Boffin Access
FOSTERING KNOWLEDGE

Open Access Journal | Volume: 3.2

International Journal of Surgical Procedures
ISSN 2517-7354

Quantified Congruence for Cancer Treatment Recommendations from Various Countries

Franz Porzsolt[1]
CJ Rhoads[2]
Robert M. Kaplan[3]

[1]Institute of Clinical Economics (ICE) e.V., 89081 Ulm, Germany
[2]Kutztown University, College of Business, Kutztown, PA 19530, United States
[3]Stanford University, Clinical Excellence Research Center, Stanford, CA 94305-6015, United States

Abstract

Clinical guidelines are an accepted tool to describe standards in health care. They help patients and health-care providers make clinical decisions. But guidelines and the recommendations within them vary.

Aims: 1) to determine if guidelines from different countries are congruent (similar) or incongruent (different), 2) if found to be incongruent, to increase the awareness of the differences, and 3) to discuss potential reasons, consequences, and solutions.

Methods: We analyzed guidelines for three different types of cancer (colon cancer, n = 10; gastric cancer, n = 8; and pancreatic cancer, n = 6) from 11 countries. In total 330 recommendations were extracted from the 24 eligible guidelines. Corresponding recommendations were categorized as congruent, incongruent, or undetermined. A congruent recommendation matched 66% of other country's recommendations, an incongruent matched less than 66%, and an undetermined recommendation did not clearly provide a recommendation in at least 66% of the modalities.

Results: Our results indicate that incongruent recommendations were 4-fold more common than congruent recommendations. For gastric cancer (n = 209 recommendations): only 7% were congruent; 86% were incongruent and 7% were undetermined. For pancreatic cancer (n = 60): 35% congruent, 18% incongruent, and 47% undetermined. For colon cancer (n = 61): 25%, 36%, and 39%. Out of the 330 recommendations, only 50 were congruent (15%).

Discussion: Corresponding analyses should be completed in other medical fields. If similar results are confirmed three reason of the incongruence should be considered; different study results, different interpretations of identical studies, and/or difference in weight given to studies versus experience.

Introduction

For about the past 40 years, clinical scientists have attempted to standardize the management of health care by introducing standard clinical guidelines. In some countries this is known as the "Standard of Care". The idea is for a respected group of knowledgeable doctor to review the available evidence, and determine the treatment recommendations based on that evidence.

These recommendations exert strong influence by setting standards. They influence doctors' decisions, patient management, and provide the basis for potential court decisions.

Unfortunately, there is a lot of variation in the quality of the available evidence. Despite this fact, in this century of evidence-based medicine and randomized controlled trials (RCTs), health-care professionals expect clinical guidelines to be valid and reliable (whether they actually are or not). Several groups have suggested that published clinical guidelines are not as useful, practicable or available as they should be [1,2].

In the literature, comparisons of different guidelines from different countries for the treatment of gastric cancer [3], respiratory problems [4], hepatocellular carcinoma [5], and carotid artery stenting [6], have demonstrated differences in their recommendations. By looking at a wide variety of recommendations in one narrow field (cancer), this study investigates the contention that there is too much variability in treatment guidelines to be able to rely upon them as-is.

This study identifies and quantifies the congruence of recommendations for cancer treatment from different countries. Potential reasons for the incongruence are discussed as well as potential opportunities to reduce the causes of the incongruence.

Methods

This study selects eleven guidelines from different countries on three types of cancer.

Article Information

DOI:	10.31021/ijsp.20203136
Article Type:	Research article
Journal Type:	Open Access
Volume:	3 **Issue:** 2
Manuscript ID:	IJSP-3-136
Publisher:	Boffin Access Limited
Received Date:	22 Mar 2020
Accepted Date:	08 June 2020
Published Date:	10 June 2020

*Corresponding author:

Franz Porzsolt
Institute of Clinical Economics (ICE) e.V.
Schwarzenbergstrasse 135
890981 Ulm / Germany
Tel: +49 173 329 9774
E-Mail: f.porzsolt@gmx.de

Citation: Porzsolt F, Rhoads CJ, Kaplan RM. Quantified Congruence for Cancer Treatment Recommendations from Various Countries. Int J Surg Proced. 2020 Jun;3(2):136.

Copyright: © 2020 Porzsolt F, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 international License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

The provided recommendations were extracted, categorized and their congruence was quantified.

