than 10% of the recommendations endorsed by the ATA guidelines stem from high-quality evidence while over 50% is based on low-quality evidence. Only one out of three knowledge gaps identified in management recommendations of common thyroid disorders are actively being researched.

Comparison with previous studies

Our findings align with previous evidence elucidating the extent of which clinical guidelines are based on low-quality of evidence. [22, 23] Murad et. al. determined that 65% of

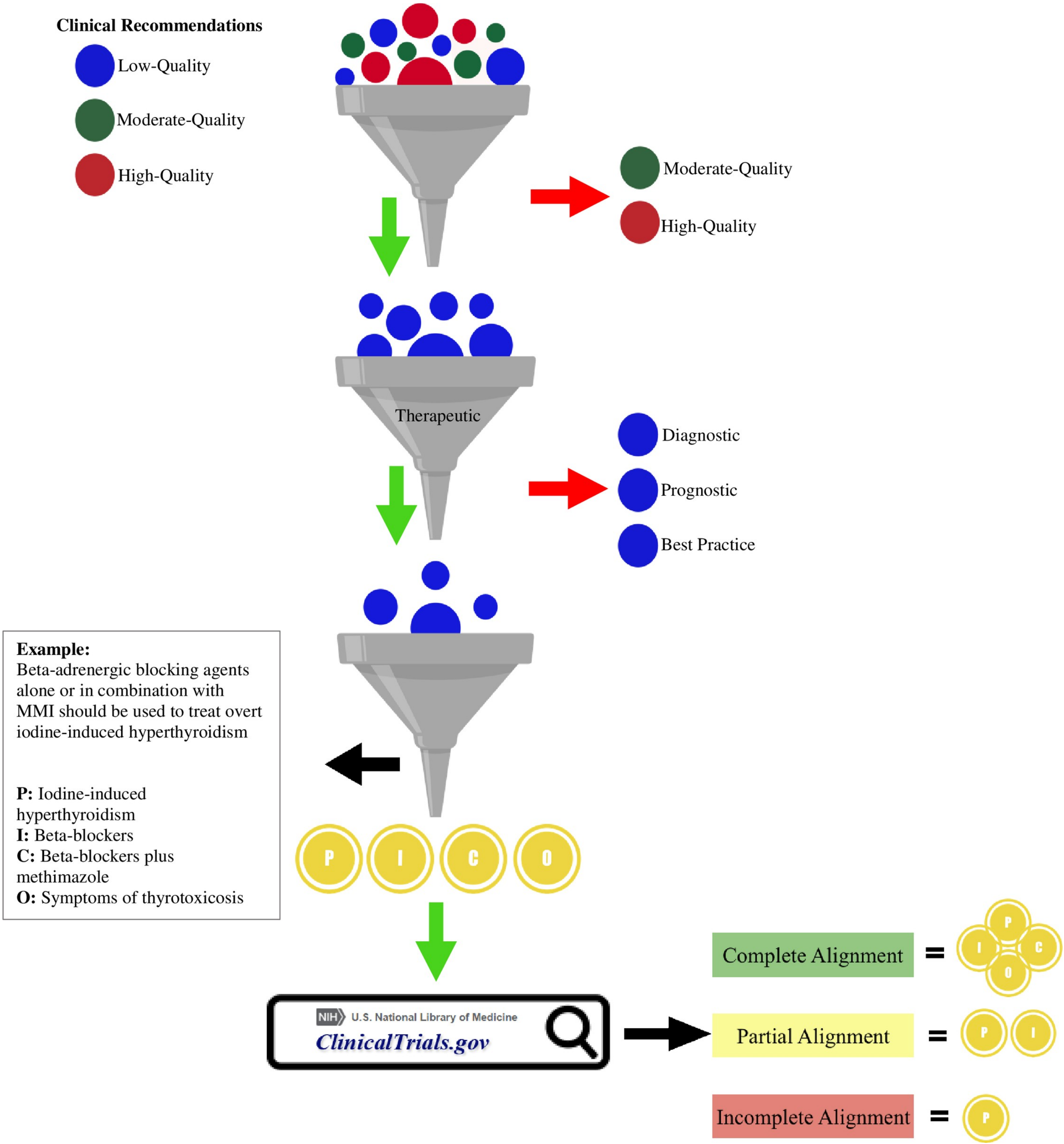
Clinical Recommendations
Low-Quality
Moderate-Quality
High-Quality
Moderate-Quality
High-Quality
Therapeutic
Diagnostic
Prognostic
Best Practice
Example:
Beta-adrenergic blocking agents alone or in combination with MMI should be used to treat overt iodine-induced hyperthyroidism
P: Iodine-induced hyperthyroidism
I: Beta-blockers
C: Beta-blockers plus methimazole
O: Symptoms of thyrotoxicosis
P
I
C
O
NIH U.S. National Library of Medicine
ClinicalTrials.gov
Complete Alignment
=
Partial Alignment
=
Incomplete Alignment
=

