



Analysis of Not Structurable Oncological Study Eligibility Criteria for Improved Patient-Trial Matching

Julia Dieter^{1,*} Friederike Dominick^{1,*} Alexander Knurr¹ Janko Ahlbrandt¹ Frank Ückert¹

¹Department of Medical Informatics for Translational Oncology, German Cancer Research Center, Heidelberg, Germany

Address for correspondence Julia Dieter, PhD, Department of Medical Informatics for Translational Oncology, German Cancer Research Center, Heidelberg 69120, Germany (e-mail: j.dieter@dkfz-heidelberg.de).

Methods Inf Med 2021;60:9–20.

Abstract

Background Higher enrolment rates of cancer patients into clinical trials are necessary to increase cancer survival. As a prerequisite, an improved semiautomated matching of patient characteristics with clinical trial eligibility criteria is needed. This is based on the computer interpretability, i.e., structurability of eligibility criteria texts. To increase structurability, the common content, phrasing, and structuring problems of oncological eligibility criteria need to be better understood.

Objectives We aimed to identify oncological eligibility criteria that were not possible to be structured by our manual approach and categorize them by the underlying structuring problem. Our results shall contribute to improved criteria phrasing in the future as a prerequisite for increased structurability.

Methods The inclusion and exclusion criteria of 159 oncological studies from the Clinical Trial Information System of the National Center for Tumor Diseases Heidelberg were manually structured and grouped into content-related subcategories. Criteria identified as not structurable were analyzed further and manually categorized by the underlying structuring problem.

Results The structuring of criteria resulted in 4,742 smallest meaningful components (SMCs) distributed across seven main categories (Diagnosis, Therapy, Laboratory, Study, Findings, Demographics, and Lifestyle, Others). A proportion of 645 SMCs (13.60%) was not possible to be structured due to content- and structure-related issues. Of these, a subset of 415 SMCs (64.34%) was considered not remediable, as supplementary medical knowledge would have been needed or the linkage among the sentence components was too complex. The main category “Diagnosis and Study” contained these two subcategories to the largest parts and thus were the least structurable. In the inclusion criteria, reasons for lacking structurability varied, while missing supplementary medical knowledge was the largest factor within the exclusion criteria.

Conclusion Our results suggest that further improvement of eligibility criterion phrasing only marginally contributes to increased structurability. Instead, physician-based confirmation of the matching results and the exclusion of factors harming the patient or biasing the study is needed.

Keywords

- ▶ oncological eligibility criteria
- ▶ structurability
- ▶ patient-trial matching

* *Equally contributing authors.*

received
August 19, 2020
accepted
December 8, 2020
published online
April 22, 2021

DOI <https://doi.org/10.1055/s-0041-1724107>.
ISSN 0026-1270.

© 2021. The Author(s).

This is an open access article published by Thieme under the terms of the Creative Commons Attribution-NonDerivative-NonCommercial-License, permitting copying and reproduction so long as the original work is given appropriate credit. Contents may not be used for commercial purposes, or adapted, remixed, transformed or built upon. (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)
Georg Thieme Verlag KG, Rüdigerstraße 14, 70469 Stuttgart, Germany

Introduction

Clinical oncological trials are needed to improve the cancer treatment.^{1–3} Suited patients for these trials are determined by matching their characteristics with the trial eligibility criteria. Originally, this was a manual process performed by physicians or study nurses. Semiautomated matching algorithms increasingly aid the laborious work. As a prerequisite for this, eligibility criteria (and patient characteristics) need to be computer readable (i.e., “structured”). The higher the proportion of structured criteria is, the higher the degree of automated matching can be. However, not all eligibility criteria can be fully structured. Therefore, we aimed to analyze the underlying reasons that complicate the structuring of oncological eligibility criteria. Understanding these reasons might help to increase the structured portion in the future by improving criteria phrasing.

Within the literature, various approaches for criteria structuring have been reported. These knowledge representations range from content-oriented classifications^{4–6} to expression languages representing eligibility rules (e.g., object-orientated such as an object-oriented expression language [GELLO],⁷ ad hoc expressions such as Eligibility Rule Grammar and Ontology [ERGO]^{5,8}). Naturally, the focus of studies on eligibility criteria structuring was on those criteria that were possible to be structured. For instance, several studies explored the secondary use of electronic health record (EHR) elements for purposes within clinical trials, such as feasibility analysis and semiautomated patient recruitment.^{9–11} However, to increase the proportion of computer-readable criteria, a thorough analysis of the criteria that cannot be structured is equally important. This has been done by Ross et al,¹² characterizing complex criteria in more detail. Within their study, Ross et al¹² manually analyzed 1,000 eligibility criteria without a disease-specific focus from ClinicalTrials.gov regarding their complexity, semantic patterns, clinical content, and data sources. A significant semantic and clinical variability was found across the criteria. About 77% of criteria remained complex and hard to evaluate, as they contained temporal constraints (40%) that were not defined properly in many cases, required a link to study metadata (24%) or clinical judgment (19%), or involved semantic connectors (9%) that are not captured by current representation languages. Additionally, it was remarked that criteria based on radiographic or histologic data (7%) might require further efforts such as natural language processing to be structured and included into automated matching processes. Thus, it has to be considered that not all criteria can be structured (e.g., Ross et al¹²) and therefore cannot be processed in automated matching systems. The authors argued that their analysis of criteria across several medical domains allows a wider range of variance in the eligibility criteria.¹² However, it might also be important to analyze disease-specific eligibility criteria in-depth to cover characteristic variations. To the best of our knowledge, a manual structuring of clinical, oncology-specific criteria texts in huge numbers and a thorough disease-specific analysis of criteria that are not structurable have not been conducted yet. We aimed to assess the proportion of not structurable criteria, the underlying reasons for missing structurability, and possible ways to rephrase eligibility criteria.

Methods

A list of 159 original oncological clinical studies from the local clinical trial information system (Clinical Trial Information System of the NCT [CTIS] of the National Center for Tumor Diseases) was assembled. The studies' recruitment phase began in 2001 or later. They were of interventional type and had the aim of improving oncological treatment. The studies mainly comprised trials of phase II (32.10%), III (25.80%), or unspecified phase (14.50%). Various cancer types are covered; most often multiple myeloma, breast cancer or several different oncological diseases at once (e.g., “Histologically confirmed locally advanced [stage III or IV], nonmetastatic squamous cell carcinoma of oropharynx, hypopharynx, and larynx [T2–4, any N, M0]; TPF-C-HIT study).

We manually structured the criteria and grouped them by their content. We refer to this step as “structuring” (see Structurable SMCs). The portion of not structurable criteria identified hereby was analyzed in detail and categorized by the underlying structuring problem within a second step. We refer to this step as “categorizing” (see Not Structurable SMCs).

The analyses were conducted via Microsoft Office Excel in 2016, R for Windows (version 4.0.3), and RStudio (version 1.2.1335; RStudio Team, 2018. RStudio: Integrated Development for R. RStudio, Inc., Boston, Massachusetts, URL: <http://www.rstudio.com/>). As indicated in the Introduction, “structured” refers to an eligibility criterion that is computer readable. For the purpose of our manual criteria structuring approach, we denoted criteria as “structurable” when they could be represented by a combination of independent, objective, unambiguous, and measurable statements. For instance, the criterion “Platelets <75,000/mm³” can be structured into “Laboratory parameter,” “Laboratory value,” and “Laboratory unit.” Opposed to the smallest meaningful component (SMC, see below), an independent statement (e.g., “laboratory value = 75,000”) does not necessarily make medical sense on its own.

