A novel method for continuous measurements of clinical practice guideline adherence

Ebben, K.C.W.J.; Kroon, C.D. de; Schmeink, C.E.; Hel, O.L. van der; Vegchel, T. van; Moncada-Torres, A.; ... ; Werf, J. van der

Received: 29 March 2023 | Revised: 24 July 2023 | Accepted: 26 July 2023

DOI: 10.1002/lrh2.10384

RESEARCH REPORT

Learning Health Systems

A novel method for continuous measurements of clinical practice guideline adherence

Kees C.W.J. Ebben[1] | Cornelis D. de Kroon[2] | Channa E. Schmeink[3] | Olga L. van der Hel[1] | Thijs van Vegchel[1] | Arturo Moncada-Torres[1] | Ignace H.J.T. de Hingh[1,4] | Jurrian van der Werf[1]

[1]Department of Research and Development, Netherlands Comprehensive Cancer Organization (IKNL), Utrecht, The Netherlands

[2]Department of Obstetrics and Gynecology, Leiden University Medical Center, Leiden, The Netherlands

[3]Department of Gynecology, Anna Hospital, Geldrop, The Netherlands

[4]Department of Surgical Oncology, Catharina Hospital, Eindhoven, The Netherlands

Correspondence
Kees C.W.J. Ebben, Netherlands Comprehensive Cancer Organization, p.a. Hatertseweg 1, Nijmegen 6533 AA, The Netherlands.
Email: k.ebben@iknl.nl

Abstract

Introduction: Clinical practice guidelines (hereafter 'guidelines') are crucial in providing evidence-based recommendations for physicians and multidisciplinary teams to make informed decisions regarding diagnostics and treatment in various diseases, including cancer. While guideline implementation has been shown to reduce (unwanted) variability and improve outcome of care, monitoring of adherence to guidelines remains challenging. Real-world data collected from cancer registries can provide a continuous source for monitoring adherence levels. In this work, we describe a novel structured approach to guideline evaluation using real-world data that enables continuous monitoring. This method was applied to endometrial cancer patients in the Netherlands and implemented through a prototype web-based dashboard that enables interactive usage and supports various analyses.

Method: The guideline under study was parsed into clinical decision trees (CDTs) and an information standard was drawn up. A dataset from the Netherlands Cancer Registry (NCR) was used and data items from both instruments were mapped. By comparing guideline recommendations with real-world data an adherence classification was determined. The developed prototype can be used to identify and prioritize potential topics for guideline updates.

Results: CDTs revealed 68 data items for recording in an information standard. Thirty-two data items from the NCR were mapped onto information standard data items. Four CDTs could sufficiently be populated with NCR data.

Conclusion: The developed methodology can evaluate a guideline to identify potential improvements in recommendations and the success of the implementation strategy. In addition, it is able to identify patient and disease characteristics that influence decision-making in clinical practice. The method supports a cyclical process of developing, implementing and evaluating guidelines and can be scaled to other diseases and settings. It contributes to a learning healthcare cycle that integrates real-world data with external knowledge.

KEYWORDS
cancer, clinical decision trees, real-world data

1 | INTRODUCTION

Clinical practice guidelines (hereafter 'guidelines') play an essential role in summarizing evidence into recommendations for application in clinical practice.[1] They support physicians and multidisciplinary teams in making substantiated decisions for diagnostics, treatment, and follow-up along the entire care pathway.[2,3] In recent decades, such guidelines have been refined and made available for a wide variety of conditions and diseases, including cancer. It is now well established that guideline implementation reduces unwanted variability in clinical practice and contributes to better outcome of care.[4] However, implementation of guidelines in daily routine is demanding and hampered by many well-studied barriers at the level of the treating physician (eg, lack of familiarity and awareness), the guideline (eg, poor lay-out, poor access, low level of evidence, and lack of applicability) and external factors (eg, lack of resources and social norms).[5] Moreover, difficulties in adherence measurements is challenging for several reasons, including the availability of data and the difficulty in interpreting the results of adherence measures.[6,7]

The research community seeks to produce high quality evidence synthesis for better health care decision-making, proposing an evidence ecosystem.[8] Extensive research has been conducted into measuring guideline adherence in relation to various diseases, with many studies focusing on cancer.[9-13] However, almost without exception, these studies have been one-off investigations that have used time and (healthcare) setting specific performance measurements or indicators. These studies have generated important insights, yet their methods cannot be applied on a continuous scale. Therefore, there remains a lack of knowledge about the time it takes to implement recommendations into clinical practice and the optimal level or range of adherence for different (types of) recommendations and patient (sub)populations.

One way of continuously monitoring guideline adherence levels is through real-world data (RWD).[14,15] RWD comprise patient data generated during routine care in daily practice and is stored in various digital sources.[16] In many regions and countries, data are collected from these sources for cancer registries, which aim to provide insight in the characteristics and magnitude of cancer.[17] The completeness of many cancer registries is estimated to be high, representing the vast majority of cancer patients in real-life.[18]

When guidelines are developed such that decision points are computer-interpretable and align with RWD registries, continuous evaluation of the guideline is theoretically possible. Over the years, several methods for computer interpretable guidelines have been developed.[19-21] However, typically, guidelines are still not designed in a computer interpretable format and are usually presented in unstructured free text. Recently, a new methodology was developed to translate guidelines into clinical decision trees (CDTs) to tackle this problem,[22] with the CDTs being human readable while allowing for computer assisted evaluation of the underlying algorithm.

In the current study we expand on this development by introducing a novel methodology for continuous guideline adherence measurements using RWD. We applied the methodology first to the Dutch multidisciplinary endometrial cancer guideline using data from the Dutch cancer registry. The developed methodology was implemented in a prototype of a web-based dashboard, which enables an interactive usage and supports a variety of analyses.

We successfully demonstrated continuous monitoring of guideline adherence is possible using this novel methodology. This creates a closed loop in guideline development, implementation, and evaluation and healthcare delivery, making it a valuable part of the evidence ecosystem.[8]

2 | METHOD

Continuously comparing guideline recommendations and RWD requires the data items from both instruments to be aligned according to predefined steps (Figure 1). Evidently, both must cover a population with an identical condition or disease and ideally incidence dates from RWD cases and the prevailing period of the guideline overlap. To develop our methodology, we used the Dutch guideline for endometrial cancer and real-world data from the Netherlands Cancer Registry (NCR). This project was carried out in the context of the Alertness project and funded by ZonMw.[23] [Correction added on 19 September 2023, after first online publication: In the preceding sentence, funder 'ZonMw' was added in this version.]