Selection of guidelines

Using the process described in Supplement 1 Search Guideline Search Flow Diagram, guidelines were found from ten countries and one international group for gastric cancer, pancreatic cancer and colon cancer. We searched the databases Medline, Embase and The Cochrane Library. Additionally, several websites [7-10] were searched for relevant publications, as well as Google Scholar and Metager. We also contacted a member of the Canadian Gastrointestinal Cancer Disease Site Group directly to obtain access to the Canadian guideline for pancreatic cancer. We examined the eligibility of all titles and abstracts of studies identified by electronic or bibliographic scanning in order to identify the initial group of 99 potential guidelines. Only the guidelines that were established by experts in scientific societies and published in English or German language were selected for our analysis. Guidelines without references were excluded. After applying criteria, there were only 24 guidelines that met the selection criteria.

Quantification of the congruence of guidelines

We collected and processed in a standardized way several types of information provided by each guideline for the quantification of the congruence of the recommendations. After the different types of information were obtained, we combined them into a final statement i.e. a comparable recommendation provided by each guideline. These different types of information were:

- The stage of the disease,
- The content of the recommendation (use treatment "A" or "B" or none),
- A standardized description of the recommendation,
- The interpretation of the result.

Corresponding recommendations were categorized as congruent, incongruent, or undetermined. We felt that two-thirds majority (66%) would suffice to show congruence. A congruent recommendation matched 66% of other country's recommendations, an incongruent matched less than 66%, and an undetermined recommendation did not clearly provide a recommendation in at least 66% of the modalities. The detailed description of the rules for quantification of the congruence including examples is shown in Supplement 37.

Results

The search resulted in a total of 99 guidelines (30 on gastric cancer, 29 on pancreatic cancer, 40 on colon cancer) from 11 countries (i.e. 10 countries and one European). The guidelines were reviewed, reduced, and analysed.

Results of the guideline search

Only 24 of the 99 guidelines provided specific enough recommendations for each diagnosis and stage of disease. These 24 eligible guidelines (eight on gastric cancer, six on pancreatic cancer, ten on colon cancer) are listed by diseases, countries, scientific organisations, and the year in Table 1.

Appraisal of congruence

From these 24 guidelines 330 recommendations were extracted, grouped by type of cancer and stage of the disease which resulted in 34 sets of recommendations. The details are shown in 34 documents included in the Supplements 2 through 35. The procedure to identify congruent, incongruent, an undetermined guidelines (including examples) is described in Supplement 36. The summary of the resulting frequencies of the 330 classified recommendations is shown in Table 2.

This summary table shows large variations among the results of recommendations and among the different types of cancer. By looking at the underlying recommendations by types and stages of cancer, additional information can be discerned. While most recommendations for surgery and for early stages of the diseases were congruent, larger differences were observed in the recommendations of systemic treatments (chemotherapy with/without radiotherapy). Also, there was a high number of undetermined recommendations for most late stage diseases (Supplements 1-37).

Discussion

The aim of this study is to investigate the congruence of recommendations of guidelines for treatment of different stages of three different diseases in different countries around the world. Only 15% (50 of the 330) of recommendations were congruent. Of the remaining recommendations, 65% were incongruent and 20% were undetermined.

This observation may be related to the issue of availability of high quality evidence-based recommendations and the resulting outcomes based on real world evidence. Previous attempts to confirm the correlation of adherence with guidelines and outcomes of care have demonstrated difficulties [11,12], perhaps due to low quality evidence. O'Sullivan et al. [13] found that the quality of evidence used correlated with higher adherence to guideline recommendations [13].

No one expects all guidelines to match exactly; the existence of differences has always been known but has been assumed to be a "tolerable" level of variation for cultural and economic reasons. The high variation of guideline recommendations cannot be used as indicator of poor quality recommendations because there are too many reasons that can explain the variation. However, there may be some reasons that can be eliminated, and it would be beneficial to identify them.

Evidence-based reasons for incongruence

There may be three reasons related to the evidence to explain the incongruence in the recommendations:

1. Lack of agreement among the results in clinical studies on the same diagnosis and stage.
2. Differences in the interpretation of studies with similar results, and
3. Design requirements of experimental or observational studies that demonstrate either efficacy or effectiveness (but not both).

This list of reasons isn't comprehensive. But for purposes of this discussion, these are the main reasons to be addressed.

Lack of agreement among the results of clinical studies

The editors of a guideline may have used different studies that observed and described different results even though the same study question was investigated. Lack of agreement in the results of different trials that study the same diagnosis and stage is common in healthcare research. There are many potential reasons for the lack of agreement. There might be small differences in a number of things; the study questions, in the defined diagnoses and stages, the selected study populations; and/or the methods used for assessment of the selected treatments.

The amount of time that teams of experienced clinicians and methodologists have to contribute for the critical appraisal of clinical trials from different expert perspectives might be important. This information was quantified in an earlier study [14].