General Criteria Structuring

Most of the 159 studies had already been divided into inclusion as well as exclusion eligibility criteria and were saved in an Excel file as unstructured text.

Three project members (F.D., A.K., J.D.) conducted and reviewed the process of eligibility criteria structuring, which was performed according to the steps described in the following:

In the first step, the criteria were restructured in the Excel sheet into SMCs. Here, an SMC was defined as the smallest, sense-making combination of words that independently represented a contentual unit within one criterion. This step is a prerequisite for criteria structuring, as the latter might be impeded by the presence of two or more different contentual components in one criterion. This approach is similar to previous studies, such as Ateya et al or Köpcke et al,^{13,14} who applied “criteria elemental statements or CES” and “independent patient characteristics” to separate a criterion into its single components. For instance, in our case, the inclusion criterion “(1) Histologically or

cytologically confirmed diagnosis of: epithelial ovarian cancer which is platinum resistant (...), (2) platinum refractory (...), (3) cancer of the fallopian tube, (4) peritoneal cancer (...)" of the PACOVAR study (NCT01238770) has:

1. "Histologically or cytologically confirmed diagnosis of epithelial ovarian cancer which is platinum resistant,"
2. "Histologically or cytologically confirmed diagnosis of epithelial ovarian cancer which is platinum refractory" as well as,
3. "Cancer of the fallopian tube," and
4. "Peritoneal cancer"

as SMCs. Throughout this study, we will use the term "SMC" instead of "criterion" (unless stated otherwise), to underline that our analysis was based on the contentual fragments within the inclusion and exclusion criteria and not on the inclusion and exclusion criteria texts as a whole.

Based on the content identified within the SMCs, main categories and subcategories were formed in iterative cycles in a consensus-driven way between the three group members. Diagnosis, Therapy, Laboratory, Study, Findings, Demographics and Lifestyle, and Others were identified as main categories (see the section "Structurable SMCs"), each potentially containing subcategories of different levels. The subcategories had underlying catalogues where applicable, thereby building a further sublevel. The catalogues were either official such as the ICD10 (International Classification of Diseases, 10th edition), or unofficial (containing grouped items identified within the eligibility texts). **Fig. 1** shows an exemplary excerpt of this structure for Diagnosis.

Thus, an SMC might contain elements of different main categories and subcategories of different levels. Ambiguous assignments to main or subcategories were discussed in the team and assigned to the category evaluated to fit best. A complete overview of main categories and their subcategories as identified in the eligibility criteria is given in the supplement (Supplementary Materials S1–S7, available in the online version).

Based on this hierarchy of main categories and subcategories, a template for criteria structuring was created within Excel (for a schematic excerpt see **Table 1**). It contained columns for main category and subcategory, study name, study ID, the linkage type, and the "structurability." The linkage type (here: "AND" and "OR") indicated the type of linkage between two or more lines of an SMC within the structuring template. The column "Structurability" indicated whether a line of an SMC was either "Structurable" or "Not structurable" (for a definition on "structurability," see the section "Methods"). SMCs that only contained "Examples and headings," such as "No cardiovascular disorder including, but not limited to, any of the following" in the TAVAREC study (EORTC-26091) were not used for further analyses.

It is presented through a structuring example for the exclusion criterion "Active secondary malignancy requiring treatment (except basal cell carcinoma or tumor curatively treated by surgery)" of the T-PLL2 study (NCT01186640). The header of the structuring template is indicated by bold text. Depending on its content, each SMC of the study was entered into one or more lines of the structuring template. Here, the SMCs were assigned to the main categories Diagnosis and Therapy as well as to the respective subcategories. The columns of a specific line were filled whenever applicable. Three lines of the example criterion were structurable, while one line remained not structurable due to the term "active" in the context of a disease, requiring supplementary medical knowledge for correct classification (for a definition on "Supplementary medical knowledge" see the section Not structurable SMCs). The linkage types were numbered within the linkage type column to identify rows belonging together.

Whenever possible and reasonable, the structuring was guided by pre-existing graduations such as global catalogues like the ICD10.¹⁵ Besides, local catalogues from the NCT or from the specifically developed clinical datamodel of the NCT DataThereHouse (DTH) project (located at the German Cancer Research Center) were applied.¹⁶

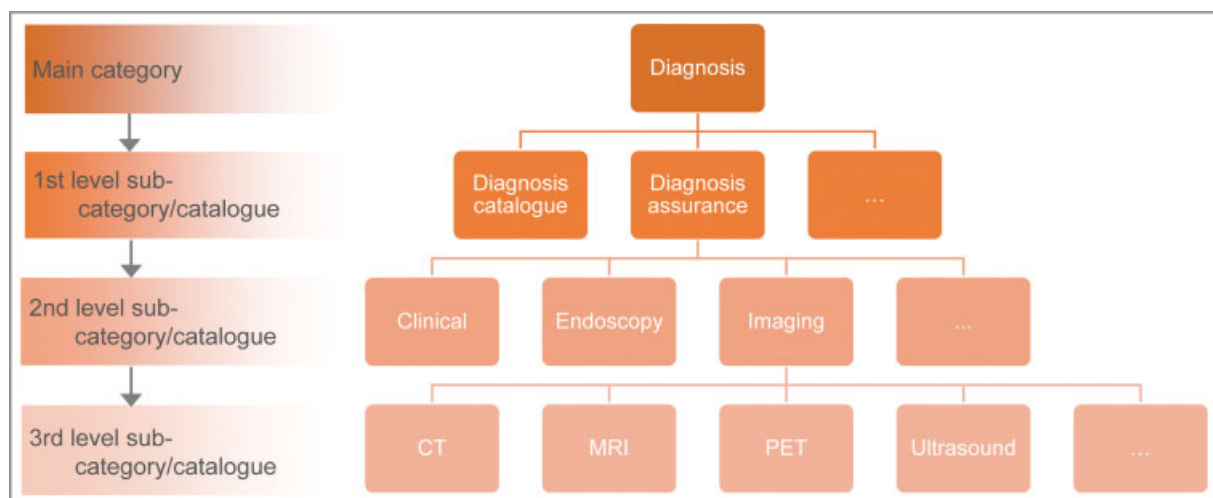


Fig. 1 Excerpt and exemplary structure of the diagnosis main category, subcategories, and catalogues. In the first level subcategory, the official ICD10-catalogue was attached to the diagnosis main category. Another first level element, "diagnosis assurance," had attached the unofficial second level catalogue with items such as "clinical," "endoscopy," "imaging," etc. The latter, in turn, had attached a third level unofficial catalogue. It contained different imaging techniques. ICD10, International Classification of Diseases for Oncology, 10th edition.

Table 1 Schematic excerpt of the Excel structuring template

Study name	Study ID	SMC	Structurability	Linkage type between main categories of following column	Main category	Level depending on main category	Level depending on main category	Level depending on main category
T-PLL2	21	Active secondary malignancy	Not structurable	Diagnosis AND Therapy 1	Diagnosis	ICD-10 (multiple codes)		Diagnosis-related: "Active"
T-PLL2	21	Requiring treatment	Structurable	Diagnosis AND Therapy 1	Therapy			Requiring therapy
T-PLL2	21	Except basal cell carcinoma or tumor	Structurable	Diagnosis AND Therapy 2	Diagnosis	ICD10 (multiple codes)	Excluded diagnosis	
T-PLL2	21	Curatively treated by surgery	Structurable	Diagnosis AND Therapy 2	Therapy	Surgery	Tumor-related	Curative treatment

Abbreviations: ICD10, International Classification of Diseases for Oncology, 10th edition; SMCs, smallest meaningful components.