2.1 | Guidelines

Guidelines are ideally developed using the PICO-methodology[24] and GRADE-system,[25] leading to evidence based recommendations for specific populations. Figure 2 brings these components together, including their interrelationships. These relations are denoted by cardinality, which represents the number of entities that can exist on each side of a relationship. In order to align the guideline and the RWD registry, the endometrial cancer guideline was parsed into machine readable CDTs. The corresponding method has been described by Hendriks et al.[22] In short, CDTs are composed of the following primitives: nodes (data items representing patient and disease characteristics), branches (representing the possible values of the data items) and leaves (representing one or multiple guideline recommendations) (Figure 3). Generally, guidelines include multiple subsequent decision moments during the care continuum and thus generate multiple CDTs. The CDTs are connected head-to-tail and consequently form a care pathway.

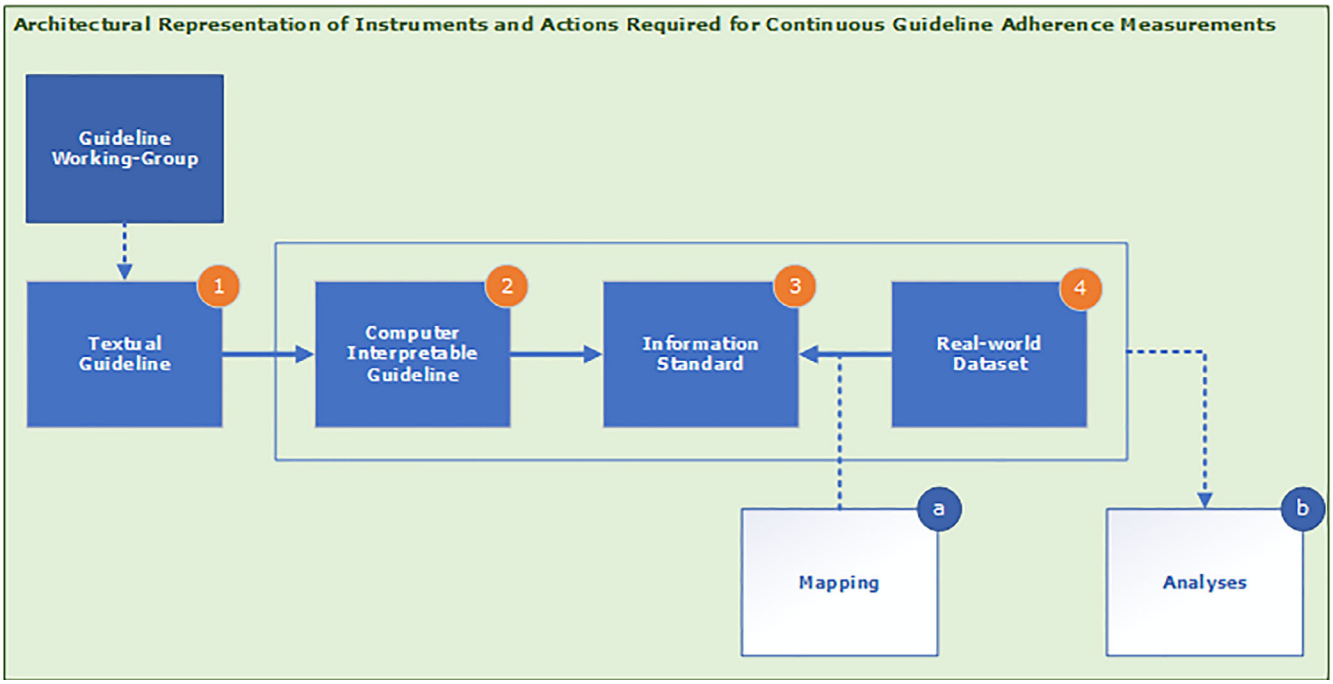

FIGURE 1 A textual guideline[1] is parsed into a computer interpretable format.[2] All patient and disease characteristics, and interventions are recorded in an Information Standard.[3] A Real-world dataset[4] is then the mapped (A) onto the Information standard. Subsequently, analyses (B) regarding guideline adherence can be performed.