Fig 1. Summary of methods. Balls represent clinical recommendations. Red arrows mean exclusion criteria, whereas green arrows means inclusion criteria. P: Patient; I: Intervention; C: Comparison; O: Outcome.

https://doi.org/10.1371/journal.pone.0234297.g001

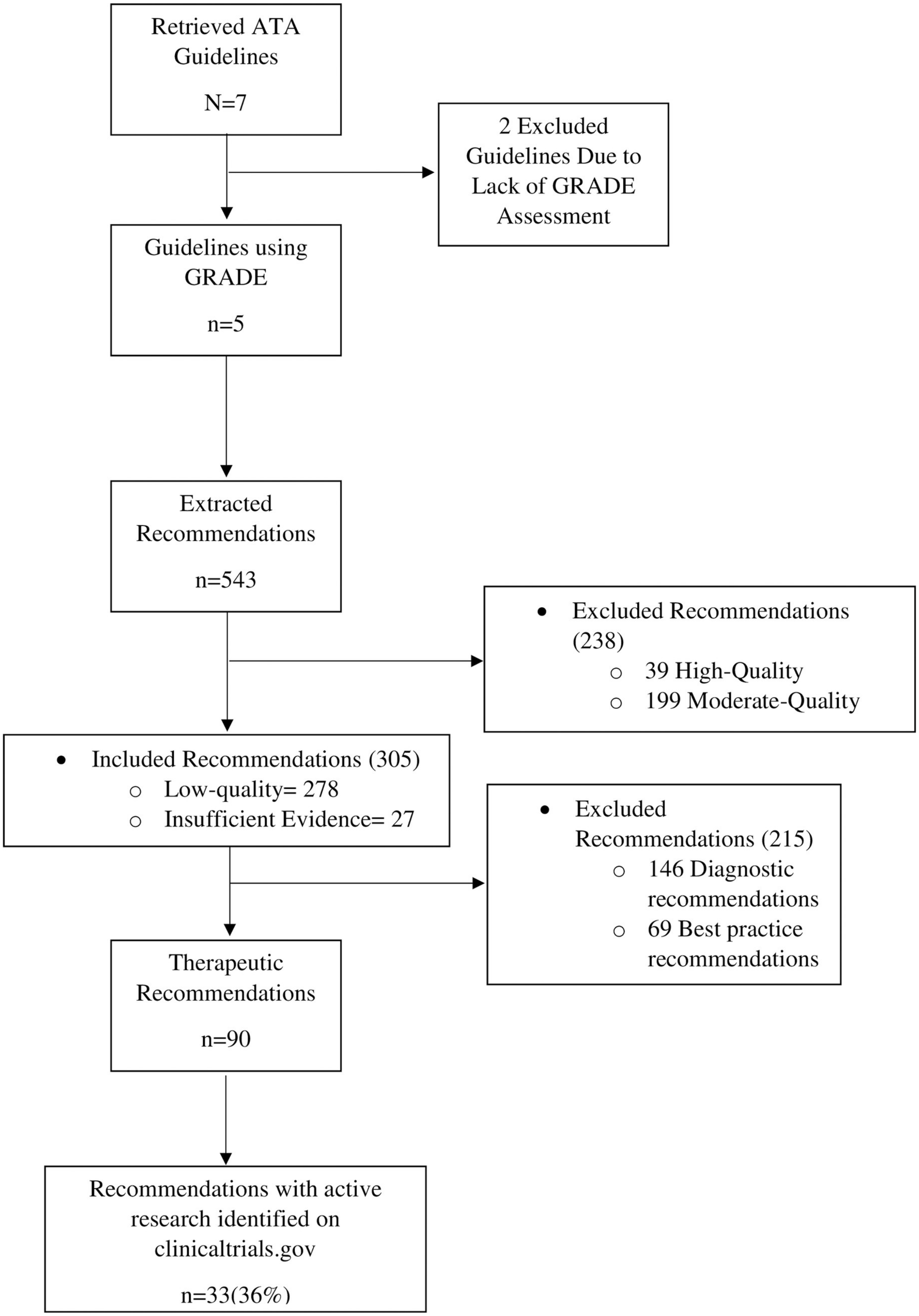
Retrieved ATA Guidelines
N=7
2 Excluded Guidelines Due to Lack of GRADE Assessment
Guidelines using GRADE
n=5
Extracted Recommendations
n=543
• Excluded Recommendations (238)
○ 39 High-Quality
○ 199 Moderate-Quality
• Included Recommendations (305)
○ Low-quality= 278
○ Insufficient Evidence= 27
• Excluded Recommendations (215)
○ 146 Diagnostic recommendations
○ 69 Best practice recommendations
Therapeutic Recommendations
n=90
Recommendations with active research identified on clinicaltrials.gov
n=33(36%)