Meta-analyses and Health Technology Assessment (HTA) Reports are completed by large institutions to summarize and harmonize the results of corresponding studies. These meta-data and the results from single studies provide different sources of support for the wording of the recommendations in clinical guidelines, and they don't always agree. Although there are frameworks designed to minimize this variation such as GRADE [15], not all guideline developers

	Publishing Scientific Community*		
	Gastric cancer	Pancreatic cancer	Colon cancer
Belgium	KCE, 2008	KCE, 2009	n/a
Canada	AHS, 2013	CCO, 2011/2013	CCO, 2008
Europe	ESMO, 2010	ESDO, 2012	ESMO, 2010
Germany	AWMF, 2012	AWMF, 2013	AWMF, 2013
Great Britain	BSG/BASO, 2011	n/a	NICE, 2011
Japan	JGCA, 2010	n/a	JSCCR, 2012
Netherlands	n/a	n/a	IKNL, 2008
New Zealand	n/a	n/a	NZGG, 2011
Scotland	SIGN, 2006	n/a	SIGN, 2011
Spain	n/a	SEOM, 2011	CCOG, 2010
USA	NCCN, 2013	NCCN, 2013	NCCN, 2014

Table 1: Description of 24 clinical guidelines

Note: *The description of the organisations for the acronyms can be found in supplement 36. Guideline producing organisations.

	Gastric cancer	Pancreatic cancer	Colon cancer	Total
Congruent	14 (7%)	21 (35%)	15 (25%)	50 (15%)
Incongruent	180 (86%)	11 (18%)	22 (36%)	213 (65%)
Undetermined	15 (7%)	28 (47%)	24 (39%)	67 (20%)
Total	209 (100%)	60 (100%)	61 (100%)	330 (100%)

Table 2: Recommendation classification frequencies.

adhere to these frameworks. Furthermore, recommendations within guidelines cannot be selected without some subjective decisions [16,17] and those subjective decisions produce variations that lead to incongruence in treatment recommendations.

Differences in the interpretation of studies with similar results

Another reason may be that two different editors may have used the same identical study to support their recommendations, but each interpreted the results differently. Different interpretations may be caused by differences in the perceived validity of a study. While one researcher reads a study and concludes there is evidence that the treatment should be recommended, another might only say that it should be considered (i.e. a WEAK recommendation which is not the same as a clear FOR recommendation).

Even strict quality criteria for the design, conduct, and evaluation of clinical trials cannot guarantee a predictable interpretation of the results. Different researchers reading the results may view the conclusions differently.

For example, the use of adjuvant therapy with FOLFOX in stage-IIa colon cancer patients without additional risk-factors was discussed by the Spanish and the American guidelines by quoting the same two studies [18,19]. The Spanish guideline recommends the use of FOLFOX because of a significant lengthening of the 6-year survival rate, while the American guideline "does not consider FOLFOX [...] to be an appropriate adjuvant therapy option for patients with stage II disease without high-risk features". A second example is related to the adjuvant therapy with FLOX in stage-III colon cancer patients. Apart from the European guideline, all guidelines base their recommendations on the study by Kuebler et al. [20]. Both the American and the Canadian guidelines list FLOX therapy as a possible option for adjuvant therapy in stage-III colon cancer. The German and the Spanish guidelines do not recommend FLOX therapy because of higher toxicity compared to FOLFOX4 treatment. In our study, we identified 13 examples (accounting for 4.0% of total recommendations) in which identical studies resulted in different recommendations.

Design requirement of experimental or observational clinical studies

Sir Archie Cochrane [21] and Sir Austin Bredford Hill [22] requested answers to three questions before a new intervention should be introduced to routine care. "Can it work?", "Does it work?", "Is it worth it?". The first answer reflects the demonstration of efficacy, i.e. can it work under Ideal Study Conditions (ISC). The second answer is expected to reflect the demonstration of effectiveness, i.e. does it work under Real World Conditions (RWC).

Each health care provider makes individual decisions based on their subjective perception, which is a result of available scientific evidence tempered by their real world experience related to three item characteristics; the investigated patients, the interventions, and the outcomes. The reduction of the real world variation by including more RWC studies in the evidence used would help guidelines to be more congruent. In other words, the experience of physicians generated under RWC can more easily be shared if the RWC were transformed to an experimental or observational study design [23,24] so that a broader base of evidence can be utilized when developing guidelines.

There is an important difference between experimental and observational studies; the assignment to groups. In experimental studies the investigator uses random allocation to the treatment groups. Often in observational studies the attending physician selects the treatment group according to medical needs, selecting the treatment most appropriate rather than randomly. Observational studies may provide outcomes closer to RWC because often patients will not agree to randomization (i.e. the chance they won't get their preferred treatment).

