The challenge of defining suffering in pain.
A systematic review on pain-related suffering using natural language processing

Niklas Noe-Steinmüller¹, Dmitry Scherbakov², Alexandra Zhuravlyova², Tor D. Wager¹, Pavel Goldstein²*, Jonas Tesarz¹*

¹ Department of General Internal Medicine and Psychosomatics, University Hospital Heidelberg, Germany
² School of Public Health, University of Haifa, Israel
³ Dartmouth College, Hanover, NH, USA

* These authors contributed equally to this manuscript

Author contributions: Conceived and designed the study: NNS, JT and PG; analyzed the data: NNS, JT DS, and AZ; wrote the paper: NNS, JT and DS; critically revised the manuscript: NNS, JT, DS, PG and TW. All authors read and approved the final manuscript.

Pages: xxx
Number of tables: 2
Number of figures: 3

Address correspondence to:
Prof. (apl.) Dr. Jonas Tesarz, Department of General Internal Medicine and Psychosomatics
Medical Hospital, University of Heidelberg
Im Neuenheimer Feld 410,
D-69120 Heidelberg, Germany
Tel.: +49 6221 56 37862; Fax: +49 6221 56 5749
E-mail: jonas.tesarz@med.uni-heidelberg.de (JT)
Abstract

Understanding, measuring, and mitigating pain-related suffering is one of the key challenges for both clinical care and pain research. However, there is no consensus on what exactly the concept of suffering includes and it is often not precisely operationalized in empirical studies. Here, we 1) systematically review the conceptualization of pain-related suffering in the existing literature, 2) develop an operational definition grounded in multidimensional concepts of suffering, and 3) use machine learning to cross-validate the results.

We identified 111 articles in systematic search of Web of Science, Pubmed, PsychInfo and PhilPapers for peer-reviewed articles containing conceptual contributions about the experience of suffering. We developed a new procedure for extracting and synthesizing study information based on cross-validation of qualitative research results with an artificial intelligence-based approach grounded in Large Language Models (LMM) and topic modeling.

We derived a definition from the literature that is representative of current theoretical views and describes suffering as a psychological experience of distress grounded in the affected person’s conception of physical, affective, social, and existential threats. We also offer a systematic summary of the literature in a multilevel multidimensional model with eight core dimensions, which are divided into 18 subdimensions at the second and 49 subdimensions at the third level.

Our data show that suffering in the context of pain is a multidimensional phenomenon and can be conceptualized at different levels. The present analysis provides a roadmap for further theoretical and empirical development.

Results:

Key words: pain, suffering, definition, natural language processing, machine learning, ChatGPT, GPT-3, topic modeling, LDA
Introduction

Suffering as a result of pain is one of the key challenges for both clinical care and pain research (De Ridder, Adhia, & Vanneste, 2021; Sullivan, Sturgeon, Lumley, & Ballantyne, 2023; Turk & Wilson, 2009). However, there is currently no consensus on how to conceptualize pain-related suffering.

The most frequently cited definition of suffering has been put forward by Eric Cassell (1982), who describes it as “the state of severe distress associated with events that threaten the intactness of the person” (p. 640). However, although this definition is widely considered as a milestone and merits recognition for introducing the concept of pain-related suffering into pain research and clinical practice, it has repeatedly been criticized for excluding certain populations (Tate, 2020), and for lacking clarity (Bueno-Gómez, 2017).

There is no established instrument to measure suffering in the clinical context, and in a recent review of existing instruments most have been criticized for severe methodological shortcomings (Gutiérrez-Sánchez, Gómez-García, Cuesta-Vargas, & Pérez-Cruzado, 2020). An exception is Büchi and Sensky’s (1999) approach to measure suffering nonverbally with a pictogram representing pain and its impact on the intactness of the self, which is inspired by Cassell’s (1982) definition. However, it has been stressed by various authors that suffering is a multidimensional experience (del Giglio, 2020; Montoya-Juarez et al., 2013; Smith, Amella, Edlund, & Mueller, 2014) and while a pictorial representation might be a valuable approximation it cannot replace a differentiated measurement of different dimensions of pain-related suffering.