Fig 2. Flow diagram of included guidelines and recommendations.

https://doi.org/10.1371/journal.pone.0234297.g002

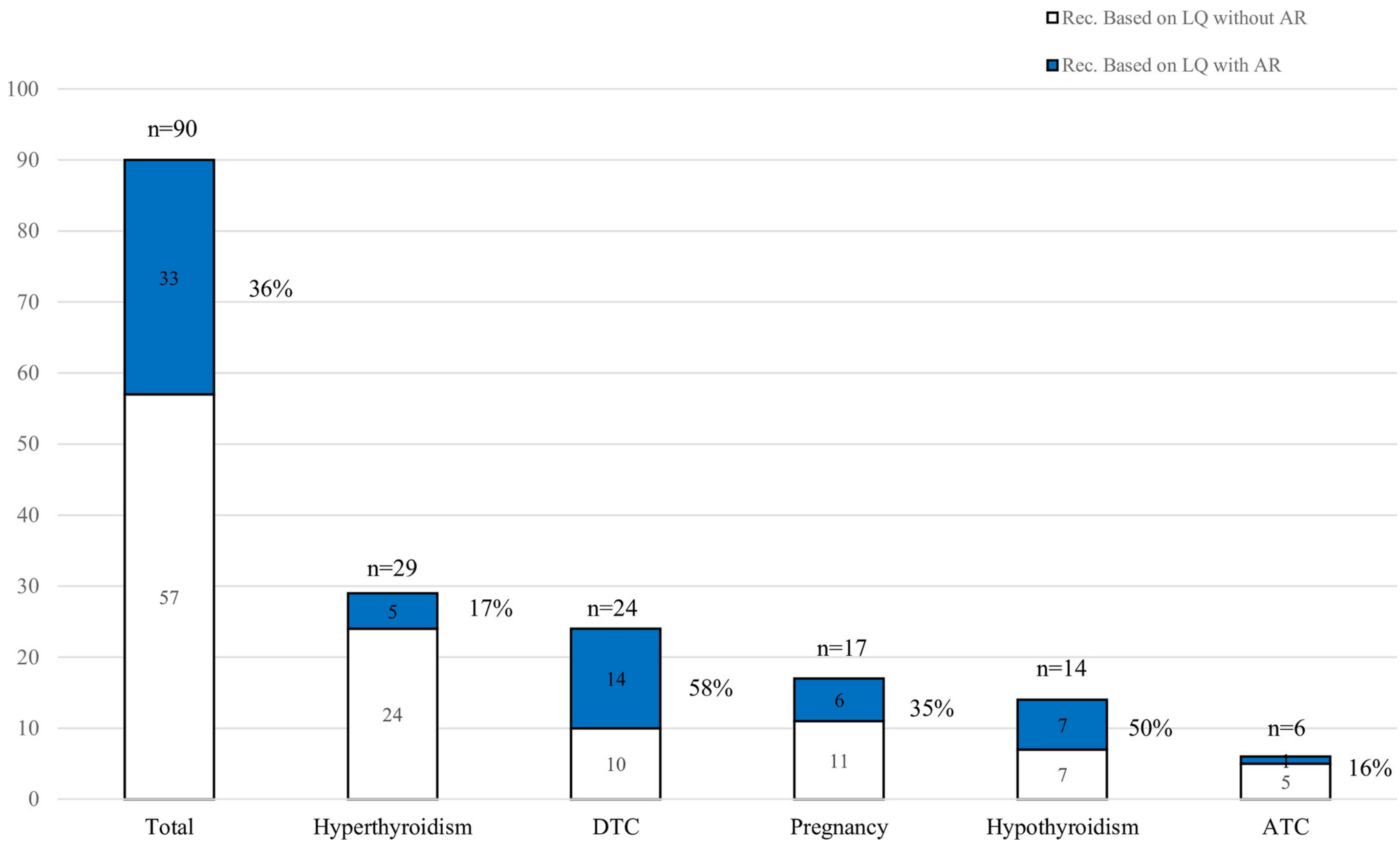

Fig 3. Proportion of recommendations based on low-quality evidence with and without active research. Percentages represents recommendations with low-quality of evidence with active research. LQ: Low-quality; Rec: Recommendation; AR: Active Research; DTC: Differentiates Thyroid Cancer; ATC: Anaplastic Thyroid Cancer.