Not Structurable Criteria

The SMCs that were not possible to be structured by our general structuring approach were collected. We then manually identified the reasons why SMCs were not structurable and grouped these reasons into categories and subcategories (conducted by team member F.D. and reviewed by the two other group members A.K. and J.D.). It was then analyzed, which categories of the identified structuring problems might be remedied, how huge this proportion would be, and which measures might be taken to enable their structuring. Ambiguities in any working step were discussed; arguments were considered carefully and then determined in accordance within the team.

Referring back to our example in **Table 1**, the not structurable SMC "Active secondary malignancy" was assigned to the category "Additional information beyond criterion text necessary" and the subcategory "Supplementary medical knowledge necessary." The reason therefore was that the description of a disease as "active" is not measurable or objective. Only a physician can judge whether the course of a patient's disease is suitable for a clinical trial. The information necessary for this decision cannot be deposited in the text of an eligibility criterion.

Results

Terms

Structurable SMCs

The first step, i.e., the structuring of the eligibility criteria texts, resulted in the formation of SMCs, out of which seven main categories with several subcategories were created, i.e., Diagnosis (with 20 subcategories), Therapy (28), Laboratory (13), Study (4), Findings (18), Demographics and Lifestyle (9), and Others (1). Hence, we refer to this first step as "structuring" (i.e., to represent an eligibility criterion by a combination of independent, objective, unambiguous, and measurable statements; for a definition see Methods).

Not Structurable SMCs

In the second step, SMCs that were not possible to be structured were categorized by the underlying reason for that. This resulted in structural and contentual categories with altogether eight subcategories. We refer to this second step as "categorizing."

Basic Statistics

In total, 159 studies were manually structured. This resulted in 4,742 SMCs overall (2,260 in the inclusion criteria and 2,482 in the exclusion criteria). The proportions of the main categories, the not structurable SMCs, and the SMCs containing only examples or headings are shown in **Fig. 2**.

The highest number of SMCs was assigned to the main category of Diagnosis (27.20%), which comprised subcategories such as diagnoses (e.g., ICD10, International Classification of Diseases for Oncology, 3rd edition [ICD-O-3]), classifications (e.g., World Health Organization [WHO], Ann Arbor, Union International Contre le Cancer), gradings

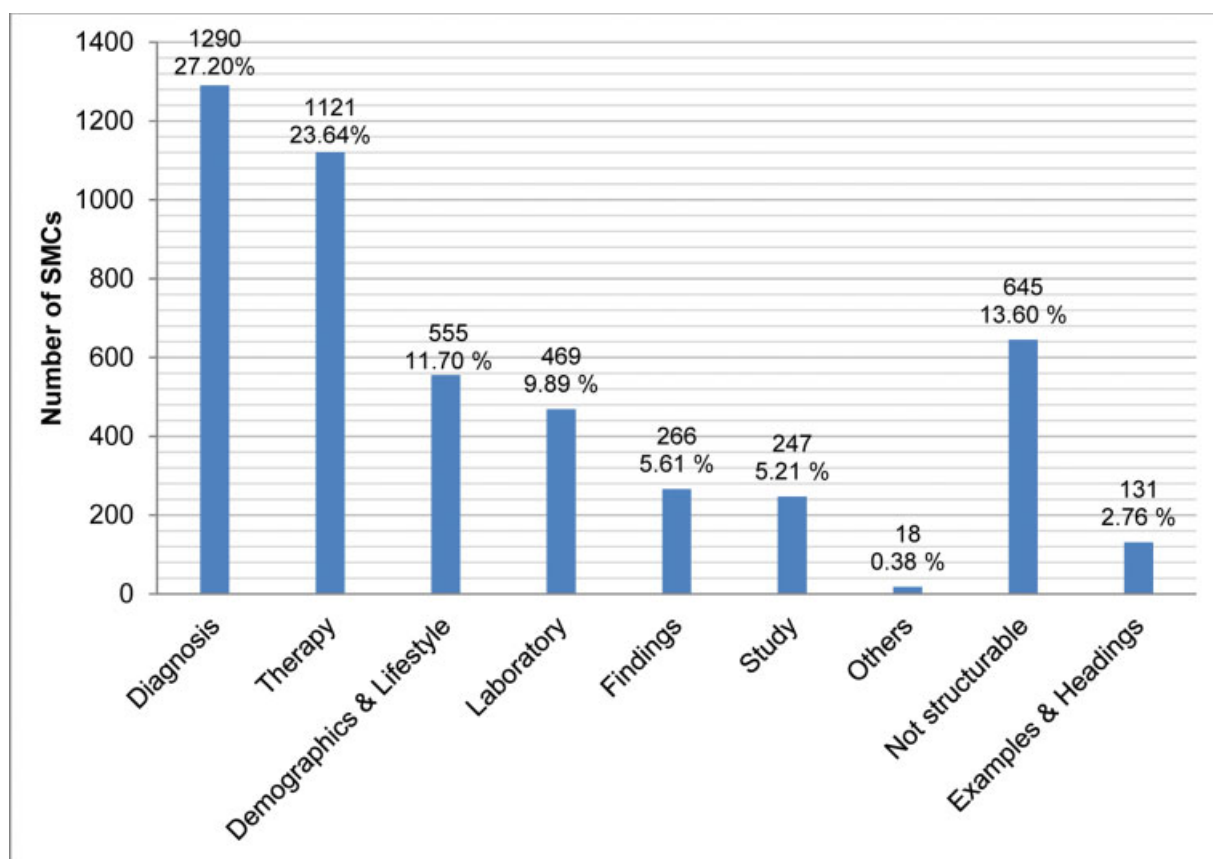


Fig. 2 Categorization of all 4,742 SMCs. The main categories (seven leftmost bars in the Figure) contained structurable SMCs. The amount of not structurable SMCs was 645 (13.60%). The SMCs within “Examples and Headings” were not considered in our study. The percentages given refer to the proportion on the overall number of SMCs, i.e., 4,742. SMC, smallest meaningful component.

(e.g., Gleason Score), or remission status (e.g., complete remission, partial remission).

Therapy was the second largest main category (23.64%) and comprised subcategories such as therapy methods (e.g., chemotherapy, radiotherapy), therapy types (e.g., first-line, adjuvant), substances (e.g., doxorubicin), and substance classes (e.g., anthracyclines), or doses (e.g., “ \leq ,” “ \geq ”).

The main category Demographics and Lifestyle (11.70%) contained subcategories such as patient data (e.g., ethnicity, age, gender) and patient decisions on lifestyle (e.g., smoking, contraception).

Laboratory (9.89%) contained subcategories such as laboratory parameter (e.g., bilirubin, calcium, blasts), parameter value (e.g., $>$, \geq , $<$, \leq , positive, negative) or sample type (blood, bone marrow, plasma). For instance, SMCs in this category help to ensure that a patient is clinically suited for enrolment into a clinical trial (e.g., exclusion criterion “inadequate liver function: bilirubin greater than two times normal, alanine aminotransferase or aspartate aminotransferase greater than five times normal”; TPF-C-HIT study).

The main category “Findings” (5.61%) comprised subcategories such as information on the examination type (e.g., echocardiography, pulmonary function test), outcome parameter (e.g., corrected QT interval, ejection fraction), or outcome value (e.g., adequate, clinically significant). SMCs in the Findings main category often serve to exclude comorbidities that might prevent study participation, e.g., exclusion criterion

“Clinically significant abnormal 12-lead ECG or electrocardiography, QT interval corrected using Fridericia’s method (QTcF) $>$ 450 milliseconds,” TIGER-3 study.