Furthermore, it appears that different empirical studies measured different constructs under the label “pain-related suffering”. For instance, Wade, Riddle, Price, and Dumenci (2011) operationalized suffering using 5 visual analogue scales measuring pain related depression, anxiety, frustration, fear and anger. By contrast, Baines and Norlander (2000) measured suffering by directly asking patients with chronic pain to indicate the extent of their own suffering in different areas (such as spiritual distress or concern for loved ones). This illustrates the multidimensional character of suffering and the importance of a definition that integrates distinct but interrelated facets of suffering. However, the conceptual models described above have not been empirically validated, and as a result, a multidimensional conceptualization of pain-related suffering supported by empirical data is still lacking.

The current study aims (1) to systematically search the existing literature for conceptualizations of pain-related suffering, (2) to use manual qualitative research methods to synthesize the literature into an integrative definition and develop a multilevel and multidimensional model of pain-related suffering, and finally (3) to use a machine learning approach to cross-validate both the model and the integrative definition. Accordingly, our first research question is how pain-related suffering can be defined according to the current literature. Our second research question is how pain-related suffering can be described in a multilevel and multidimension model.
Methods

Procedure

This study is based on a systematic review using natural language processing to analyze the concept of pain-related suffering in the current literature. In the first step, a transdisciplinary systematic literature search was conducted to identify relevant studies on pain-related suffering. In the second step, a systematic data extraction of verbatim quotations with relevant information for the conceptualization of suffering was conducted. The verbatim quotations were translated into key words, which served as a basis for (a) the derivation of common core elements across different definitions, as well as for (b) the identification of the different dimensions of pain-related suffering. In the third step, natural language processing was used to cross-validate the results. Topic modeling (Blei, Ng, & Jordan, 2003) was applied to the text corpus to cross-validate the different dimensions of pain-related suffering identified in the second step. To cross-validate the integrative definition of suffering obtained in the second step, we used ChatGPT and GPT-3 models (Brown et al., 2020). The procedure and analytical plan of the current study were predesigned. A detailed description of protocol is available from the authors upon request.

Step 1: Systematic search

This review was performed and reported according to the recommendations of the PRISMA statement (Page et al., 2021) when appropriate. We searched Web of Science, Pubmed, PsychInfo, and PhilPapers. The cutoff date of the search was September 7th, 2021. In addition, a citation search on the included articles was performed. The search strings were adapted for each database (see Appendix A). Two reviewers (N.N.S. and a scientific assistant) independently scanned the titles and abstracts of eligible studies and, in a second step, the full text articles, to determine whether the articles met the selection criteria. Disagreements between the two reviewers were resolved by discussion, and if agreements could not be achieved, a third reviewer was consulted (J.T.). Study selection was performed using Covidence systematic review software (Veritas Health Innovation, Australia).

We selected all studies that focused on the conceptualization of suffering in the context of physical pain. To be included, an article had to (1) use the term “suffering” in a conceptual way, which we took to be the case if and only if this article specified an answer to at least three of 14 predefined conceptual questions in our data extraction form (see Appendix E). In addition, the article also had to (2) discuss physical pain, or it had to be evident that the phenomenon of interest involved physical pain. We only included (3) peer-reviewed articles (4) published in English.

As this study focused on human suffering as a consequence of pain experiences, studies were excluded if they (1) only dealt with the suffering of animals or (2) only discussed the effect of suffering on others, e.g., on a caregiver. We did not exclude articles because of their date of appearance. A flow diagram of the study selection process is shown in Appendix B.

Step 2: Data extraction and synthesis with manual qualitative research methods

We developed a systematic procedure for manually extracting and analyzing data from the selected articles, with the goal of identifying the common core elements and dimensions across different conceptualizations of suffering. In addition to general article information (e.g.,
author, discipline, or outcome, see Appendix D), data extraction was structured by conceptual questions regarding pain-related suffering (e.g., how each article defined suffering, what dimensions were specified, or what antecedents or consequences of suffering were postulated, see Appendix E). For the conceptual questions, verbatim quotations were extracted from the articles that were then summarized into key words — i.e., words from the quotations representing the core elements of suffering according to the respective article. The key words summarizing the extracted verbatim quotations were used for all further analyses using the manual text processing method.

Data extraction and the summarization into key words were conducted by one of the authors trained in psychology and phenomenological philosophy (N.N.S). The summarization of verbatim quotations into key words was also conducted independently by a second author trained in internal medicine and psychosomatics (J.T.). Disagreements were resolved in direct discussion.