https://doi.org/10.1371/journal.pone.0234297.g003

recommendations crafted by the Society for Vascular Surgery were strong but based on low-quality of evidence. [24] Similar findings have been found in the American College of Cardiology/American Heart Association and the Infectious Diseases Society of America guidelines in which nearly 48% and 50% of recommendations were drawn based on low-quality of evidence, respectively. [25, 26] More recently, an assessment made by Alexander et. al. on the World Health Organization (WHO) guidelines find that more than half (55.5%) of strong recommendations were based on either very- or low quality of evidence. [27] In endocrine, Brito et. al. evaluated the Endocrine Society guidelines and found that 58% of recommendations were predominantly based on observational studies and thus, conferring low-quality of evidence. [28] Conversely, an analytical survey performed by Agoristas et. al. showed that most recommendations spawned in UpToDate were based on moderate- or high-quality evidence. [29] None of the aforementioned studies, however, sought to determine the extent to which these knowledge gaps were being assessed by the research enterprise.

An analytical assessment performed by Checketts et. al. on the American Society of Plastic Surgeon's clinical guidelines for breast reconstruction found that 6 out of 10 identified research gaps were being evaluated by active research. [30] While, Singh Ospina et. al. found that 3 out of 10 recommendations based on low-quality of evidence the Endocrine Society guidelines were being assessed by active clinical trials. [10] Our findings are more in line with those of Singh-Ospina; we found that for the ATA guidelines, active research was being conducted on around one-third of the recommendations based on low-quality evidence.

Table 1. Description of interventional studies.

Trial Characteristics		Included Trials (53)
Thyroid Topic		
	Pregnancy	9 (17)
	Hyperthyroidism	5 (9)
	TN and DTC	26 (49)
	Hypothyroidism	11 (21)
	ATC	2 (4)
Funding		
	Non-Profit Sources	47 (89)
	Industry	6 (11)
Continent		
	America	20 (38)
	Europe	20 (38)
	Africa	3 (6)
	Asia	10 (19)
Intervention Masking		
	Parallel	30 (57)
	Single Group	16 (30)
	Factorial	3 (6)
	Crossover	4 (7)
Allocation		
	Not Randomized	7 (13)
	Randomized	32 (60)
	No reported	13 (24)
Phase		
	1	4 (7)
	2	12 (23)
	2 & 3	5 (9)
	3	16 (30)
	4	7 (13)
	N/A	9 (17)
Status		
	Active, Not Recruiting	12 (23)
	Recruiting	21 (40)
	Completed	20 (38)
Alignment		
	Complete	33 (62)
	Partial	17 (32)
	Incomplete	3 (6)

Data is presented as frequencies (percentages). TN: Thyroid Nodule; DTC: Differentiated Thyroid Cancer; ATC: Anaplastic Thyroid Cancer; N/A: Not Applicable

https://doi.org/10.1371/journal.pone.0234297.t001

Strengths and limitations

We utilized a protocol-driven method, that facilitates reproducibility however, our results have several limitations that may reduce the confidence in our estimates. First, there are several instances in which performing a clinical study may be unethical or not feasible and by excluding observational evidence, our results may underestimate the extent to which research is addressing current knowledge gaps. However, we purposefully excluded best-practice recommendations which reduce this possibility. Second, we excluded other trial registries but due to the scope of clinicaltrials.gov we are confident that excluding additional registries did not appreciably alter our estimates. Additionally, most guidelines were using the ACP system to qualify recommendations, and thus, lacked a "best practice" category. But as this system mimics the rationale and criteria of the GRADE system and experienced methodologists made the final evaluation, we consider this limitation to be insignificant. Despite these limitations, this approach has successfully been used previously and experts in methodology and clinical research planned and guided the conduct of this study, [10] thereby supporting the validity of this approach.

Implications for clinical practice and research

Based on our findings, we have found that most clinical recommendations, put forth by the ATA guidelines, are based on observational studies. This high prevalence may be due to several perks of thyroid research. For instance, conducting a multinational RCT requires for a multidisciplinary team that helps with the design, recruitment, analysis, and publication, and all these can only be afforded by proper funding. Yet, recent evidence has shown that thyroid research lacks the proper funding when compared to other areas with more prevalence and impact on morbidity and mortality. [31] This lack of funding impairs the proper conduction of RCT's, and therefore, limits panelists to utilize the best available evidence to create clinical guidelines, despite the best stands for mostly being observational studies. But are observational studies always deemed to deploy low-quality of evidence? The answer is no.