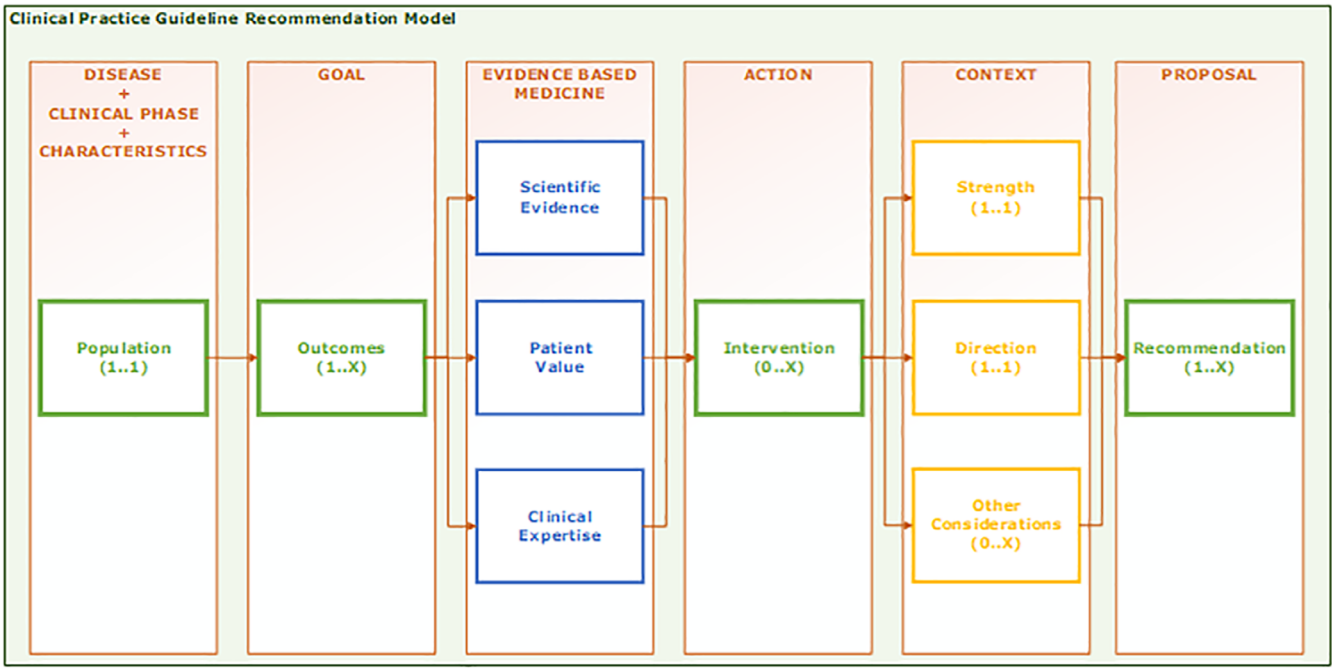

FIGURE 2 The clinical practice guideline recommendation model. In the development of guideline recommendations for a specific population, at a specific point in the care pathway, the aim is always to achieve optimal care outcomes. Interventions are assessed for appropriateness based on the evidence-based medicine method components. A recommendation is then created by providing the right context for each intervention. Cardinality (eg, 0 … X) represents the number of entities that can exist on each side of a relation.

An information standard is also drawn up iteratively while the CDTs are developed. In an information standard, all data items should be provided with metadata according to the FAIR-data principles[26] and be encoded using international terminologies (eg, SNOMED CT[27]) where possible. Consequently, an information standard facilitates mapping of data items between the guideline

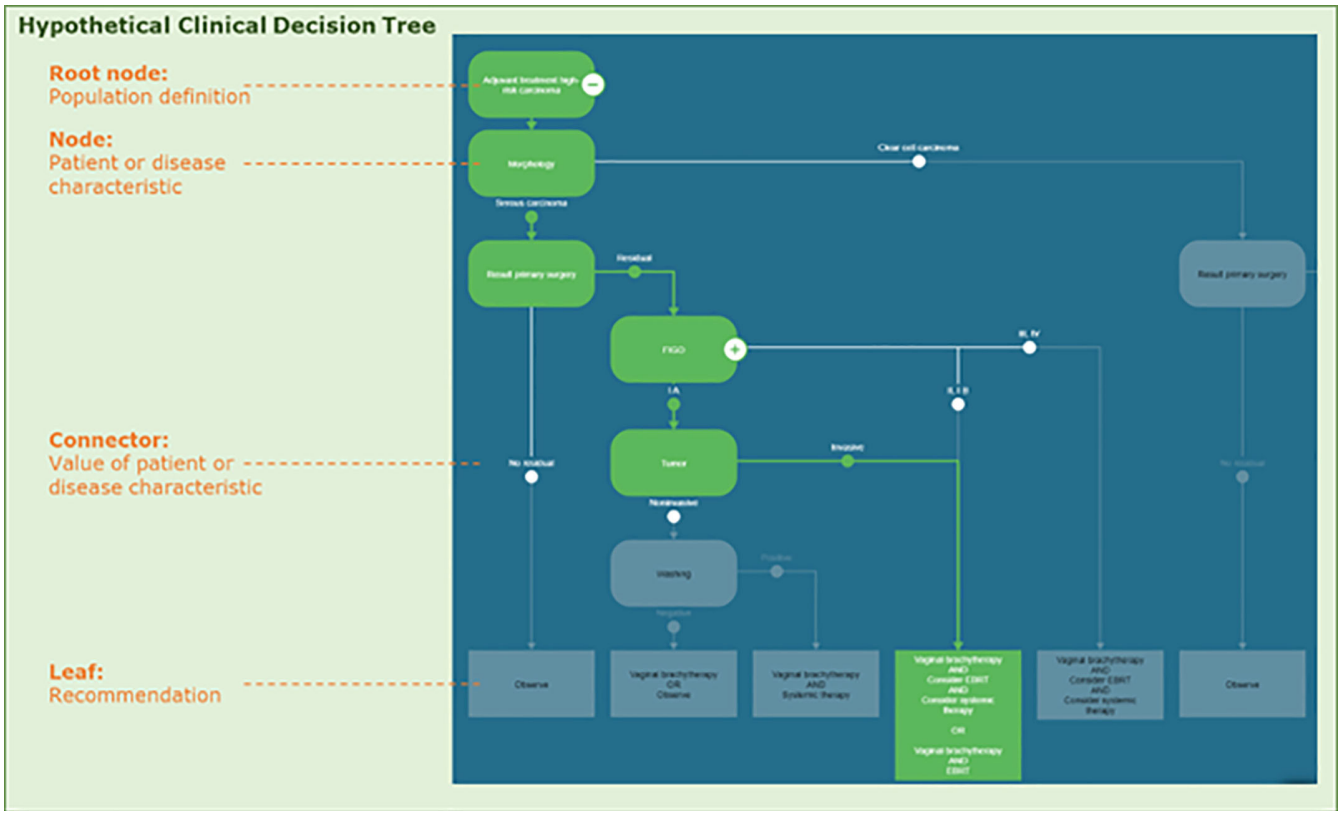

FIGURE 3 Part of a hypothetical clinical decision tree for endometrial cancer. This tree applies to patients after surgery for which, depending on the patient and disease characteristics, a variant of adjuvant treatment can be recommended. The green highlighted path concerns the population: serous carcinoma, residual disease after surgery, FIGO stage IA, and an invasive tumor. Based on these characteristics, the appropriate recommendation is identified by the tree.

and other datasets, even when future data items or datasets are added to the comparison.

2.2 | Real-world data

A valid and reliable dataset should be used to carry out a high-quality evaluation of guideline use, since the quality of these data determines the relevance and usability of the results of the adherence evaluation. Cancer registries are a reliable and comprehensive method system for gathering information about cancer patients. We used the NCR as a source for real-world data (RWD) that most closely overlaps with the population to which the guideline is applied. The NCR is a nationwide registration that has been providing statistics on cancer in the Netherlands since 1989.[28] It includes data on diagnosis, incidence, tumor site (topography), morphology (histology), tumor staging, treatment, and survival. Within the registry, datasets are compiled per tumor type in close consultation with clinicians and other stakeholders. Data are collected for the NCR by specially trained data managers on an ongoing basis, which allows for continuous comparisons with the guideline recommendations. In our project, we initially retrieved a dataset from the NCR as defined by the guideline based on age and morphology (> = 18 years old and pure endometrioid adenocarcinoma, serous carcinoma, and clear cell carcinoma) with an incidence date from 2010 through 2020 ($n = 21\,602$), which has been continuously updated during the project with the most recent available complete cases (incidence years 2021, 2022, etc.) in the NCR.