**Research question 1: How is suffering defined in current pain research?**

Sixty of the 111 selected articles explicitly provided a definition of suffering. For those articles, we identified common elements of the definitions by applying a modified version of concept analysis (Walker & Avant, 1995), a well-founded method of theory development that has already been used to examine the concept of suffering (Best, Aldridge, Butow, Olver, & Webster, 2015).

For this, we used the key word summaries of the 60 extracted definitions. If two or more key words from different definitions (1) were identical, (2) were based on the same lemma (e.g., “body” and “bodily”), or (3) were clearly synonymous in the present context (e.g., “bodily” and “physical”), they were given the same code. It was then counted how often each code was used. Only key words that were part of at least two definitions, i.e. whose code was used at least twice, were kept and summarized into our integrative definition of pain-related suffering (see Table 1).

**Research question 2: Summarizing the literature into a multilevel and multidimensional model**

To summarize the literature into a multilevel and multi-dimensional model, we first used the procedure described above for summarizing the definitions. However, this time we used the verbatim quotations and key words that were extracted for answering another conceptual question from our data extraction form (“What dimensions or types of suffering are differentiated?”). Eighty-two articles specified an answer to this question. As described above for the core elements of the definitions, it was examined if different key words referred to the same core dimension of suffering, and the number of articles describing this dimension was counted. If a dimension was mentioned by at least two articles, it was added to the first (highest) level of our model.

To anticipate, we identified eight core dimensions of suffering with this procedure. However, as can be seen in Figure 1, those dimensions are very broad and call for further differentiation.

In a next step, we applied the above-described categorization procedure to the key words extracted for yet two other conceptual questions from our data extraction form. We analyzed which antecedents and consequences of suffering were mentioned repeatedly in the literature. Seventy-five articles provided an answer to at least one of these questions. This
way, beyond the explicit mentioning of dimensions of suffering, additional information from the full texts could be used for our model. The result was again a list of core aspects of suffering, although, in this case, not core dimensions or defining core elements, but core antecedents and consequences. These were allocated as 58 subdimensions to the eight core dimensions identified in the first step of the dimensional modeling based on theoretical considerations.

In an additional step – for further exploratory qualitative reasons – we added an intermediary level to the model. Based again on theoretical considerations, the 58 core antecedents and consequences were grouped into 22 subdimensions, which were in turn allocated to the eight core dimensions.

**Step 3: Cross-validation using machine learning**

Since a qualitative approach, despite all efforts to be objective and systematic, is inevitably based to some extent on subjective judgements, we used the possibilities of machine learning via natural language processing (NLP) for cross-validation. **For this purpose, we used unsupervised learning techniques to process our text corpora.**

**Research question 1: How is suffering defined in current pain research?**

To validate the integrative definition of pain-related suffering obtained with manual qualitative methods (Table 2), we applied GPT-3.5, which is based on the generative pre-trained transformer (GPT) large language model (Brown et al., 2020), release from January 30, 2023, to generate a definition of suffering. For this purpose, we used the same verbatim quotations extracted for the manual analysis described above.

In addition, we used text-davinci-003 (GPT-3 model to generate an integrative definition of suffering based on the full texts of 111 articles, to validate the extracted verbatim quotations. For each mentioning of the term “suffering” found in text we extracted 25 words before and after the term “suffering”. The range of ±25 words was chosen because this is approximately the length of two average sentences in scientific publications (Gotti, 2008). This resulted in 8910 matching text strings each one consisting of 51 words. The resulting 8910 strings were fed iteratively into GPT-3, asking it to define suffering from the new portion of string while taking into account already generated definition from the previous iteration.

To estimate how closely the resulting definitions from GPT language models matched our reference definition obtained manually, we used three commonly used paragraph proximity metrics, namely *vector similarity*, using embeddings from text-embedding-ada-002 GPT-3 model (e.g., Le & Mikolov, 2014), as well as lexical similarity, using ROUGE-1 F1 score based on tokens as described in Lin, C. Y. (2004, July), as well as a variation of ROUGE-1 F1 score using only lemmas of nouns and adjectives. The results can be found in Table 3.