There is an unquestionable valuableness of observational studies to yield causal results whenever conducted properly and adjusting for possible confounders. [32, 33] Using observational data to make causal statements, however, requires the proper methods; causal questions requires causal method to answer them. For this, a rogue group of statisticians and epidemiologists defied the adage of "association is not causation" and stated that causality could be estimated from observational studies when proper methods are applied. [34–37] When this premises are violated, however, observational studies may yield an association between a treatment and an outcome even if this physiological plausibility fails to hold in real-world scenarios. For instance, one of the most famous cases that bespeak the latter concept was the use of hormone replacement therapy (HRT) to prevent CVD in postmenopausal women. [38] The Framingham Cohort Study showed that the incidence of myocardial infarction (MI) among premenopausal women was virtually inexistent. This led to the assumption that by treating postmenopausal women with HRT, the incidence of MI will decrease. [38] Yet, this surmise failed when tested in a RCT. [39] Along this example, there are several other studies that highlight the precautions of using observational studies to provide care when analyzed incorrectly. [39, 40]

On the other hand, the conduction of an RCT's are far from being exempt of bias. [41] Blindly inputting our faith towards everything that is labeled "randomized clinical trial" spurs for an inadequate application of evidence-based medicine to patient care. [42] A thorough and peruse appraisal of the evidence is necessary for clinicians to make confident clinical decisions and for panelists to develop trustworthy clinical recommendations. Additionally, for some thyroid condition, especially rare diseases, with low prevalence, or high mortality, conducting a RCT would demand a tremendous effort, funding, recruitment, and some sort of fortunateness, and most likely, the resulting RCT will be underpowered, unrepresentative, and with misleading results. Observational studies may sometimes be the best and most feasible option to provide a causal inference and implement it to patient care.

Our results have also denoted that even for diseases with high prevalence, such as hyperthyroidism, hypothyroidism, thyroid disease during pregnancy, and differentiated thyroid cancer, this knowledge gap is undesirably prevalent. In best case scenario, only half of these recommendations were being evaluated by a RCT's. A conjecture of our own is that the research enterprise has probably deemed these conditions as being of low priority due to their benign and relatively good prognosis. For instance, the 10-year mortality for papillary thyroid cancer has been estimated to be up to 97% in some studies, [43] whereas for hypo- and hyperthyroidism the prognosis are also considered to be fair when treated. Nonetheless, good prognosis is unrelated to a decrease in patients' expenditures and healthcare costs or an avoidance of the hindrance of life-long treatments or the permanent scarring of surgery. *De facto*, the increasing prevalence of these conditions would be putative of causing billion dollars in healthcare costs in the following years. [31]

Opportunities for future research and enhancement of clinical guidelines

Beyond identifying knowledge gaps, we identified redundancy in some areas of research. While multiple studies on similar topics can be useful to explore heterogeneity and decrease imprecision, it may also be an indicator of misusage of resources. In efforts to avoid this, researchers may guide their research pipeline by using systematic reviews to justify new studies. [44–46] Furthermore, some of the ATA guidelines have highlighted areas of opportunities in which future research should direct their efforts. This section was found in 3 out of 5 guidelines (thyroid nodules and differentiated thyroid cancer, thyroid disease during pregnancy and postpartum, and hypothyroidism), [13–15] and we encourage panelists to add this same section for all other guidelines as it poses a trustable guidance for researchers to direct their efforts towards meaningful clinical queries.

Lastly, the ATA, as mentioned previously, primarily uses the ACP system to develop recommendations. Although a recent statement mentioned that there are piloting with the GRADE system, [47] full endorsement would aid in the advocacy of transparency and informativeness. For instance, the rating of the evidence in GRADE has 4 categories (very low, low, moderate, and high), whereas the ATA guidelines only used 3 (low, moderate and high). [4, 11] While this may simplify rating for clinicians, it may also obscure the most necessary areas of research.

Conclusions

Less than one-third of low quality recommendations are actively being researched. Our results suggest that these guidelines may serve as an opportunity for the research enterprise to aim their research, after evaluating the feasibility, funding, and priorities, at these knowledge gaps.

Supporting information

S1 Table. Reporting guidelines for meta-epidemiological studies.
(DOC)

S2 Table. Paradigmatic situations in which panels may reasonably offer strong based on low confidence in effect estimates.
(DOCX)

Author Contributions

Conceptualization: Naykky M. Singh Ospina, Rene Rodriguez-Gutierrez.