2.3 | Mapping

2.3.1 | General mapping issues

A challenge in mapping data from different sources, like an information standard and real-world data, are potential differences in the 'level of detail' in which an intervention is described. Interventions can be described in a more generic or in a more specific manner. For example, 'chemotherapy' is more generic than 'Cisplatin', and 'drug therapy' is more generic than 'chemotherapy' or 'targeted therapy'. Similarly, interventions can be described with additional information, like the treatment schedule and dose, which can be added to the 'Cisplatin' intervention. For each specific adherence measurement, agreements must be made on how to deal with this phenomenon to ensure that the results are interpreted correctly. When the guideline provided insufficient clarity in the level of detail, we relied on clinical knowledge from the guideline-working group. Moreover, the

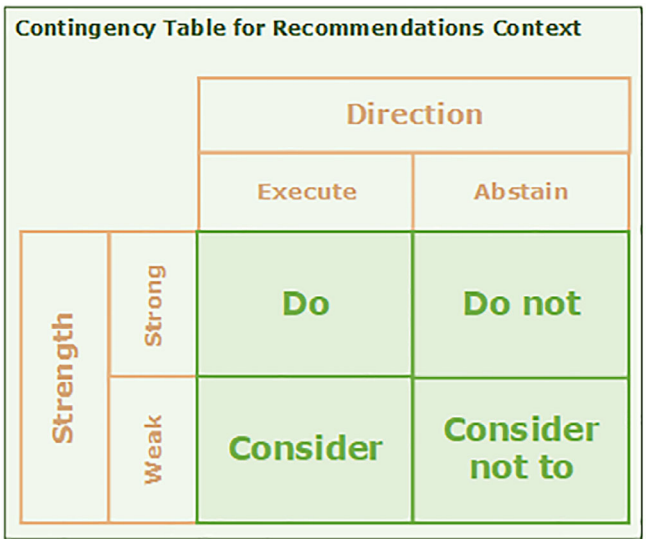

FIGURE 4 Contingency table to address the mandatory context of interventions, which together form a recommendation. Based on this table, 2 × 2 types of context can be defined.

intentions of interventions (eg, *palliative* chemotherapy) often implicitly refer to a limited number of treatment options that are well known to the experts in the relevant field, but not necessarily made explicit in the guideline. The intention itself is not measurable, but again we relied on the guideline working group's clinical knowledge on intend and relevant treatment options to make informed mapping decisions.

Another general mapping issue is related to successive phases in the care pathway. Each medical decision requires specific information relevant for a specific patient population for that specific moment in the care process. Some of this information may be unique to that moment. Some information may be relevant at later stages and will not be updated. Critically, some information may be relevant at later stages and might be updated with new observations in the course of the care pathway. In this instance, for guideline evaluation, it is essential to use the information drawn from the clinical context the recommendation is relevant in. For example, a 'tumor grade' can be determined on a biopsy during the diagnostic phase but also on an excision specimen from surgery during the therapeutic phase. This context sensitivity requires accuracy during mapping. The correct variant of 'tumor grade' must be used in the mapping and aligned with the step in the care pathway.

2.3.2 | Guideline specific mapping issues

A guideline has a specific period over which it is prevailing and all cases from the RWD must have their incidence date in the same period to accurately evaluate adherence. As long as a guideline (version) is prevailing, it will be used as a reference for measuring high-quality care. In our analysis we used the 2011 version of the Dutch endometrial cancer guideline.

2.3.3 | Guideline recommendations specific mapping issues

Guideline recommendations are more than simply proposed interventions. They contain one or more interventions, supplemented with context, namely 'direction', 'strength', and 'other considerations'. The first two types of context are mandatory; a direction of action indicates whether or not to execute a specific recommendation, and the strength indicates how much underlying evidence supports the intervention for this population (Figure 4). The third type of context, 'other considerations', is conditional and may include a wide range of additional information. Consequently, guideline recommendations can be expressed as shown in Equation (1):

$$R = \Sigma(I + D + S) \pm C \quad (1)$$

R, Recommendation, I, Intervention, D, Direction, S, Strength, C, 'Other considerations'.

Equation (1) is the comparison of a guideline recommendations, composed as a formula.

A number of recommendation specific phenomena have to be taken into account in particular during the mapping process. These concern 'weak recommendations', 'other considerations', and 'temporal context'. This subsection describes how we dealt with these phenomena in guideline adherence measurements.

'Weak recommendations' state that an intervention *can be considered*. If an intervention is considered but not performed, then it is likely that, besides the obvious absence of the intervention, also the 'consideration' is not available in the RWD. To classify weak recommendations on adherence, multidisciplinary team (MDT) discussions are used as a proxy. Discussions in MDT meetings are available in the NCR. MDTs are a common instrument to support decision making in oncology, and often base their advice on guidelines.[29,30] In the analyses for adherence classification, it is therefore assumed for all patients that are discussed by a MDT, at least the weak recommendations from the guideline are by definition being discussed. Consequently, if a 'weak recommendation' applies to a case and this patient is discussed in a MDT it is classified as 'adherent' (Table 1).

'Other considerations' is an optional recommendation context provided along with an intervention. There are several acknowledged and often used types of 'other considerations': clinical relevance, safety, patient perspective, professional perspective, cost effectivity, organization perspective, and societal perspective. In the development of guideline recommendations, the 'other considerations' can influence how the recommendations are formulated, they can strengthen or weaken the recommendation, and theoretically even change the direction of the recommendation. 'Other considerations' can be so diverse that they are unlikely to be consistently available a RWD. How to deal with 'other considerations' in adherence measurements must therefore be determined per recommendation.

'Temporal context' can occur within a recommendation in two ways. First within a single recommendation (eg, surgery followed by

TABLE 1 Adherence classification for all five types of recommendations, applied to hypothetical examples of guideline recommendations and real-world data interventions.