**Research question 2: Summarizing the literature into a multilevel and multidimensional model**

To cross-validate the dimension and levels of suffering, we used topic modeling, an unsupervised machine learning approach for the automatic discovery of topics in large text corpora (Blei et al., 2003). For a detailed description see Bittermann (2022).
**Brief description of the NLP methods**

For scientific application, Latent Dirichlet Allocation is the most frequently used algorithm for topic modeling (Liu, Tang, Dong, Yao, & Zhou, 2016). It rests on what is called the “bag-of-words’ assumption” (Blei et al., 2003, p. 994): It is assumed that the order of words is neglectable and the algorithm focuses on the quantity and co-occurrence of words within a given document. If a group of words appears together in many documents of a text corpus, those words are assigned a high probability of belonging to the same topic. As a result, one can identify lists of words, known as top words, that most likely are characteristic for the actual topics within the text corpus.

The top words depend on several decisions that must be made before feeding the data to the algorithm. The model is based on two hyperparameters: \( \alpha \) defines the number of topics each document of the text corpus is likely to contain (\( \alpha \rightarrow 0 \) only one topic per document, higher values indicate multiple topics), \( \varepsilon \) defines, how likely that a word can appear in more than one topic (\( \varepsilon \rightarrow 0 \) the topics do not share any words at all). Furthermore, the number of topics \( K \) must be determined in advance, as well a \( \lambda \), a parameter determining the display of top words.

We used the Latent Dirichlet Allocation (LDA) algorithm to generate topic models. For the implementation in R (Version 4.1.2.), we used the packages “quanteda”, “text2vec” and “tidyverse”.

In text pre-processing, stop words, special characters, numbers, as well as punctuation were removed from the corpus. We also tested the effect of both word-stemming (Xu & Croft, 1998) and the use of n-grams (Jurafsky & Martin, 2021) in a probatory model, but neither technique improved the interpretability of the data.

We employed identical word windows to those utilized in the analysis conducted with the text-davinci-003 model mentioned earlier. It is worth noting that each word window was treated as an individual document by the algorithm. This approach allowed us to focus on specific sections of the articles rather than seeking topics across the articles as a whole.

Data cleaning was conducted by two researchers (J.T. and N.N.S) based on the results of a “standard model”, which extracted 30 topics from the corpus using standard settings with \( \alpha = 1.67, \varepsilon = 0.03 \) and \( \lambda = 0.4 \) (Griffiths & Steyvers, 2004; Sievert & Shirley, 2014). In an iterative process, we removed not only undetected stop words, but also terms that were expressive of the shared (academic) style of the examined articles, but did not carry significant meaning for our research question, as well as proper names.

Since neither the number of topics, nor lambda, nor the hyperparameters can be determined theoretically, N.N.S. and J.T. independently rated probatory extractions with varying settings and agreed on the defining of 30 topics, using standard values for lambda and other hyperparameters.

**Comparing the topics and the manually identified dimensions**

To compare the results of topic modeling to the dimensions of pain-related suffering identified by manual methods, we developed a variation of what is called topic labeling. Usually, topics are given labels based on what is a shared element of most top words (see for instance, Lau, Grieser, Newman, & Baldwin, 2011), this can be done by humans or automatically. Giving a straightforward example, if the top words are “chair”, “table”, and
“couch”, an obvious choice for a topic label would be “furniture”. We changed this procedure slightly, insofar as we not only looked for a commonality between the top words, but also determined, which of the topics showed a clear similarity to one of the 22 second-level dimensions of suffering identified by our multilevel multidimensional model.

Two researchers (J.T. and N.N.S.) conducted the comparison between the manual and the machine learning approach independently. The agreement rate between both researchers was 90%, i.e., for 27 of the 30 identified topics there was agreement about whether they corresponded to one of the second-level dimensions of the model. Disagreements were resolved in direct discussion.
Results

Step 1: Study characteristics

The initial database search identified 6179 articles. After the removal of duplicates, 3813 articles remained. Three-thousand-four-hundred-thirty-four of these were excluded based on title and abstract screening, because they did not meet the selection criteria. Of the remaining 379 articles, only 271 could be retrieved, and were examined in detail. In the full text screening, another 160 articles were removed from the text corpus because they did not meet the inclusion criteria. One-hundred-and-eighteen articles were excluded because they did not use the term suffering in a conceptual way, i.e., did not provide an answer to at least three of the conceptual questions shown in Appendix E. Twenty-two articles were excluded because pain was not sufficiently discussed in their notion of suffering. A complete report about all reasons for exclusion in the full-text screening can be seen in Appendix C.

Of the remaining 111 articles from 26 different countries, 45 were theoretical studies in an essay format. Forty-seven articles were empirical studies, of which 23 used qualitative, 22 quantitative, and two mixed methods. The remaining 19 articles were reviews (eleven narrative, eight systematic).