Data curation: Dalia A. Castillo-Gonzalez, Edgar G. Dorsey-Trevino, Mariana Garcia-Leal, Karen G. Bautista-Orduño, Karina Raygoza, Rene Rodriguez-Gutierrez.

Formal analysis: Dalia A. Castillo-Gonzalez, Edgar G. Dorsey-Trevino.

Funding acquisition: Jose G. Gonzalez-Gonzalez.

Investigation: Jose G. Gonzalez-Gonzalez, Mariana Garcia-Leal, Karen G. Bautista-Orduño, Karina Raygoza, Rene Rodriguez-Gutierrez.

Methodology: Dalia A. Castillo-Gonzalez, Edgar G. Dorsey-Trevino, Jose G. Gonzalez-Gonzalez, Mariana Garcia-Leal, Karen G. Bautista-Orduño, Karina Raygoza, Rene Rodriguez-Gutierrez.

Supervision: Rene Rodriguez-Gutierrez.

Validation: Rene Rodriguez-Gutierrez.

Visualization: Rene Rodriguez-Gutierrez.

Writing – original draft: Dalia A. Castillo-Gonzalez, Edgar G. Dorsey-Trevino.

Writing – review & editing: Edgar G. Dorsey-Trevino, Jose G. Gonzalez-Gonzalez, Michael R. Gionfriddo, Naykky M. Singh Ospina, Rene Rodriguez-Gutierrez.

References

1. Institute of Medicine Committee on Standards for Developing Trustworthy Clinical Practice G. In: Graham R, Mancher M, Miller Wolman D, Greenfield S, Steinberg E, editors. Clinical Practice Guidelines We Can Trust. Washington (DC): National Academies Press (US) Copyright 2011 by the National Academy of Sciences. All rights reserved.; 2011.
2. Brito JP, Montori VM, Domecq JP, Murad MH, Guyatt GH. The Endocrine Society Guidelines: When the Confidence Cart Goes Before the Evidence Horse. The Journal of Clinical Endocrinology & Metabolism. 2013; 98:3246–52.
3. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008; 336:924–6. https://doi.org/10.1136/bmj.39489.470347.AD PMID: 18436948
4. Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. Journal of clinical epidemiology. 2011; 64:401–6. https://doi.org/10.1016/j.jclinepi.2010.07.015 PMID: 21208779
5. Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, et al. GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias). Journal of clinical epidemiology. 2011; 64:407–15. https://doi.org/10.1016/j.jclinepi.2010.07.017 PMID: 21247734
6. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 7. Rating the quality of evidence—inconsistency. Journal of clinical epidemiology. 2011; 64:1294–302. https://doi.org/10.1016/j.jclinepi.2011.03.017 PMID: 21803546
7. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 8. Rating the quality of evidence—indirectness. Journal of clinical epidemiology. 2011; 64:1303–10. https://doi.org/10.1016/j.jclinepi.2011.04.014 PMID: 21802903
8. 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; 66:726–35. https://doi.org/10.1016/j.jclinepi.2013.02.003 PMID: 23570745
9. Andrews J, Guyatt G, Oxman AD, Alderson P, Dahm P, Falck-Ytter Y, et al. GRADE guidelines: 14. Going from evidence to recommendations: the significance and presentation of recommendations. Journal of clinical epidemiology. 2013; 66:719–25. https://doi.org/10.1016/j.jclinepi.2012.03.013 PMID: 23312392
10. Singh Ospina N, Rodriguez-Gutierrez R, Brito JP, Young WF Jr., Montori VM. Is the endocrine research pipeline broken? A systematic evaluation of the Endocrine Society clinical practice guidelines and trial registration. BMC Med. 2015; 13:187. https://doi.org/10.1186/s12916-015-0435-z PMID: 26265226

11. Qaseem A, Snow V, Owens DK, Shekelle P. The development of clinical practice guidelines and guidance statements of the American College of Physicians: summary of methods. Ann Intern Med. 2010; 153:194–9. https://doi.org/10.7326/0003-4819-153-3-201008030-00010 PMID: 20679562

12. Ross DS, Burch HB, Cooper DS, Greenlee MC, Laurberg P, Maia AL, et al. 2016 American Thyroid Association Guidelines for Diagnosis and Management of Hyperthyroidism and Other Causes of Thyrotoxicosis. Thyroid. 2016; 26:1343–421. https://doi.org/10.1089/thy.2016.0229 PMID: 27521067

13. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, et al. 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. Thyroid. 2016; 26:1–133. https://doi.org/10.1089/thy.2015.0020 PMID: 26462967