Another characteristic is different in experimental and observational studies. Exclusion criteria of ISC means that many potential subjects are excluded from the study, whereas these subjects will be included in observational studies. The latter may provide more RWC results.

Of course, to account for lack of randomization and exclusion criteria, additional data have to be recorded and an appropriate study design has to be applied that will enable the unbiased presentation of outcomes. This may increase study costs and brings added challenge. The challenge is to test real world effectiveness of the treatment using an observational tool that can collect the necessary information without excluding the majority of eligible patients under ISC.

The optimum process of a new intervention would be to first test using experimental efficacy studies under ISC. These studies would confirm the proof of principle. Then effectiveness studies would confirm that the benefit observed under ISC can also be demonstrated under RWC. RWC includes patients with comorbidities, co-treatments or other conditions that excluded them from efficacy studies.

Utilizing both ISC and RWC would provide confirmation of the patient benefit. Relying upon ISC alone has sometimes resulted in increased treatment for no discernable patient benefit. For example, the change for definition of hypertension from >140/90 to >130/80 increased the treated population but in the real world, has not necessarily demonstrated any patient benefit [25]. Similar concerns were published for primary care [26].

One type of observational study derived from the Bayes Theorem has been developed that stratifies patients under RWC according to their endpoint-related risk profiles for each of the assessed endpoints. Attending physicians select the treatment according to standard care

guidelines and individual patient characteristics. The patients are put into treatment groups based on their profile. The study algorithm compares the endpoints of patients in different treatments groups but with identical baseline risks based on their profile [27,28].

Propensity score matching has confirmed the expected advantage of the endpoint-related risk allocation method [29]. If these types of observational studies are included in the development of future guidelines, they will benefit because two different outcomes, efficacy and effectiveness, would be included. Future guidelines may become more congruent.

Strengths and limitations of this study

The strength of this study is demonstrating the quantification of the incongruence of recommendations in cancer treatment guidelines from different countries. Without quantification most colleagues largely overestimate the congruence of guideline recommendations.

The strength, (domain knowledge), is also a limitation of the study. It points out the subjective nature of the currently used strategy for development of guidelines. High variability cannot be avoided when objective efficacy data derived from experimental studies is combined with subjective effectiveness data derived from personal clinical experience. This knowledge leads to the conclusion that the standardized, reliable, and valid assessment of both ideal study conditions and real world effectiveness should be included.

For further confirmation, these results would need to be reproduced by other groups in other fields of medicine.

Conclusion

In conclusion, we found a large amount of incongruence among the recommendations from guidelines from different countries. Potential reasons for this incongruence are: 1) Lack of agreement among the results in clinical studies on the same diagnosis and stage. 2) Differences in the interpretation of studies with similar results, and 3) Design requirements of experimental or observational studies that demonstrate either efficacy or effectiveness (but not both). It would be difficult to eliminate the first two reasons. The most likely solution to the incongruence would be for guideline developers to seek out unbiased real world (effectiveness) data.

Data sharing statement

All data relevant to the study are included in the article or uploaded as supplementary information.

Ethics approval and consent to participate

Not applicable. Neither the patients nor the public has been involved in this study.

Competing interests

The authors declare that they have no competing interests.

Funding statement

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Author´s contributions

Porzsolt F and Kaplan RM contributed substantial to conception and design of the study, Rhoads CJ and Porzsolt F contributed in analysis and interpretation of the data, Porzsolt F and Kaplan RM drafted the article, Rhoads CJ critically revised the manuscript. All authors gave the final approval of the version to be published.