Thirty-six articles (32%) came from the field of nursing studies, in particular from the study of palliative care (eleven articles; 10%). Thirty-five articles (32%) came from different disciplines within the humanities, in particular from the field of medical ethics (22 articles; 20%). Twenty-two articles (20%) came from behavioral and/or biological sciences, in particular from psychology (twelve articles; 11%). Finally, 16 articles (14%) came from medical science, in particular from palliative medicine (six articles; 5%). Figure 2 gives an overview of all disciplines present in the text corpus. Other study characteristics are shown in Table 1.

Step 2: Data extraction and synthesis with (manual) qualitative research methods

Research question 1: How is suffering defined in current pain research?

All articles contributed to the conceptualization of pain-related suffering, however, only 60 offered an explicit definition. Bringing together the common core elements of suffering that were mentioned consistently across these different articles, we propose an integrative definition of pain-related suffering shown in Table 2.

Figure 3 shows in detail how many articles mention each core element of suffering. For instance, 20 articles referred to the experiential, affective and existential character of suffering, while only three articles referred to ‘identity’ or the dynamic character of this experience.

Research question 2: Summarizing the literature into a multilevel and multidimensional model

The manual approach identified eight basic, first-level dimensions of pain-related suffering, that could be further differentiated into 22 second-level and 58 third-level dimensions of suffering. A visualization of this first version of our model can be seen in Figure 1. Appendix F shows the model in table format.

Among the first-level dimensions, the existential and the social aspect of suffering were most prominent in the literature. Thirty-seven articles (33%) mentioned experiences...
that could be summarized as existential concerns, such as dissatisfaction with life or the feeling of having lost one’s future. Thirty-six articles (32%) mentioned social experiences, such as isolation or the loss of autonomy. Twenty-nine articles (26%) described experiences that were summarized as the personal dimension of suffering, such as the experience of a threat to the self. Twenty-seven articles (24%) pointed out the physical dimension of suffering, e.g., by pointing out the importance of general somatic symptoms. Twenty-two articles (20%) referred to aspects that belong to the affective dimension of suffering, such as depression. Twenty articles (18%) described cognitive aspects of suffering, such as the perceived inability to cope with one’s pain. The least frequently mentioned first-level dimensions of suffering were the cultural and the spiritual. Fourteen articles (13%) described experiences that could be summarized as cultural dimension of suffering, such as the experience of being objectified by medicine. Seven articles (6%) mentioned spiritual concerns as an important aspect of suffering.

**Step 3: Cross-validation using machine learning**

**Research question 1: How is suffering defined in current pain research?**

Results indicate a very high semantic (vector) and a moderate lexical similarity between our manual integrative definition of pain-related suffering and the definitions provided by the GPT 3.5 and GPT-3 large language models. The definition given by GPT 3.5 based on the verbatim quotations extracted by the authors had a vector similarity value of 0.963 and a ROUGE-1 value of 0.471 (0.474 for the variant using only lemmas of nouns and adjectives) compared to the integrative definition. The definition given by GPT-3 based on the full texts had a vector similarity value of 0.941 and a ROUGE-1 value of 0.393 (0.328 for the variant using only lemmas of nouns and adjectives) compared to the integrative definition. Table 3 summarizes the results. The exact definitions provided by the large language models, can be seen in Appendix G.

**Research question 2: Summarizing the literature into a multilevel and multidimensional model**

**The topic model based on the LDA algorithm**

Twenty-five of the 30 top word lists extracted by our search algorithm constitute clearly interpretable topics. In the other five cases, no distinct topic could be detected. Table 4 shows the topic model. It lists all detected topics as well as the ten top words.

Among the 25 top word lists with clearly interpretable topics, in four cases both researchers agreed that two interpretations are equally viable, i.e., that the respective top word could be interpreted as either of two topics. This concerns the top word lists T1 (personal development/inability to cope), T7 (existential concerns/fear of death) T11 (concern for other/spiritual concern), as well as T30 (loss of autonomy/existential concern). Therefore, in Table 4, in those top word lists, the terms that indicate the second theme are underlined, and, in the last column, a second label is given.