Study (5.21%) comprised subcategories such as aspects of study participation (e.g., compliance with the study protocol, consent) and details on required timings for the study (e.g., exclusion criterion “Treatment in any other clinical trial within 30 days before screening,” Vaximm Studie [VXM01-03-DE] or exclusion criterion “Participation in other interventional trial within the last 30 days,” TPF-C-HIT study).

The smallest main category was Others (only containing 18 SMCs, 0.38%), which comprised general patient information (e.g., “Patients of a specific cohort”).

Thus, the sum of the structurable SMCs in the main categories amounted to 3,966 (83.64%).

Of the remaining 16.36%, 2.76% were only examples or headings (inserted due to the export from CTIS) that were not included into the analyses. A portion of 645 SMCs (13.60%) was not possible to be structured with our approach and was therefore analyzed in more detail (see Not Structurable SMCs).

The analysis of the linkage types among the statements of an SMC (i.e., between the lines of an SMC within the structuring template, see the template for criteria structuring in General Criteria Structuring) showed that “AND”-linkages were present in both the inclusion and exclusion criteria and that in both cases they were by far most prevalent between the main categories Diagnosis and Therapy.

Also, most pairwise combinations of main categories involving either Diagnosis or Laboratory were more prevalent in the inclusion criteria (e.g., Diagnosis-Diagnosis, Diagnosis-Findings, Diagnosis-Laboratory, Laboratory-Demographics and Lifestyle [D&L]), while combinations with therapy seemed to occur more often within the exclusion criteria (e.g., Therapy-Therapy, Therapy-Study, Therapy-Findings, Therapy-Laboratory).

Linkages of the “OR”-type only occurred in the inclusion criteria. In contrast to “AND,” they mainly linked equal main categories. Laboratory-Laboratory was the most prevalent combination (followed by D&L-D&L and Diagnosis-Diagnosis).

Not Structurable SMCs

The detailed analysis and categorization of the 645 SMCs (13.60%) that were not structurable by our approach revealed that the underlying reasons might be subdivided into two categories. The first one involved ambiguities referring to content and was entitled “Additional information beyond criterion text necessary.” The second category comprised problems with structural composition (“Structural ambiguity,” [Fig. 3](#)).

These two main categories were further subdivided into four subcategories each: The content-related main category comprised the subcategories “Supplementary medical knowledge necessary,” “Individual list necessary,” “Missing or unclear definitions of terms in criterion text,” as well as “Official catalogue necessary” (for details see below).

The structural main category contained the subcategories “Linkages too complex,” “Criterion cannot be detected as inclusion or exclusion criterion due to formulation,” “Logical linkage even with supplementary medical knowledge not evident,” and “Logical linkage with supplementary medical

knowledge evident” (for details see below). The distribution of not structurable SMCs across the categories of structuring problem is shown in [Fig. 3](#).

In the following, the main as well as the subcategories of the not structurable SMCs are described in detail and corresponding examples are given.

The main category “Additional information beyond criterion text necessary” (A) has the following four different subcategories (see dotted bars on the left half of [Fig. 3](#)):

A1. Subcategory “Supplementary medical knowledge necessary.” The term “supplementary medical knowledge” means that a person needs medical knowledge (e.g., a physician) to interpret an eligibility criterion or an SMC and to apply it correctly to decide on a patient’s eligibility for a clinical trial. The medical knowledge is too complex to be completely given in a criterion text. A detailed within group analysis of the reasons for missing structurability revealed that medical knowledge is especially needed regarding:

- Judgment of the severity and course of the current oncological disease, as reflected by terms such as “active” (e.g., “Richter’s transformation in current relapse or active disease,” exclusion criterion, TUD CLL X4 study).
- Judgment, whether a test outcome is clinically insignificant (e.g., “Normal 12-lead ECG or clinically insignificant abnormalities in the judgment of the investigator,” inclusion criterion, and Wilex study [WX 60-004]).
- Judgment of therapy options for individual patients (e.g., whether their neoplasia might be “resectable” or “nonresectable,” or whether there might be a standard therapy available, e.g., “Presence of nonresectable liver metastasis,” Vaximm Studie [VXM01-03-DE]).

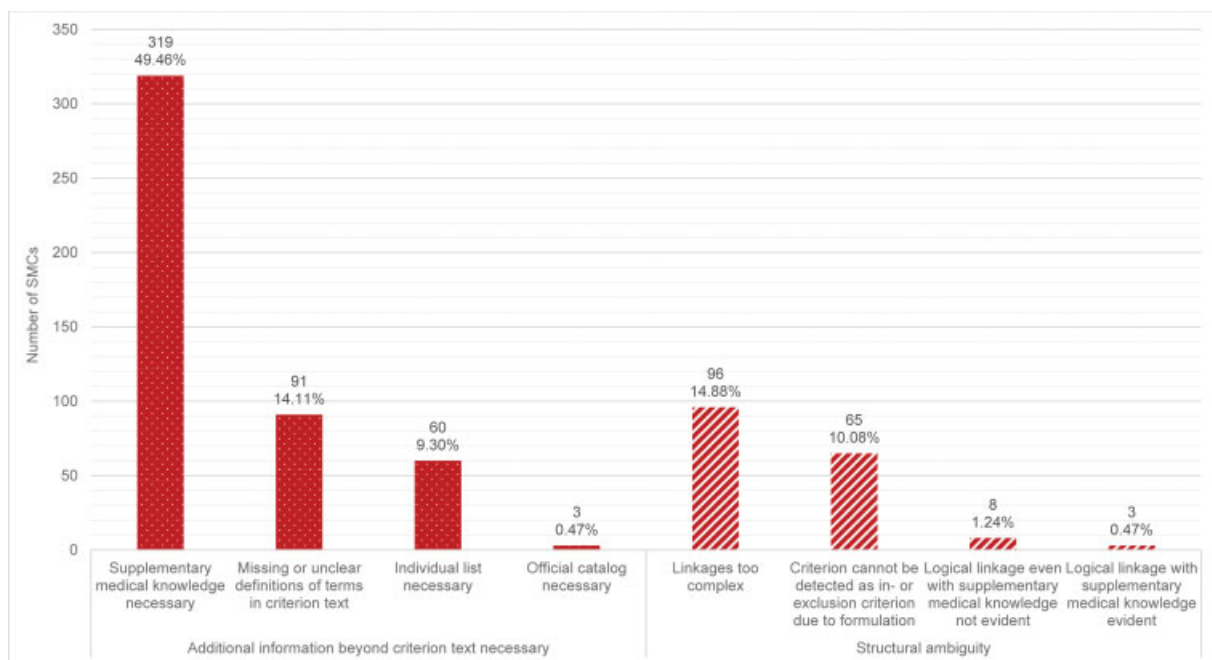


Fig. 3 Categorization of the 645 not structurable SMCs by underlying structuring problem. The given percentages refer to the proportion on the overall number of not structurable SMCs, i.e., 645. SMC, smallest meaningful component.

- Judgment of the recovery from previous treatments (related to the general condition of the patient and not to the tumor itself), in the case that a reference to standards such as the CTCAE is not given (e.g., “Full recovery from surgery, chemotherapy, or radiation therapy,” inclusion criterion of the Vicoryx [P16_37–63] study).
- Judgments of risks and contraindications for study participation, as they might affect patient or study (e.g., “or other severe concurrent disease, which, in the opinion of the investigator, would place the patient at undue risk or interfere with the study,” exclusion criterion, PH-L19IL2–03/09). In the case of comorbidities, it shall be checked whether they are adequately controlled (e.g., “Uncontrolled or insulin-dependent diabetes mellitus,” exclusion criterion, REMOTUX study).
- Knowledge regarding the effects of substances/drugs (e.g., “History of hypersensitivity to cetuximab or to any drug with similar chemical structure or to any excipient present in the pharmaceutical form of the investigational medicinal product,” exclusion criterion, and REMOTUX study). Apart from medical knowledge, interpersonal assessment and intuition are needed for the Judgment of legal competence (e.g., “Ability of subject to understand character and individual consequences of the clinical trail” inclusion criterion, TAME-AL study).
- Estimation of personal aspects, such as compliance (e.g., “Willingness and ability to comply with the scheduled visits, treatment plan, laboratory tests and other study procedures,” inclusion criterion, PH-L19IL2–03/09 study).