Guideline recommendation	Dataset value (s) (eg, NCR)	Adherence classification[f]
Strong, singular intervention recommendation		
Hysterectomy	Hysterectomy (only)	*Adherent*
	Bilateral salpingo-oophorectomy (only)	*Non-adherent*
	Hysterectomy AND Lymph node dissection	*Other*[d]
	'Other'	*Non-adherent*
	'No intervention'	*Non-adherent*
Hysterectomy OR Bilateral salpingo-oophorectomy	Hysterectomy (only)	*Adherent*
	Bilateral salpingo-oophorectomy (only)	*Adherent*
	Hysterectomy AND Lymph node dissection	*Other*[d]
	'Other'	*Non-adherent*
	'No intervention'	*Non-adherent*
Weak, singular intervention recommendation		
Consider hysterectomy	Hysterectomy (only)	*Adherent*
	Bilateral salpingo-oophorectomy (only)	*Non-adherent*
	Hysterectomy AND Lymph node dissection	*Other*[d]
	'Other'	*Non-adherent*
	'No intervention'	*Non-adherent*[a]|*Other*[b]
	Hysterectomy was considered, 'No intervention'[c]	*Adherent*
Strong, multiple intervention recommendation		
Hysterectomy AND Lymph node dissection	Hysterectomy AND Lymph node dissection	*Adherent*
	Hysterectomy (only)	*Other*[e]
	Lymph node dissection (only)	*Other*[e]
	Hysterectomy AND Lymph node dissection AND Radiotherapy	*Other*[d]
	'Other'	*Non-adherent*
	'No intervention'	*Non-adherent*
Weak, multiple intervention recommendation		
Consider hysterectomy AND Consider lymph node dissection	Hysterectomy AND Lymph node dissection	*Adherent*
	Hysterectomy (only)	*Other*
	Lymph node dissection (only)	*Other*
	Hysterectomy AND Lymph node dissection AND Radiotherapy	*Other*[d]
	Hysterectomy AND Lymph node dissection were considered, 'No intervention'[c]	*Adherent*
	'Other'	*Non-adherent*
	'No intervention'	*Non-adherent*[a]|*Other*[b]
Strong and weak, multiple intervention recommendation		
Hysterectomy AND Consider lymph node dissection	Hysterectomy AND Lymph node dissection	*Adherent*
	Hysterectomy (only)	*Other*[a]|*Adherent*[b]
	Lymph node dissection (only)	*Non-adherent*
	Hysterectomy AND Lymph node dissection AND Radiotherapy	*Other*[d]
	Hysterectomy AND Lymph node dissection is considered	*Adherent*
	'Other'	*Non-adherent*
	'No intervention'	*Non-adherent*

[a]Patient was not discussed during multidisciplinary team meeting.
[b]Patient was discussed during multidisciplinary team meeting.
[c]Recommendations provided with the context 'consider' can only be evaluated in case the consideration is explicitly available in the dataset.
[d]Potential over treatment.
[e]Potential under treatment.
[f]All cases deviating from the examples in the table are classified as 'Residual'.

radiotherapy) and second indicates a likelihood of successive treatment in the next phase of care (eg, *neo-adjuvant* radiotherapy). In the first case the challenge lies in combining the correct interventions in a database as one recommendation. Adherence classification mainly depends on meeting the expected time interval between interventions by which they should be considered as executed based on a single recommendation. In the second case, it must be assessed whether the implicitly assumed main treatment (eg, surgery) has indeed been administered within the expected time frame.

2.4 | Analyses

2.4.1 | Adherence classification

CDTs reveal all subpopulations described by the guideline and thus enable adherence measurements for these groups of patients. Ideally, the results of these adherence measurements are presented dichotomously: adherent vs non-adherent, but in reality there are also cases with uncertainty or which cannot be evaluated. In our project, we identified the following adherence classes:

1. Adherent: *the executed interventions are similar to the recommended interventions*.
2. Non-adherent: *the executed interventions deviate from the recommended interventions*.
3. Other: *the relation between the executed interventions and the recommended interventions is unclear*. The class 'other' can optionally be subdivided into:
   a. Potential under treatment: The executed interventions are part of, but not the full set of recommended interventions.
   b. Potential over treatment: The executed interventions are similar to the recommendation, however also additional interventions were executed.
   c. Residual: The relation between the executed interventions and the recommendation is not clear.

Hypothetical examples of adherence classification are given in Table 1 for all combinations of the five types of recommendations and executed interventions from a real-world dataset.

In the prototype dashboard the results of the adherence classification were expressed in percentages. For every subpopulation, thresholds (lower limit and upper limit) of acceptable adherence percentages were determined in advance. If a measured adherence level falls outside these reference values, an alert can be issued.

2.4.2 | Analyses types

Adherence measurements can be performed along two axes in the developed prototype in which the methodology described here is implemented. First, every decision moment in the care pathway, represented by a single CDT, can be evaluated in total and stratified per subpopulation. This process supports, for example, analyses of overall adherence to 'adjuvant treatment' of a disease. In addition, a specific subpopulation from a CDT can be selected and analyzed.

Second, a panel in the prototype provides an overview of all mapped data items in a project. Based on these data items, a selection of a specific population can be made, for which all executed interventions as available in the RWD are presented, including adherence classification.

In addition, more specific analyses can be performed based on additional available variables in the RWD. For example, the NCR also contains data related to time (eg, incidence dates and age) and hospitals (eg, region and type), which makes it possible to analyze trends over time and age or region specific adherence.

3 | RESULTS

Remodeling of the Dutch guideline for endometrial cancer resulted in a total of 10 CDTs. The developed CDTs revealed 22 unique patient and disease characteristics and 46 unique interventions (Table 2), which were captured in the information standard and supplemented with SNOMED-CT codes. It took a time investment for our team (expert informaticians, physicians and researchers) of 32, 12, and 16 h, to respectively: develop and validate CDTs, draw up an information standard, and perform the mapping process.

From a total of 124 NCR data items in the dataset of endometrial cancer, 22 data items were directly mappable with an information standard concept. A total of 10 data items were mappable after editing of NCR data: 'Clinical stage' (calculated from 'cT', 'cN', and 'cM'), 'Number of metastases' (calculated from 'metastases topography'), 'FIGO stage (calculated from 'pT', pN', and 'cM'), 'Age' (calculated from 'date of birth' and 'incidence data'), Lymphadenectomy (created from 'Lymph node dissection'), 'Resection of lymph node metastases' (created from 'Lymph node dissection'), 'Systemic therapy' (created from all available 'systemic therapies'), 'No adjuvant treatment' (created from absence of 'systemic therapy' and 'radiotherapy'), 'Staging result' (adapted from 'completeness of staging'), and 'Re-staging' (created from subsequent 'staging surgeries').