**Revisiting the manually obtained multilevel and multidimensional model**

Of the 25 clearly interpretable top word lists, 20 corresponded to one of the 22 subdimensions of the manually obtained multilevel multidimensional model and were
labelled accordingly. However, because in two cases, two top word lists corresponded to the same manually obtained subdimension, overall, 18 of the 22 second-level dimensions of the manually obtained multilevel multidimensional model were confirmed the topic model. Thus, we can state an agreement of 82 % between the second level dimensions of the manually obtained multilevel multidimensional model.

The four dimensions that were found only with the manual approach but not in the topic model were “impaired physical functioning”, “objectifying one’s own body”, “isolation”, and “cognitive impairment”. In Figure 1, those second-level dimensions that were not confirmed by the machine learning approach and the respective third-level dimensions are bracketed. These dimensions are not part of our final model of pain-related suffering.
Discussion

The overall aim of this study was to explore how pain-related suffering is defined in the existing literature and to identify the different dimensions used. It is therefore important to note that the aim of this study was not to develop an "optimal definition". Rather, it was to summarize how pain-related suffering is defined in the scientific and clinical literature and the dimensions that are distinguished. Based on a systematic review of conceptualizations of pain-related suffering across different disciplines, we integrated existing definitions of this phenomenon and derived a multilevel and multidimensional model from the literature. The complete integrative definition includes a total of 25 elements. Most frequently, suffering is defined as (1) a psychological experience (2) taking place to varying degrees on a physical, affective, social, and existential level, (3) causing distress and (4) posing a threat to the affected person. Our final model of pain-related suffering distinguishes three levels with eight, 18, and 49 dimensions respectively. This model can be directly used for scale construction and at the very least serve as a point of orientation for evaluating future and existing operationalizations.

Research question 1: How is suffering defined in current pain research?

One of the main findings of this study is that there is no consensus on a single definition of suffering in the existing literature on pain-related suffering. In total, we were able to identify 60 different definitions of suffering in the context of pain. Despite this variety of definitions, most literature refers to Eric Cassell (1982), who has been cited by 101 of the 111 articles in our text corpus and defines suffering as “the state of severe distress associated with events that threaten the intactness of the person” (p. 640). However, Cassell’s definition has also been criticized for being overly exclusive (Tate, 2020) and for leaving open what exactly a person is and what it means that its intactness can be threatened (Bueno-Gómez, 2017). What is more, he does not specify what exactly the experience of suffering includes and on what dimensions it takes place. Our integrative definition accommodates for this problem by specifying precisely on which dimensions suffering takes place. Nevertheless, because it is representative of huge parts of the literature, our definition also includes Cassell’s description of suffering as resulting from a threat to personal intactness (“integrity as a self”), despite above-mentioned criticisms.

Recent psychological and neuroscientific research defines suffering as “an unpleasant experience associated with negative cognitive, emotional, and autonomic response to a stimulus” (De Ridder et al., 2021). However, many authors discuss suffering in a much wider context than just emotion and cognition, referring to sociolinguistic (Charmaz, 1983; Strong, 1999), existential (Best et al., 2015; Saunders, 1996), and even spiritual aspects (Eriksson, 2006). Our integrative definition points out the multidimensional character of suffering, explicitly mentioning social, existential, and spiritual aspects. Importantly, this does not contradict but rather amends the view that suffering is a psychological and autonomic response to a stimulus.

The multidimensional character of suffering has been pointed out by numerous authors in the examined text corpus (del Giglio, 2020; Montoya-Juarez et al., 2013; Smith et al., 2014). Nevertheless, the comparison of our integrative definition to individual definitions from the literature has shown that existing definitions often either leave out important dimensions or refrain from formulating the concrete dimensions of suffering altogether.

Importantly, the comparison of our definition to the definitions provided by ChatGPT and GPT-3 large language models based on the same textual basis showed a very high semantic similarity (with a vector similarity metric of 0.963 and 0.941 respectively). Obtaining such a
similar synthesis of the current literature with a completely different methodological approach cross-validates our findings and support our conclusion about the importance of conceptualizing the multidimensional character of pain-related suffering in its full width.