A2. The subcategory “Individual list necessary” indicates that even an unofficial, locally developed list of study-specific characteristics would contribute to changing a formally not structurable criterion into a structurable one. For instance, a list of laboratory parameters together with their accepted value ranges reflecting sufficient organ function for the respective study will enable the structuring of inclusion criteria such as “Have adequate organ function” (Tabalumab-study [H9S-MC-JDCG]).

A3. The subcategory “Missing or unclear definitions of terms in criterion text” represent a group of structuring problems, where ambiguous interpretations might arise from missing definitions. An example is given in the exclusion criterion “Recent participation in another interventional trial” (ParvOryx02 study), where a definite time specification might replace “recent.” Another example is the exclusion criterion “Known allergy against the IMP or drugs with similar chemical structure or additives” of the “SAHA-I” study (NCT00918489), where “similar” might allow substantial variations in interpretation. Nonetheless, this category was classified as one that might be remediable by adding corresponding definitions during criterion phrasing.

A4. The subcategory “Official catalogue necessary” contains eligibility criteria that might become structurable with the help of an official catalogue. The latter could be stored in the

local system and support criteria structuring by supplementing information on side effects of drugs or half-lives of chemical components. An example for this subcategory is the exclusion criterion “Conflicts in patient recruitment arisen from warning messages, safety precautions, and contraindications as listed in the valid specialist information for Capecitabin (Xeloda)” (translated from the original German text “Konflikte bei der Patientenauswahl mit Warnhinweisen, Vorsichtsmaßnahmen und Kontraindikationen, aufgeführt in der gültigen Fachinformation zu Capecitabin [Xeloda]”) of the WF10-study.

The main category “Structural ambiguity” (B) contains the following four subcategories (see striped bars on the right side of [Fig. 3](#)):

B1. As revealed by a detailed within-group analysis, the category “Linkages too complex” represented several structuring obstacles, including causalities (e.g., “...unless caused by...”), comparisons (e.g., “...14 days or five half-lives of a drug [whichever is longer]”), linkages across inclusion and exclusion criteria (e.g., “any other exon not allowed by the inclusion criteria”), exclusion of a specific case from a group (e.g., “steroids for adrenal failure”: in our approach, we can exclude the substance class of steroids, but we cannot exclude the steroids for adrenal failure specifically) as well as further constellations that go beyond the possibilities of our structuring approach (see Discussion). An example for the latter is the extract from the exclusion criterion of the Vaximm Study (VXM01–03–DE): “Participants with treated brain metastases are eligible if they are clinically stable with regard to neurologic function, off steroids after cranial irradiation ending at least 2 weeks prior to randomization, or after surgical resection performed at least 28 days prior to randomization.”

B2. An example for the category “Logical linkage even with supplementary medical knowledge not evident” is the exclusion criterion “Brain metastases or spinal cord compression unless asymptomatic, treated and stable” (study “MYSTIC”; NCT02453282). It remains unclear whether the descriptive adjectives “asymptomatic, treated, and stable” only refer to the spinal cord or also to the brain metastases. However, this subcategory might be remediated by using alternative formulations.

B3. The subcategory “Criterion cannot be detected as inclusion or exclusion criterion due to formulation” contains criteria that have been assigned to the inclusion or exclusion criteria. However, their formulation does not make them explicit and therefore they cannot be handled by our approach. Representative terms include “...is permitted,” “...is allowed” or “...is eligible for,” as found in the inclusion criterion “Placement of a central vascular access device performed = 2 days prior to bevacizumab administration is allowed” (study “TAVAREC [EORTC-26091]”; NCT01164189) or in the exclusion criterion “Ongoing infection > Grade 2 according to NCI-CTCAE (National Cancer Institute–Common Terminology Criteria for Adverse Events) v. 4.0. Hepatitis B is allowed if no active

replication is present. Hepatitis C is allowed if no antiviral treatment is required” (study “RESORCE [Bayer BAY73-4506, BAY15982]”; NCT01774344). Nonetheless, we consider this subcategory as remediable by rephrasing.

B4. The subcategory “Logical linkage with supplementary medical knowledge evident” comprises criteria that have been phrased ambiguously but which might be structured nonetheless by resolving the underlying logical linkage with the help of medical knowledge. For instance, within the exclusion criterion “Clinically significant autoimmune cytopenia or clinically significant hemolytic anemia with suspicion of immune origin, even if Coombs test is negative” (t-pll2; NCT01186640), it might not be clear without medical knowledge whether the “Coombs test” only refers to the hemolytic anemia or also to the autoimmune cytopenia. Nonetheless, we consider this subcategory as a remediable one.

Taken together, six of these eight subcategories of the not structurable criteria were considered to be remediable (corresponding to 35.66% of the 645 underlying not structurable SMCs). Merely the two subcategories “Supplementary medical knowledge is necessary” (A1) and “Linkages too complex” (B1) were regarded as impossible to be remediated within our approach, which corresponded to 64.34% of the not structurable SMCs (see Discussion).

The proportion of potentially remediable SMCs in relation to those considered as remaining not structurable within our approach is shown per main category in ▶Fig. 4. In accordance with being the largest main category (for sizes refer to ▶Fig. 2), Diagnosis contained the largest proportion of not structurable SMCs. Interestingly, Study as the second smallest main category, also showed an unproportionally high amount of not structurable SMCs. In both Diagnosis and Study (together with Laboratory), the number of potentially remediable SMCs was low compared with most other main categories, which had potential remediation rates of >50%.

Interestingly, the reasons that SMCs were not possible to be structured differed within and between the inclusion and exclusion criteria as well as the main categories (▶Fig. 5). The exclusion criteria roughly contained twice as much not structurable SMCs as the inclusion criteria (421 and 224, respectively). The most prominent structuring problem regarding “Additional information beyond criterion text necessary” (A in this section) across the inclusion and exclusion criteria (upper line of ▶Fig. 5) was that supplementary medical knowledge would have been needed. This was especially frequent in the main categories Diagnosis (184) and Study (78) of the exclusion criteria (279; ▶Fig. 5, top right). In the inclusion criteria, the most common structuring problems were “Missing or unclear definitions” (49) and