A total of four CDTs could be populated with sufficient data items for measurement of adherence.

Three of these could directly be completely populated with data from the NCR: 'Staging Evaluation', 'Adjuvant treatment for endometrioid type', and 'Adjuvant treatment sereus and clearcell type'. In the CDT 'Primary treatment' two data items were missing in the NCR dataset. The first, 'Location tumor (Parametrium/Vagina)' consequently was added to the NCR dataset from incidence date January 1, 2022 onwards. The second, 'Radical treatment possible (Yes/No)', is a subjective data item which is considered not registrable from EHRs and consequently omitted in the analyses.

The NCR contains 92 additional data items enabling potential relevant filtering options in the prototype dashboard.

TABLE 2 All patient and disease characteristics, and interventions present in the guideline based clinical decision trees. These concepts were recorded in the endometrial cancer information standard. Data items from the NCR dataset were then mapped onto the concepts in the information standards if appropriate.

Clinical decision tree	Patient and disease characteristic	Intervention
Initial diagnostics		
	-	Anamnesis[c]
		Gynecological examination[c]
		Transvaginal ultrasound[c]
Additional diagnostics for suspicion of endometrial carcinoma		
	Ultrasound result[c]	Endometrial sampling[c]
		Endocervical curettage[c]
Staging diagnostics		
	Endometrial sampling result[c]	Individualize treatment[d]
	Grade (biopsy)[c]	Chest x-ray[c]
	Suspicion on invasion[c]	CA-125[c]
		Cystoscopy[c]
		Proctoscopy[c]
		CT-scan[c]
		CT-scan abdomen[c]
		MRI-scan[c]
Primary treatment		
	Clinical stage[b]	Total abdominal hysterectomy[a]
	Histology[a]	Total laparoscopic hysterectomy[a]
	Number of metastases[b]	Bilateral salpingo-oophorectomy[a]
	Tumor location[c]	Radical hysterectomy[a]
	Radical treatment feasible[c]	Lymphadenectomy[b]
		External beam radiotherapy[a]
		Internal radiotherapy[a]
		Resection of lymph node metastases[b]
		Individualize treatment[d]
		Complete staging[a]
		Local policy matching uterine status[d]
		Radiotherapy on metastases[a]
		Systemic therapy[b]
Adjuvant therapy endometrioid type		
	FIGO stage[b]	No adjuvant therapy[b]
	Grade (surgery)[a]	Vaginal brachytherapy[a]
	Age[b]	External beam radiotherapy[a]
	Prior surgery type[a]	Chemotherapy[a]
	Lymph-vascular space invasion[a]	Individualize treatment[d]
Staging evaluation		
	Staging result[b]	Re-staging[b]
Adjuvant treatment sereus and clearcell type		
	FIGO stage[b]	Vaginal brachytherapy[a]
	Grade (surgery)[a]	Radiotherapy[a]
	Lymph-vascular space invasion[a]	Chemotherapy[a]
		External beam radiotherapy[a]

TABLE 2 (Continued)

Clinical decision tree	Patient and disease characteristic	Intervention
Follow-up		
	Years since primary treatment[c]	Physical examination[c]
		Gynecological examination[c]
		Counseling[c]
		Vaginatop cytology[c]
Recurrence diagnostics		
	Recurrence type[c]	Biopsy[c]
		CT-chest/abdomen[c]
		Laboratory blood test[c]
		CA-125[c]
		Cystoscopy[c]
		Proctoscopy[c]
		Bone scintigraphy[c]
Recurrence treatment		
	Recurrence type[c]	Radiotherapy[c]
	Prior treatment[c]	Debulking[c]
	Local recurrence diameter[c]	Exenteration[c]
	Optional treatment radicality[c]	Hormonal therapy[c]
	Progesteron receptor status[c]	Systemic chemotherapy[c]
	Location of regional recurrence[c]	Surgery[c]
		Low-dose medroxyprogesterone (200 mg/day)[c]
		Megestrol acetate (160 mg/day)[c]

[a]Available in NCR.
[b]Derivable from or calculated with NCR data.
[c]Unavailable in NCR.
[d]Non evaluable.

4 | DISCUSSION

We developed and successfully implemented a methodology to continuously monitor guideline adherence. The Dutch endometrial guideline represented as CDTs were linked with real world data from the NCR. The novel methodology expands on earlier work by Hendriks and colleagues. In this work, Hendriks describes a methodology to parse textual guidelines into CDTs, making them both human readable while allowing for computer assisted data driven evaluation of the underlying algorithm. To our knowledge, a methodology to actually leverage this potential in relation to evaluation of guideline adherence has not been reported.

Continuous monitoring of guidelines has a 2 fold relevance for guideline developers. First, an evaluation with RWD enables a quality assessment of the guideline, which leads to the identification of potential improvements in the recommendations. Second, it is possible to quantify how successful a recommendation has been implemented, which enables consideration of changing the implementation strategy. The methodology supports a cyclical process of developing (or updating), implementing, and evaluating guidelines and consequently is a substantial component of a collaborative learning health environment.[31,32] This is supported by the longitudinal nature of the real-world data, that enables analyzes regarding the duration of implementation of interventions ('evidence to practice')[33] and potentially to initiate targeted actions to shorten the anticipated time span.

In addition to measuring adherence, a structured analysis of the guideline using RWD has another important advantage. Real-world data analyses address the generalizability of the research conclusions that guideline recommendations are initially based upon.[34] Guideline recommendations are preferably based on the highest level of evidence available; often (systematic reviews of) RCTs. However, patient populations included in many RCTs do not adequately reflect the patients in the real world, for instance by only including patients of a certain age, race, or social background. In contrast, real-world datasets may include older patients with more comorbidity and challenging social and demographic circumstances by default. Including all patients from the applicable population in guideline adherence analyses results in a more complete understanding of the impacts of the recommendations under study in real life.