14. Alexander EK, Pearce EN, Brent GA, Brown RS, Chen H, Dosiou C, et al. 2017 Guidelines of the American Thyroid Association for the Diagnosis and Management of Thyroid Disease During Pregnancy and the Postpartum. Thyroid. 2017; 27:315–89. https://doi.org/10.1089/thy.2016.0457 PMID: 28056690

15. Jonklaas J, Bianco AC, Bauer AJ, Burman KD, Cappola AR, Celi FS, et al. Guidelines for the treatment of hypothyroidism: prepared by the american thyroid association task force on thyroid hormone replacement. Thyroid. 2014; 24:1670–751. https://doi.org/10.1089/thy.2014.0028 PMID: 25266247

16. Smallridge RC, Ain KB, Asa SL, Bible KC, Brierley JD, Burman KD, et al. American Thyroid Association guidelines for management of patients with anaplastic thyroid cancer. Thyroid. 2012; 22:1104–39. https://doi.org/10.1089/thy.2012.0302 PMID: 23130564

17. Murad MH, Wang Z. Guidelines for reporting meta-epidemiological methodology research. Evid Based Med. 2017; 22:139–42. https://doi.org/10.1136/ebmed-2017-110713 PMID: 28701372

18. Jonklaas J, Bianco AC, Bauer AJ, Burman KD, Cappola AR, Celi FS, et al. Guidelines for the treatment of hypothyroidism: prepared by the american thyroid association task force on thyroid hormone replacement. Thyroid. 2014; 24:1670–751. https://doi.org/10.1089/thy.2014.0028 PMID: 25266247

19. Spencer-Bonilla G, Singh Ospina N, Rodriguez-Gutierrez R, Brito JP, Iniguez-Ariza N, Tamhane S, et al. Systematic reviews of diagnostic tests in endocrinology: an audit of methods, reporting, and performance. Endocrine. 2017; 57:18–34. https://doi.org/10.1007/s12020-017-1298-1 PMID: 28585154

20. Guyatt GH, Schunemann HJ, Djulbegovic B, Akl EA. Guideline panels should not GRADE good practice statements. Journal of clinical epidemiology. 2015; 68:597–600. https://doi.org/10.1016/j.jclinepi.2014.12.011 PMID: 25660962

21. ClinicalTrials.gov.

22. Hazlehurst JM, Armstrong MJ, Sherlock M, Rowe IA, O'Reilly MW, Franklyn JA, et al. A comparative quality assessment of evidence-based clinical guidelines in endocrinology. Clin Endocrinol (Oxf). 2013; 78:183–90.

23. Nagler EV, Ball S. Quality assessment of clinical practice guidelines: who watches the watchmen? Clin Endocrinol (Oxf). 2013; 78:181–2.

24. Murad MH, Montori VM, Sidawy AN, Ascher E, Meissner MH, Chaikof EL, et al. Guideline methodology of the Society for Vascular Surgery including the experience with the GRADE framework. Journal of Vascular Surgery. 2011; 53:1375–80. https://doi.org/10.1016/j.jvs.2011.01.036 PMID: 21575755

25. Khan AR, Khan S, Zimmerman V, Baddour LM, Tleyjeh IM. Quality and Strength of Evidence of the Infectious Diseases Society of America Clinical Practice Guidelines. Clinical Infectious Diseases. 2010; 51:1147–56. https://doi.org/10.1086/656735 PMID: 20946067

26. Tricoci P, Allen JM, Kramer JM, Califf RM, Smith SC Jr. Scientific evidence underlying the ACC/AHA clinical practice guidelines. Jama. 2009; 301:831–41. https://doi.org/10.1001/jama.2009.205 PMID: 19244190

27. Alexander PE, Bero L, Montori VM, Brito JP, Stoltzfus R, Djulbegovic B, et al. World Health Organization recommendations are often strong based on low confidence in effect estimates. Journal of clinical epidemiology. 2014; 67:629–34. https://doi.org/10.1016/j.jclinepi.2013.09.020 PMID: 24388966

28. Brito JP, Domecq JP, Murad MH, Guyatt GH, Montori VM. The Endocrine Society guidelines: when the confidence cart goes before the evidence horse. J Clin Endocrinol Metab. 2013; 98:3246–52. https://doi.org/10.1210/jc.2013-1814 PMID: 23783104

29. Agoritsas T, Merglen A, Heen AF, Kristiansen A, Neumann I, Brito JP, et al. UpToDate adherence to GRADE criteria for strong recommendations: an analytical survey. BMJ open. 2017; 7:e018593. https://doi.org/10.1136/bmjopen-2017-018593 PMID: 29150475