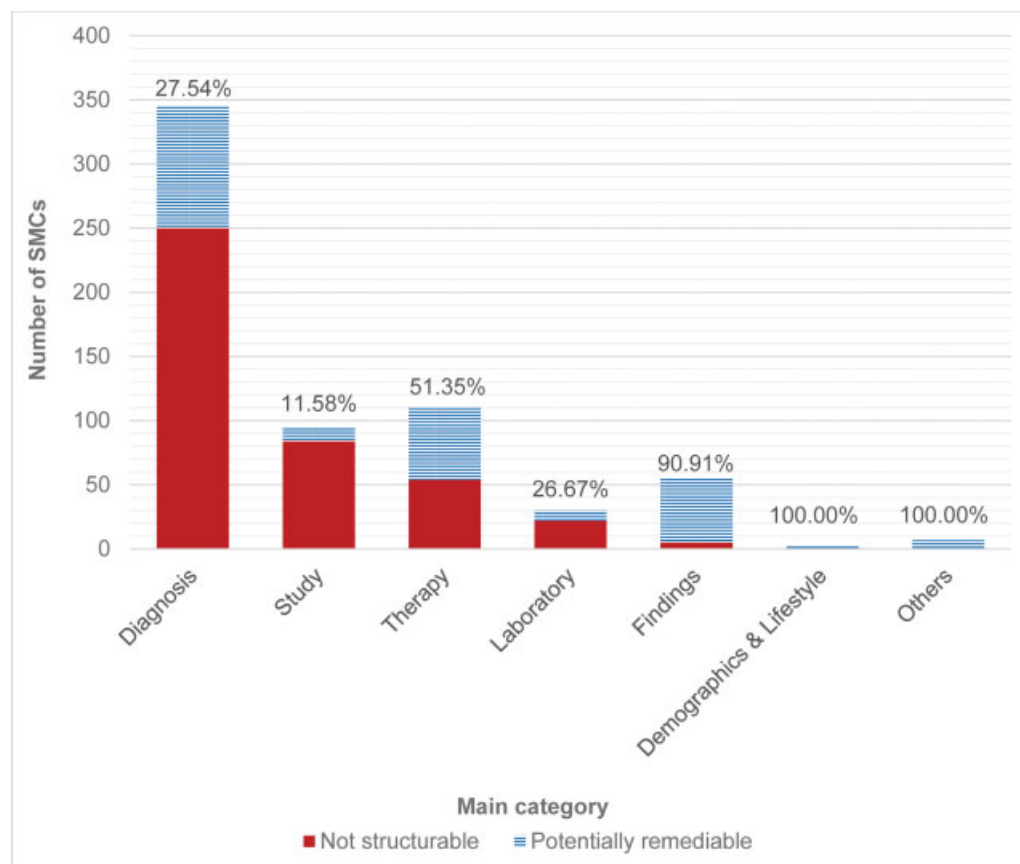


Fig. 4 Distribution of all 645 not structurable SMCs across the main categories and proportions of SMCs that might be turned into structurable ones. The proportion of SMCs remaining not structurable within our approach is indicated in red, while the proportion that might be transformed into structurable SMCs is indicated in horizontally striped bars on top. The percentages given refer to the proportion of SMCs (on the overall number of not structurable SMCs, i.e., 645) that might be turned into structurable SMCs in the corresponding main category. SMC, smallest meaningful component.

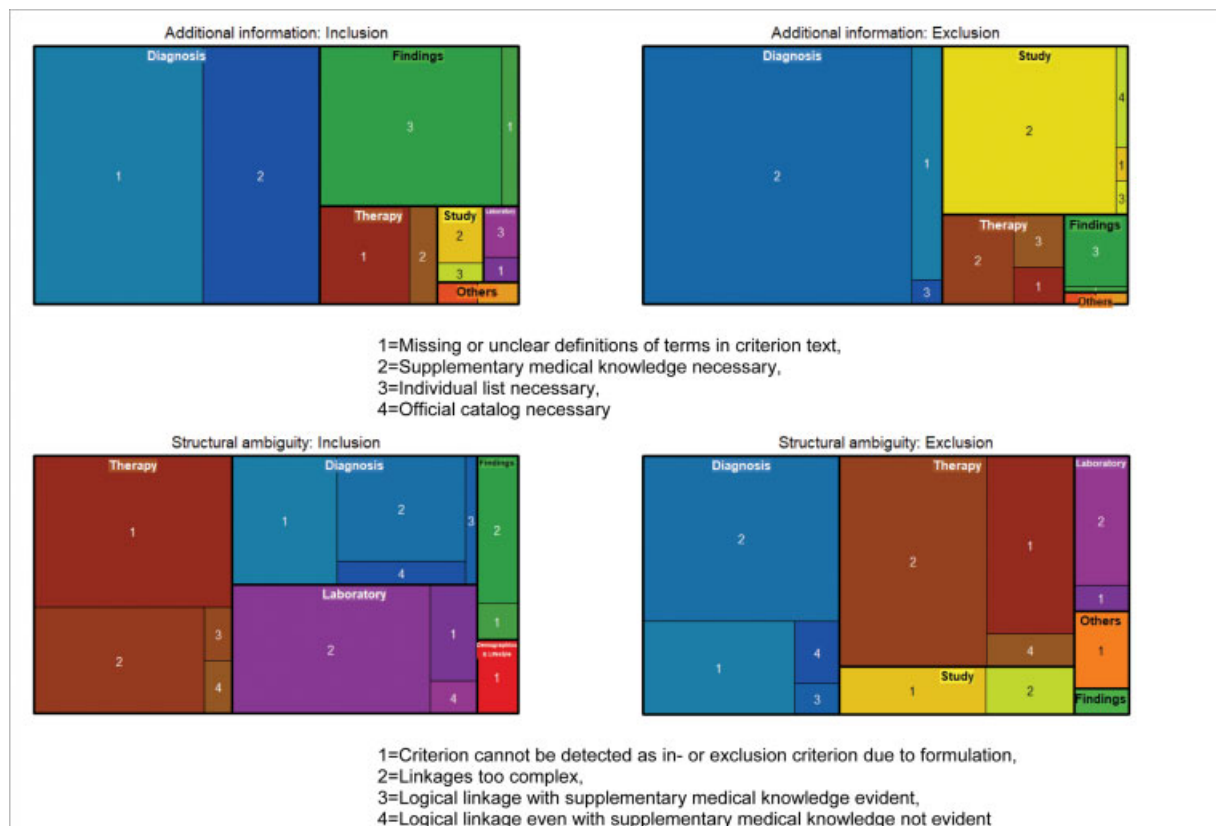


Fig. 5 Reasons for SMC structuring problems per main category and across inclusion and exclusion criteria. SMC, smallest meaningful components.

“Supplementary medical knowledge necessary” (40) in Diagnosis, followed by missing individual lists in Findings (33; ▶Fig. 5, top left).

Regarding the main category of structuring problems “Structural ambiguity” (see B in this section), the subcategories of structuring problems were more evenly distributed within and between the inclusion and exclusion criteria. Diagnosis, Therapy, and Laboratory had most issues and the problem that linkages were too complex was the most frequent structuring problem among them (23, 22 and 5 cases, respectively; ▶Fig. 5, lower line).

Discussion

This study aimed to analyze the structuring problems in oncological eligibility criteria (here: SMCs). Central characteristics of our study were its experimental character, as well as the aim of placing no a priori assumptions or restrictions to criteria structuring. Therefore, none of the previous eligibility criteria structuring approaches (either based on content or on semantic pattern representations⁵) was applied.

The identification of not structurable SMCs within this study’s frame was based on our general criteria structuring approach (described in General Criteria Structuring), which resulted in the seven main categories: Diagnosis, Therapy, Laboratory, Study, Findings, Demographics and Lifestyle, and Others. They correspond well to other studies such as to the six topic groups of Luo et al,^{10,17} comprising Health Status

(e.g., disease or organ status), Treatment or Health Care (e.g., drug), Diagnostic or Laboratory Tests (e.g., creatinine), Ethical Consideration (e.g., willing to consent), Lifestyle Choice (e.g., diet or exercise), and Demographics (e.g., age or gender). Merely the contents of Luo et al’s topic group Diagnostic or Laboratory Tests did not match directly with our main categories. In our study, we had split these contents into the two main categories: Findings and Laboratory. Our main category, Others only contained few and special elements, which might be why it did not have a matching counterpart in Luo et al.^{10,17} These topic groups of Luo et al resulted from 3,400 randomly selected eligibility criteria from Clinicaltrials.gov without disease-specific focus. On these criteria, hierarchical clustering with manual cluster granularity adjustment based on the Unified Medical Language System (UMLS) had been applied. Luo et al’s topic groups seem to be profound, as they were applied in various other studies, for instance to compare own results or as a template for criteria classification.^{10,13,14,17–19}

Other examples for comparable content groupings are the studies of Bruland et al¹¹ and Doods et al,⁹ who identified data elements commonly appearing in both EHRs and clinical trials and grouped them by their content. Bruland et al¹¹ reported the top 14 of their groups (i.e., “form domains”), sorted by the frequency of occurrence. Although these were textually separated to a higher degree than in our study, they were assignable to our seven main categories. The work of Doods et al⁹ resulted in nine data groups that correspond

relatively well to our main categories, as four of them (Diagnosis, Laboratory findings, Findings, and Patient characteristics) can directly be compared with their counterparts in our study (Patient characteristics corresponding to Demographics and Lifestyle).