With a guideline as a reference, this methodology supports an additional structured and data driven analysis of quality of care

provided to the population of interest. By remodeling the guideline text into CDTs, all defined subpopulations are identified and provided with valid recommendations. The identified subpopulations can then be further stratified using additional variables from the real-world dataset that are not included in the guideline-based CDTs; enabling exploration of more detailed subpopulations that may alter adherence levels. For example, if 'age' is not a criterion for a specific recommendation, but RWD analysis shows that 90% of all non-adherent cases are over 70 years of age, this may be relevant information to adjust the guideline by adding 'age' as a steering variable. Such analyses identify patient and disease characteristics that influence decision making in clinical practice, but are not mentioned in the recommendations. Evidently, an extensive dataset supports a multitude of possible analyses, and hence the potential to gain relevant new insights, contributing to real-world evidence.[35]

To the best of our knowledge, no studies have been conducted to determine the desired level or range of adherence to oncological guidelines. Guideline adherence measurements as described here should prompt further discussion of acceptable adherence levels. Determining factors for guideline adherence are (1) patient-related (eg, comorbidities, age, patient preferences) or influenced by (2) the level of evidence underlying a recommendation. Guideline recommendations are formulated at the population level and are evaluated by clinicians to support optimal decision making for individual patients, taking into account their specific characteristics. Definitive conclusions for practice are only possible if one can distinguish between justified and non-justified non-adherence to guidelines.[36] Conscious deviation from guideline recommendations should be considered as justified and contributing to quality of care. Accurate clinical reporting of the motivation to deviate from a recommendation is of high importance because it is a key factor in the assessment of quality of care. In addition, it is plausible that a lower level of evidence for an intervention, in principle resulting in a weak recommendation, generally leads to lower levels of adherence. A low level of evidence leaves more room for physicians to ensure optimal care for individual patients, based on practice-based evidence. For every identified subpopulation in a guideline, the aforementioned criteria should be taken into account to debate on an optimal range or proportion of adherence levels. The methodology and prototype developed in this project can generate insights that advance this debate.

By analyzing all variables from a guideline, it is possible to identify the steering parameters for clinical decision making; these should be regarded as the minimum data set required for assessing adherence accurately. This project used NCR data, which is a uniform, high-quality dataset. However, users should be aware of the nature and origin of the data they are using and interpret the generated results accordingly. RWD is retrospective in nature and may be prone to misclassification and selection bias, which has to be considered in the analyses and interpretation of the results. In other words, the quality of the real-world data determines the validity and reliability of the analyses.[37]

In addition to data quality, data availability is another major hurdle to come to a self-improving healthcare system. For example, we were only able to populate 4 out of 10 CDT's using the Dutch Cancer Registry. As it stands, EHR data, without validation and cleansing, is unfit for secondary use, and manually collecting more data in cancer registries is not scalable. A path forward should come from a combination of policy and technical innovation. On the local, national and European level there are already initiatives to improve data availability in the healthcare space. More use of semantic and technical information standards, together with functional, legal, and financial incentives to implement and use these standards in a meaningful way should provide a steady path toward better availability of high-quality data for secondary use. However, even with the proper standards, incentives and policies in place, standardized reporting may not be of added value for physicians in their workflow and only add registration burden. Technical innovations such as, for example, large-language models trained on the medical domain may be of use to support physicians with encoding their free-text note-taking without adding extra registration burden.

The described methodology focuses on the guideline working group as primary users and a single oncological condition. However, CDTs are also suitable for other conditions and diseases, which enables application in the newly developed adherence measurements methodology. Underlying calculations according to the methodology remain identical for other diseases. Although other areas of healthcare may be organized differently. For example, MDTs are standard within oncology, which is why we were able to include them as a proxy in our classification of adherence for 'weak recommendations'. Moreover, the prototype can be continuously supplemented with more recent cases, and thus identify (adherence) trends over time. Additionally, when hospital and geographical data is included in the RWD, it enables analyses between (type of) hospitals, regions, and countries. Because CDTs identify all relevant subpopulations, this method is also able to continuously monitor adherence for all these groups if a guideline is (modularly) updated. These aspects may enhance guideline development, guideline quality assurance and improvement, and will ultimately improve quality of care.

The method was implemented through a dashboard prototype that requires further testing to evaluate its functionality. The prototype is currently capable of applying the method described here and performing additional analyzes based on filtering with input from additional data from the NCR. It enables interactive usage and supports various analyses. By adjusting the presentation of the results this application could be useful for other target groups, like patients, policy makers, healthcare professionals and the general audience.

5 | CONCLUSION

We successfully demonstrated continuous monitoring of guideline adherence is possible using this novel methodology. This creates a closed loop in guideline development, implementation, and evaluation and healthcare delivery, making it a valuable part of the evidence ecosystem.[8]

Having continuous insights into guideline adherence is a crucial, first step to provide high-quality care. This methodology could be used as a base and extended to cover health outcomes (as defined in PICOs, such as adverse events, progression-free survival, and overall survival) as actual endpoints to evaluate quality of care. This would add even more value to this application to be further used in a value-based healthcare environment.[38]

This method closes the circle of the evidence ecosystem by linking knowledge sources to real-world data. The knowledge source is thus continuously fed with results from clinical practice and forms the linking cog of a learning healthcare system. Availability, completeness and quality of real world data determines the validity and reliability of the methodology, and should have ongoing attention when further implementing the evidence ecosystem.

CONFLICT OF INTEREST STATEMENT

The authors declare that there is no conflict of interest.