30. Checketts JX, Gordon J, Crawford JH, Adams H, Duckett L, Vassar M. Is the Right Research Being Conducted to Advance Knowledge about Breast Reconstruction? An Analysis of the Research Pipeline. Plast Reconstr Surg. 2018; 141:566–77. https://doi.org/10.1097/PRS.0000000000004107 PMID: 29481388

31. Aschebrook-Kilfoy B, Schechter RB, Shih YC, Kaplan EL, Chiu BC, Angelos P, et al. The clinical and economic burden of a sustained increase in thyroid cancer incidence. Cancer Epidemiol Biomarkers Prev. 2013; 22:1252–9. https://doi.org/10.1158/1055-9965.EPI-13-0242 PMID: 23677575

32. van Vollenhoven RF, Severens JL. Observational studies: a valuable source for data on the true value of RA therapies. Clin Rheumatol. 2011; 30 Suppl 1:S19–24.

33. Song JW, Chung KC. Observational studies: cohort and case-control studies. Plast Reconstr Surg. 2010; 126:2234–42. https://doi.org/10.1097/PRS.0b013e3181f44abc PMID: 20697313

34. Hernán MA, Robins JM. Estimating causal effects from epidemiological data. Journal of epidemiology and community health. 2006; 60:578–86. https://doi.org/10.1136/jech.2004.029496 PMID: 16790829

35. Hernán MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? Epidemiology (Cambridge, Mass). 2006; 17:360–72.

36. Hernán MA. The C-Word: Scientific Euphemisms Do Not Improve Causal Inference From Observational Data. American journal of public health. 2018; 108:616–9. https://doi.org/10.2105/AJPH.2018.304337 PMID: 29565659

37. Glass TA, Goodman SN, Hernán MA, Samet JM. Causal inference in public health. Annual review of public health. 2013; 34:61–75. https://doi.org/10.1146/annurev-publhealth-031811-124606 PMID: 23297653

38. GORDON T, KANNEL WB, HJORTLAND MC, McNAMARA PM. Menopause and Coronary Heart Disease: The Framingham Study. Annals of Internal Medicine. 1978; 89:157–61. https://doi.org/10.7326/0003-4819-89-2-157 PMID: 677576

39. Manson JE, Hsia J, Johnson KC, Rossouw JE, Assaf AR, Lasser NL, et al. Estrogen plus Progestin and the Risk of Coronary Heart Disease. New England Journal of Medicine. 2003; 349:523–34. https://doi.org/10.1056/NEJMoa030808 PMID: 12904517

40. Shikata S, Nakayama T, Noguchi Y, Taji Y, Yamagishi H. Comparison of effects in randomized controlled trials with observational studies in digestive surgery. Ann Surg. 2006; 244:668–76. https://doi.org/10.1097/01.sla.0000225356.04304.bc PMID: 17060757

41. Mansournia MA, Higgins JP, Sterne JA, Hernán MA. Biases in Randomized Trials: A Conversation Between Trialists and Epidemiologists. Epidemiology (Cambridge, Mass). 2017; 28:54–9.

42. González-González JG, Dorsey-Treviño EG, Alvarez-Villalobos N, Barrera-Flores FJ, Díaz González-Colmenero A, Quintanilla-Sánchez C, et al. Trustworthiness of randomized trials in endocrinology—A systematic survey. PLOS ONE. 2019; 14:e0212360. https://doi.org/10.1371/journal.pone.0212360 PMID: 30779814

43. Cancer Epidemiology and Prevention. 4 ed. New York: Oxford University Press; 2017.

44. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. Lancet. 2009; 374:86–9. https://doi.org/10.1016/S0140-6736(09)60329-9 PMID: 19525005

45. Jones AP, Conroy E, Williamson PR, Clarke M, Gamble C. The use of systematic reviews in the planning, design and conduct of randomised trials: a retrospective cohort of NIHR HTA funded trials. BMC Medical Research Methodology. 2013; 13:50. https://doi.org/10.1186/1471-2288-13-50 PMID: 23530582

46. Chalmers I, Bracken MB, Djulbegovic B, Garattini S, Grant J, Gulmezoglu AM, et al. How to increase value and reduce waste when research priorities are set. Lancet. 2014; 383:156–65. https://doi.org/10.1016/S0140-6736(13)62229-1 PMID: 24411644

47. SA M., CS E., HB R., HJ V., KP A., PE N., et al. American Thyroid Association Guidelines and Statements: Past, Present, and Future. Thyroid. 2018; 28:692–706. https://doi.org/10.1089/thy.2018.0070 PMID: 29698130