Taken together, the comparison of our main categories with those of other studies suggests that although there are interstudy differences due to subjectivity, the groupings share a similar division by content. Therefore, the seven main categories resulting from our first step-structuring are a valid basis for identifying and exploring not structurable criteria in our study.

As described above, among the 4,742 overall SMCs, a proportion of 645 SMCs (13.60%) was identified as not structurable within our approach. This proportion was subsequently analyzed in detail for underlying structuring problems and potential remediation approaches, i.e., possibilities for conversion into structurable criteria.

Merely SMCs requiring supplementary medical knowledge or that had too complex linkages were considered not remediable (64.34%, corresponding to 8.75% of all SMCs). "Supplementary medical knowledge necessary" was the largest subgroup of the not structurable SMCs by far (76.87% of the not remediable criteria, 6.73% of the total of 4,742 criteria) and contained formulations asking for individual opinions and estimations of the study team or physicians ("in the opinion of the investigator"). Hence, the knowledge of the investigator is essential and does not seem replaceable by information given in the study text. The assumption that criteria asking for professional judgment or describing patient preferences is not possible to be structured in accordance with Ateya et al¹⁴ and Ross et al.¹² In the latter study, a proportion of 19% of all 932 comprehensible criteria (without disease-specific focus) consisted of criteria requiring clinical judgment. In Ross et al's study, it was assigned to a group of "complex criteria." This group comprised 718 criteria (77%) and involved semantic connectors as well as definitions of temporal constraints (among others). The sum of all criteria identified as remaining too complex to evaluate within this study¹² added up to 786 criteria (78.6%). Compared with our study (8.75% of criteria remaining not structurable overall), the proportion of Ross et al¹² is quite high. However, we had subtracted SMCs that might be manually remediated in future studies, e.g., by supplementing information such as individual lists or definitions. Without this assumption, our results suggest a portion of not structurable SMCs of 13.60% overall. The remaining discrepancy to Ross et al¹² might emanate from more basic differences in study approaches. For instance, in contrast to Ross et al's¹² groups of "incomprehensible or nonselective" and "complex" for not structurable criteria, we constructed one single pool of SMCs that differed in the degree of structurability and the reasons, therefore. This pool served as a basis for grouping the obstacles of criteria structuring into different categories as well as for deriving possibilities of remediation.

The comparison with proportions of not structured eligibility criteria in studies of other structuring approaches also shows variations of results. Ateya et al¹⁴ explored the extent to which CES (comparable to our SMCs, see General Criteria

Structuring) without disease-specific focus from primary care clinical trials could be represented by structured EHR elements. A proportion of 26% was not possible to be covered. The authors found divergent results compared with Köpcke et al,¹³ who reported a proportion of 45% not being possible to be represented.¹⁴ Ateya et al¹⁴ compared the content of their CES categories with those of Köpcke et al¹³ It was concluded that the CES categories overlapped but were not identical.¹⁴ This also applies to the categories of our SMCs. Similar to other studies, we did not achieve clear and non-overlapping SMC formulation and categorization, as this is subjective.

Taken together, it can be assumed that differences in structuring approaches had caused the observed differences regarding the proportion of structurable eligibility criteria (e.g., manually, semiautomated, varying underlying knowledge representations). Furthermore, differences in the definition and handling of abstract concepts such as "not structurable," "active," or "adequate" might lead to divergent results. For instance, the SMC "adequate liver function" was not covered within our approach. We presupposed an exact value range given for "adequate" concerning liver function measured by laboratory results. In contrast, Luo et al¹⁷ subsumed criteria such as "Patient must have adequate organ function" into their semiautomatically induced semantic class of "Organ and Tissue Status."

However, despite differences in the degree of structurability, the necessity of complementing the automated patient-trial matching process with person-to-person interviews by the physician before study enrolment was underlined in multiple previous studies (e.g., Köpcke et al¹³, Weng et al¹⁸, Sahoo et al²⁰). This is in line with our result, suggesting that the physicians' medical knowledge cannot be replaced by or put into eligibility criteria texts.

The subcategory "Linkages too complex" was also judged as not structurable and remediable within our approach. Interestingly, as stated above, the more detailed analysis within this subgroup revealed several reasons therefore (i.e., causalities, comparisons, linkages across in- and exclusion criteria, exclusion of a specific case from a group as well as further constellations that go beyond the possibilities of our structuring approach). For instance, these issues are reflected in the excerpt of the following exclusion criterion of the Vaximm Study (VXM01-03-DE):

"Untreated central nervous system (CNS) metastases: Participants with treated brain metastases are eligible if they are clinically stable with regard to neurologic function, off steroids after cranial irradiation ending at least 2 weeks prior to randomization, or after surgical resection performed at least 28 days prior to randomization."

The complex linkages among the parts of the criterion can be formulated as follows (the part "Untreated CNS metastases" is not considered):

"(M1 AND ICDO3_localized) therapy_conducted (AND NOT clinically_stable = neurologic_function AND NOT steroids_exclusion after (radiotherapy_localized ≥2 weeks before randomization OR [surgery AND R0] ≥28 days before randomization))." This underlines the complexity of

eligibility criteria phrasings, which exceeds the possibilities of our structuring template.

Beyond our approach, some semantic issues might be handled by expression languages such as ERGO.⁸ Nonetheless, a certain part of not structurable criteria involving too complex linkages might remain, as indicated by Ross et al¹²: Exploring eligibility criteria from ClinicalTrials.gov regarding their semantic complexity (among other aspects), Ross et al¹² identified a proportion of 9% among the complex criteria (718; see above) that were not possible to be evaluated, as they involved the use of semantic connectors that are not covered by current representation languages.

Exploring the types of structuring problems between inclusion and exclusion criteria in dependence of the main categories (e.g., Diagnosis, Therapy, and so on) revealed a variety of different structuring obstacles within the inclusion criteria (missing or unclear definitions, vague criterion formulations impeding the recognition of a definite inclusion or exclusion criterion, too complex linkages as well as missing individual lists). In contrast, the only structuring obstacle mainly appearing within the exclusions was missing supplementary medical knowledge. As the largest, not remediable structuring obstacle identified in our study, its prevalence in Diagnosis and Study explains the high proportions of not structurable criteria in these main categories displayed in **Fig. 4**. We identified a clearly higher proportion of missing supplementary medical knowledge within the exclusions (43.26% of the amount of 645 not structurable SMCs) compared with the inclusion criteria (6.20%; for a definition of “Supplementary medical knowledge” see Not Structurable SMCs). This suggests that physician's medical expertise is essential for confirming the matching results and identifying factors that might harm the patient or bias the study.