ORCID

Kees C.W.J. Ebben https://orcid.org/0000-0003-1069-2357

REFERENCES

1. Graham R, Mancher M, Miller D, Greenfield S, Steinberg E. *Clinical Practice Guidelines we Can Trust; Institute of Medicine*. Washington (DC): National Academies Press (US); 2011:300.
2. Glatzer M, Panje CM, Sirén C, Cihoric N, Putora PM. Decision making criteria in oncology. *Oncology*. 2020;98:370-378.
3. Charara RN, Kreidieh FY, Farhat RA, et al. Practice and impact of multidisciplinary tumor boards on patient management: a prospective study. *J Glob Oncol*. 2017;3:242-249.
4. Heins MJ, de Jong JD, Spronk I, Ho VKY, Brink M, Korevaar JC. Adherence to cancer treatment guidelines: influence of general and cancer-specific guideline characteristics. *Eur J Pub Health*. 2017;27(4):616-620.
5. Fischer F, Lange K, Klose K, Greiner W, Kraemer A. Barriers and strategies in guideline implementation—a scoping review. *Healthcare*. 2016;4(3):36.
6. van de Klundert J, Gorissen P, Zeemering S. Measuring clinical pathway adherence. *J Biomed Inform*. 2010;43(6):861-872.
7. Kardas P, Aguilar-Palacio I, Almada M, et al. The need to develop standard measures of patient adherence for big data: viewpoint. *J Med Internet Res*. 2020;22(8):e18150.
8. Vandvik O, Brandt L. EVIDENCE SYNTHESIS ECOSYSTEM SERIES future of evidence ecosystem Series: evidence ecosystems and learning health systems: why bother? *J Clin Epidemiol*. 2020;123:166-170.
9. Cohen O, Brauer PR, Judson BL, Burtness BA, Earles J, Mehra S. Guideline - adherence in advanced stage head and neck cancer is associated with improved survival - a national study. *Oral Oncol*. 2022;125:105694.
10. van den Berg I, van de Weerd S, van Klaveren D, et al. Daily practice in guideline adherence to adjuvant chemotherapy in stage III colon cancer and predictors of outcome. *Eur J Surg Oncol*. 2021;47(8):2060-2068.
11. Jochum F, De Rozario T, Lecointre L, et al. Adherence to European ovarian cancer guidelines and impact on survival: a French multicenter study (FRANCOGYN). *Int J Gynecol Cancer*. 2021;31(11):1443-1452.
12. Elegbede AA, Gibson AJ, Fu H, et al. Real-world adherence to guideline-recommended treatment for small cell lung cancer. *Am J Clin Oncol*. 2020;43:236-242.
13. White KM, Seale H, Harrison R. Enhancing ovarian cancer care: a systematic review of guideline adherence and clinical variation. *BMC Public Health*. 2019;19:1-13.
14. Willems SM, Abeln S, Feenstra KA, et al. The potential use of big data in oncology. *Potential Use Big Data Oncol*. 2019;98:8-12.
15. Chew SY, Koh MS, Loo CM, Thumboo J, Shantakumar S, Matchar DB. Making clinical practice guidelines pragmatic: how big data and real world evidence can close the gap. *Annal Acad Med Singapore*. 2018;47(12):523-527.
16. Khozin S, Blumenthal GM, Pazdur R. Real-world data for clinical evidence generation in oncology. *J Natl Cancer Inst*. 2017;109(11):djx187.
17. Parkin DM. The role of cancer registries in cancer control. *Int J Clin Oncol*. 2008;13:102-111.
18. Das B, Limin AE, Ae XC, Feuer EJ, Linda AE, Pickle W. A new method to evaluate the completeness of case ascertainment by a cancer registry. *Cancer Causes Control*. 2008;19:515-525.
19. Peleg M. Computer-interpretable clinical guidelines: A methodological review. *J Biomed Inform*. 2013;46:744-763.
20. Boxwala AA, Rocha BH, Maviglia S, et al. A multi-layered framework for disseminating knowledge for computer-based decision support. *J Am Med Inform Assoc*. 2011;18(Suppl 1):i132-i139.
21. Tso GJ, Tu SW, Oshiro C, et al. Automating guidelines for clinical decision support: knowledge engineering and implementation. *AMIA Annu Symp Proc*. 2016;2016:1189-1198.
22. Hendriks MP, Verbeek XAAM, van Vegchel T, et al. Transformation of the National Breast Cancer Guideline into Data-Driven Clinical Decision Trees. *J Clin Oncol Clin Cancer Inf*. 2019;3:1-14.
23. IKNL. ALERTNESS project page 2021. Available from: https://iknl.nl/en/guidelines/alertness;-structural-signaling-for-up-to-date-gui
24. Eriksen MB, Frandsen TF. The impact of patient, intervention, comparison, outcome (PICO) as a search strategy tool on literature search quality: a systematic review. *J Med Libr Assoc*. 2018;106:420-431.
25. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008;336(7650):924-926.
26. Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR guiding principles for scientific data management and stewardship. *Sci Data*. 2016;3:160018.
27. Lee D, de Keizer N, Lau F, Cornet R. Literature review of SNOMED CT use. *J Am Med Inform Assoc*. 2014;21(e1):e11-e19.
28. Schouten LJ, Jager JJ, van den Brandt PA. Quality of cancer registry data: a comparison of data provided by clinicians with those of registration personnel. *Br J Cancer*. 1993;68:974-977.
29. Pillay B, Wootten AC, Crowe H, et al. The impact of multidisciplinary team meetings on patient assessment, management and outcomes in oncology settings: a systematic review of the literature. *Cancer Treat Rev*. 2016;42:56-72.
30. Walraven JEW, van der Hel OL, van der Hoeven JJM, Lemmens VEPP, Verhoeven RHA, Desar IME. Factors influencing the quality and functioning of oncological multidisciplinary team meetings: results of a systematic review. *BMC Health Serv Res*. 2022;22:829.
31. Pomare C, Mahmoud Z, Vedovi A, et al. Learning health systems: a review of key topic areas and bibliometric trends. *Learn Health Systems*. 2022;6(1):e10265.
32. Ellis LA, Sarkies M, Churruca K, et al. The science of learning health systems: scoping review of empirical research. *J Med Int Res Med Inf*. 2022;10(2):e34907.
33. Beauchemin M, Cohn E, Shelton RC. Implementation of clinical practice guidelines in the health care setting: a concept analysis. *ANS Adv Nurs Sci*. 2019;42(4):307-324.
34. Sedgwick P. External and internal validity in clinical trials. *BMJ*. 2012;344:e1004.
35. Ramamoorthy A, Huang SM. What does it take to transform real-world data into real-world evidence? *Clin Pharmacol Ther*. 2019;106:10-18.

36. Arts DL, Voncken AG, Medlock S, Abu-Hanna A, van Weert HC. Reasons for intentional guideline non-adherence: a systematic review. *Int J Med Inform*. 2016;89:55-62.
37. Blacketer C, Defalco FJ, Ryan PB, Rijnbeek PR. Increasing trust in real-world evidence through evaluation of observational data quality. *J Am Med Inform Assoc*. 2021;28(10):2251-2257.
38. Porter ME. Value-based health care delivery. *Ann Surg*. 2008;248(4):503-509.

How to cite this article: Ebben KCWJ, de Kroon CD, Schmeink CE, et al. A novel method for continuous measurements of clinical practice guideline adherence. *Learn Health Sys*. 2023;7(4):e10384. doi:10.1002/lrh2.10384