References

1. Elovainio M, Mäkelä M, Sinervo T, Kivimäki M, Eccles M, et al. Effects of job characteristics, team climate, and attitudes towards clinical guidelines. Scand J Public Health. 2000 Jun;28(2):117-122.
2. Ibañez J, Arikan F, Pedraza S, Sánchez E, Poca MA, et al. Reliability of clinical guidelines in the detection of patients at risk following mild head injury: results of a prospective study. J Neurosurg. 2004 May;100:825-834.
3. Bauer K, Schroeder M, Porzsolt F, Henne-Bruns D. Comparison of international guidelines on the accompanying therapies for gastric cancer. Reasons for the differences. J Gastric Cancer. 2015 Mar;15(1):10-18.
4. Johnson AM, Smith SM. Respiratory clinical guidelines inform ward-based nurses' clinical skills and knowledge required for evidence-based care. Breathe (Sheff). 2016;12:257-266.
5. Manzini G, Henne-Bruns D, Porzsolt F, Kremer M. Is there a standard for surgical therapy of hepatocellular carcinoma? A comparison of eight international guidelines. BMJ Open Gastroenterol. 2017 Mar;24:4(1):e000129.
6. White CJ. Carotid Artery Stenting. JACC 2014;64(7):722-731.
7. AWMF online. Das Portal der wissenschaftlichen Medizin.
8. AWMF online. Das Portal der wissenschaftlichen Medizin.
9. Guidelines International Network.
10. Agency for Healthcare.
11. Sekercioglu N, Al-Khalifah R, Ewusie JE, Elias RM, Thabane L, et al. A critical appraisal of chronic kidney disease mineral and bone disorders clinical practice guidelines using the AGREE II instrument. Int Urol Nephrol. 2017 Feb;49(2):273-284.
12. Boudoulas KD, Leier CV, Geleris P, et al. The shortcomings of clinical practice guidelines. Cardiology 2015;130(3):187-200.
13. O'Sullivan JW, Albasri A, Koshiaris C, Aronson JK, Heneghan C, et al. Diagnostic test guidelines based on high-quality evidence had greater rates of adherence: a meta-epidemiological study. J Clin Epidemiol. 2018 Nov;103:40-50.
14. Porzsolt F, Braubach P, Flurschütz PI, Goller A. Medical Students Help Avoid the Expert Bias in Medicine. Creative Education. 2012;3(06):1115-1121.
15. Schünemann H, Wojtek W, Brozek J, Etxeandia-Ikobaltzeta I, Mustafa RA, et al. GRADE Evidence to Decision (EtD) frameworks for adoption, adaption, and de novo development of trustworthy recommendations: GRADE-ADOLOPMENT. J Clin Epidemiol. 2017 Jan;81:101-110.
16. Sterne JAC, Egger M. Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis. J Clin Epidemiol. 2001 Oct;54(10):1046-1055.
17. Lin L, Chu H, Hodges JS. Sensitivity to Excluding Treatments in Network Meta-analysis. Epidemiology. 2016 Jul;27(4):562-569.
18. André T, Boni C, Mounedji-Boudiaf L, Navarro M, Tabernero J, et al. Multicenter International Study of Oxaliplatin/5-Fluorouracil/ Leucovorin in the Adjuvant Treatment of Colon Cancer (MOSAIC) Investigators: Oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment for colon cancer. The New England Journal of Medicine. 2004;350:2343-2351.
19. Figer A, Perez-Staub N, Carola E, Tournigand C, Lledo G,et al. FOLFOX in patients aged between 76 and 80 years with metastatic colorectal cancer: an exploratory cohort of the OPTIMOX1 study. Cancer. 2007 Dec;110(12):2666-2671.
20. Kuebler JP, Wieand HS, O'Connell MJ, et al. Oxaliplatin combined with weekly bolus fluorouracil and leucovorin as surgical adjuvant chemotherapy for stage II and III colon cancer: results from NSABP C-07. J Clin Oncol. 2007 Jun;25(16):2198-2204.
21. Cochrane AL. Effectiveness and Efficiency: Random Reflections on the Health Services. London: Nuffield Provincial Hospital Trust1972. Control Clin Trials. 1989;10:428-33.
22. Horton R. Common sense and figures: the rhetoric of validity in medicine [Bredford Hill Memorial Lecture 1999]. Stat Med. 2000 Dec;19:3149-3164.
23. Grimes DA, Schulz KF. An overview of clinical research: the lay of the land. Lancet. 2002 Jan;359(9300):57-61.

24. Thiese MS. Observational and interventional study design types; an overview. Bioche Med. 2014 Jun;24(2):199-210.

25. Ionannidis JPA. Diagnosis and Treatment of Hypertension in the 2017 ACC/AHA Guidelines and in the Real World. JAMA. 2018 Jan;319(2):115-116.

26. Phillips D. Most primary care guidelines low quality. Medscape Medical News. 2019.

27. Porzsolt F, Eisemann M, Habs M, Wyer P. Form Follows Function: Pragmatic Controlled Trials (PCTs) have to answer different questions and require different designs than Randomized Controlled Trials (RCTs). J Publ Health. 2013 Jun;21(3):307-313.

28. Porzsolt F, Rocha NG, Toledo-Arruda AC, Thomaz TG, Moraes C, et al. Efficacy and Effectiveness Trials Have Different Goals, Use Different Tools, and Generate Different Messages. Pragmat Obs Res. 2015;6:47-54.

29. Ferdinand D, Otto M, Weiss C. Get the most from your data: a propensity score model comparison on real-life data. Int J Gen Med. 2016:9 123–131.