Taken together, a proportion of 64.34% of the SMCs identified as not structurable might not be remediable, as their structuring based on our approach was impeded by missing supplementary medical knowledge or too complex linkages. The remaining proportion of 35.66% might become structurable by taking measures such as including individual lists, giving more concrete information on definitions, or ensuring a stronger differentiation between in- and exclusion criteria. However, a more diligent phrasing of eligibility criteria does not seem to contribute significantly to an increase in the proportion of structurable criteria or SMCs. The necessity of a partially manual patient trial matching is underlined.

Our analysis of the contents and phrasings that complicate the structuring of oncological eligibility criteria might contribute to a better formulation in the future, as these identified issues can then be circumvented. In addition, the presented considerations potentially support NLP (Natural Language Processing) approaches by providing detailed insights into frequent human formulations that remain not machine readable.

Limitations

When interpreting our results, it has to be considered that the definition of “not structurable SMC” is subjective and

specific to this study. SMCs that were defined as “not structurable” by our approach might have been handled by other studies and vice versa.

Furthermore, it has to be taken into account that the assignment of SMCs to subcategories as well as the assignment of subcategories to their corresponding main categories was ambiguous in some cases, so that slightly different structuring results might be conceivable. For instance, the SMC “Participation in investigational study” might belong to the two main categories: Therapy and Study. As the assignment was to be done on exactly one main category, the SMC was assigned to Therapy after an in-team discussion. However, during a review cycle within the structuring process, it was moved to Study to pronounce the SMC's study-related content. Issues like these reflect the subjectivity within this working step and were carefully discussed within the team. However, as we maintained our concept throughout the structuring procedure and had the process reviewed by two other team members (A.K., J.D.), the degree of “structurability” can be assumed to be comparable across the SMCs within our study. Besides, an exploratory, manual structuring approach was chosen to do justice as much as possible to the oncology-specific full-text sentences of eligibility criteria in natural language and the separation into inclusion and exclusion criteria. Our aims were SMC exploration and detailed recognition of characteristics that might lead to structuring problems, in contrast to high-level logical consistency that might be the advantage of semiautomated approaches.

Conclusion

The aim of our study was the exploration of structuring problems in oncological eligibility criteria (SMCs). The basis of the identification of not structurable SMCs was our own manual eligibility criteria structuring approach. The latter might be considered suited for our aim, as the main categories identified by it were comparable with the ones of previous studies.

The three most-frequent reasons for impossible SMC structuring were missing supplementary medical knowledge (for a definition, see General Criteria Structuring), missing or unclear definitions, as well as too complex linkages between the criteria statements. Interestingly, the reasons varied within and between the inclusion and exclusion criteria as well as the main categories: While a variety of reasons for structuring problems were observed in the inclusion criteria, supplementary medical knowledge was the main reason for impossible structuring within the exclusion criteria. The latter suggests that only human beings with medical knowledge (i.e., physicians) can prevent the patient from potential harm during study participation. For instance, only a physician can decide whether a patient has sufficiently recovered from previous treatments and that the patient's risk is minimal. In addition, the physician's decision also minimizes study biasing, e.g., by checking other therapy options for a patient and contraindications. This complexity

of the considerations and decisions can currently not be achieved by algorithms.

Furthermore, our results suggest that a more diligent formulation of eligibility criteria does not seem to contribute largely to an increase in the proportion of structurable SMCs. Therefore, our results argue that human medical knowledge and judgment cannot be substituted by automatized patient-trial matching algorithms completely.

Conflict of Interest

None declared.

Acknowledgments

The authors cordially thank Kristina Nett for her support in all eligibility criteria-related issues as well as the corresponding physicians from the NCT for their feedback regarding medical questions.

References

- 1 Stiller CA, Kroll ME, Pritchard-Jones K. Population survival from childhood cancer in Britain during 1978–2005 by eras of entry to clinical trials. *Ann Oncol* 2012;23(09):2464–2469
- 2 Chow CJ, Habermann EB, Abraham A, et al. Does enrollment in cancer trials improve survival? *J Am Coll Surg* 2013;216(04):774–780, discussion 780–781
- 3 Goyal J, Nuhn P, Huang P, et al. The effect of clinical trial participation versus non-participation on overall survival in men receiving first-line docetaxel-containing chemotherapy for metastatic castration-resistant prostate cancer. *BJU Int* 2012;110(11 Pt B):E575–E582
- 4 Metz JM, Coyle C, Hudson C, Hampshire M. An Internet-based cancer clinical trials matching resource. *J Med Internet Res* 2005;7(03):e24
- 5 Weng C, Tu SW, Sim I, Richesson R. Formal representation of eligibility criteria: a literature review. *J Biomed Inform* 2010;43(03):451–467
- 6 Niland J, Dorr D, El Saadawi G, et al. Knowledge Representation of Eligibility Criteria in Clinical Trials. Chicago American Medical Informatics Association Annual Symposium 2007
- 7 Sordo M, Boxwala AA, Ogunyemi O, Greenes RA. Description and Status Update on GELLO: a Proposed Standardized Object-Oriented Expression Language for Clinical Decision Support. Amsterdam: IOS Press; 2004:164–168
- 8 Tu SW, Peleg M, Carini S, et al. A practical method for transforming free-text eligibility criteria into computable criteria. *J Biomed Inform* 2011;44(02):239–250
- 9 Doods J, Botteri F, Dugas M, Fritz FEHR4CR WP7. A European inventory of common electronic health record data elements for clinical trial feasibility. *Trials* 2014;15:18
- 10 Luo Z, Yetisgen-Yildiz M, Weng C. Dynamic categorization of clinical research eligibility criteria by hierarchical clustering. *J Biomed Inform* 2011;44(06):927–935
- 11 Bruland P, McGilchrist M, Zapletal E, et al. Common data elements for secondary use of electronic health record data for clinical trial execution and serious adverse event reporting. *BMC Med Res Methodol* 2016;16(01):159
- 12 Ross J, Tu S, Carini S, Sim I. Analysis of eligibility criteria complexity in clinical trials. Summit On Translat Bioinforma 2010; 2010:46–50
- 13 Köpcke F, Trinczek B, Majeed RW, et al. Evaluation of data completeness in the electronic health record for the purpose of patient recruitment into clinical trials: a retrospective analysis of element presence. *BMC Med Inform Decis Mak* 2013;13:37
- 14 Ateya MB, Delaney BC, Speedie SM. The value of structured data elements from electronic health records for identifying subjects for primary care clinical trials. *BMC Med Inform Decis Mak* 2016; 16:1
- 15 DIMDI, Deutsches Institut für Medizinische Dokumentation und Information. Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme, 10th revision, German Modification. Accessed April 16, 2020 at: <https://www.dimdi.de/dynamic/de/klassifikationen/icd/icd-10-gm/>
- 16 Division of Medical Informatics for Translational Oncology. Electronic Science Record. <https://www.dkfz.de/en/mitro/index.php>. Accessed April 16, 2020
- 17 Luo Z, Johnson SB, Weng C. Semi-automatically inducing semantic classes of clinical research eligibility criteria using UMLS and hierarchical clustering. *AMIA Annu Symp Proc* 2010; 2010:487–491
- 18 Weng C, Batres C, Borda T, et al. A real-time screening alert improves patient recruitment efficiency. *AMIA Annu Symp Proc* 2011;2011:1489–1498
- 19 Weng C, Wu X, Luo Z, Boland MR, Theodoratos D, Johnson SB. EliXR: an approach to eligibility criteria extraction and representation. *J Am Med Inform Assoc* 2011;18(Suppl 1):i116–i124
- 20 Sahoo SS, Tao S, Parchman A, et al. Trial prospector: matching patients with cancer research studies using an automated and scalable approach. *Cancer Inform* 2014;13:157–166