**Research question 2: Summarizing the literature into a multilevel and multidimensional model**

It has been stressed repeatedly that suffering is a wholistic experience influencing all aspects of the affected person’s life (Cassell, 1982; Kahn & Steeves, 1986; Krikorian & Limonero, 2012; van Hooft, 1998). Nevertheless, our model shows that the existential as well as the social dimension are of particular importance. Thirty-seven articles discussed the existential dimension of pain related suffering. The important role of this aspect may be related to the fact that many articles discussed pain in the context of life-threatening illness (15 articles focused exclusively on secondary pain). Best et al. (2015) even suggest using the terms “existential distress” and “existential pain” as synonyms for suffering. Thirty-six articles discussed the social dimension of pain related suffering. Recently, Sullivan et al. (2023) have argued that social factors should be seen not only ‘as modifiers of biological causes […] but as equal contributors to pain’ (p. 1). In accordance with this, our model represents the importance of the social and the existential dimension but at the same time does justice to the wholistic character of suffering by bringing the various dimensions of suffering together into one model.

We suggest that the 49 third-level dimensions of the model can be used directly to guide scale construction for the operationalization of pain-related suffering. There are existing measures from related contexts that can be used to quantify the concepts we extracted from the literature. For instance, “Exhaustion” could be measured using a subscale of the Maslach Burnout Inventory (Maslach, Jackson, & Leiter, 1997), and “Unfinished business” could be measured using the Unfinished Business Questionnaire (Masterson et al., 2018). Similarly, a straightforward operationalization should be feasible for most of the 49 third-level dimensions.

We also suggest using the first-level dimensions of our model to determine, if in a particular research context, pain-related suffering should be measured with all its dimensions or if a focus on specific dimensions is more appropriate. Although research suggests a close relationship between the different dimensions (Cassell, 1982; Krikorian & Limonero, 2012), the development of a more nuanced terminology may be helpful to avoid equivocation between them. For instance, it is possible that in cancer research there is a particular interest in existential pain-related suffering, while in psychotherapy research affective or cognitive pain-related suffering are of greater relevance. Conclusions about one type of suffering may not apply to another. Our model can not only guide the operationalization of pain-related suffering in general, but also help adjust the measuring of suffering to the respective need.

More research is needed to examine the factor structure of our proposed model. The third-level dimensions were extracted systematically from the examined literature and then summarized based on theoretical considerations. Although we cross-validated the result and identified very similar dimensions by analyzing our text corpus with an unsupervised language model, the factorization of the third-level dimensions needs to be examined empirically. The same holds for the allocation of the second-level dimensions to the first-level dimensions. Nevertheless, our model may serve as a point of orientation, that can be revised based on empirical research on the proposed factor structure.
Methodological considerations

In this systematic review we used conventional qualitative methods for concept development and cross-validated them using two different machine learning approaches. Comparing our integrative definition with the definition provided by ChatGPT and GPT-3 large language models based on the same textual basis, we found a very high semantic similarity between the results from the manual and the machine learning approaches (with a vector similarity metric of 0.963 and 0.941 respectively) and a relatively low lexical similarity (ROUGE-1 metrics between 0.328 and 0.474). The difference between the similarity measures could be explained by the fact that, although we used a standardized procedure, theoretical knowledge influenced the exact choice of terms for the integrative definition. For instance, in the integrative definition we chose the term “affective” over the term “emotional” (which is used in the definitions provided by the ChatGPT and GPT-3 large language models) because it is the more general concept in psychological research (Niven, 2013). The fact that such semantic details may still be missed by large language models also illustrates the strength of combining manual and machine learning based methods in concept development. Our combined approach allows to pay attention to conceptual detail while at the same time avoiding the risk of subjective judgment as much as possible.

Our multilevel multidimensional model of pain-related suffering was derived from the literature, based on manual extraction and summarization. While the first and third level of this model correspond directly to aspects of suffering frequently mentioned within our text corpus, the second level of the model is based on theoretical considerations and was therefore cross-validated using topic modeling with the LDA algorithm (Blei et al., 2003). The vast congruence (82%) between the second-level dimensions of the model and the topics found by the algorithm shows that topic modeling can be a useful tool for theory development. Especially for reviewing conceptual research, this can significantly reduce subjectivity. At the same time, we found that the topic model also identifies topics unrelated to the concept of interest. This cannot not be prevented because – although pain-related suffering was the shared theme of all articles in the text corpus – it is likely that there are topics shared by a subgroup of articles that are unrelated to suffering but are nevertheless detected by the algorithm. Accordingly, we suggest that the cross-validation of conventional qualitative methods and machine learning is the most promising account for systematically reviewing conceptual data.

Limitations

There are some limitations of our study that need to be considered. Our search strategy was to look for articles that used the terms “pain” and “suffering” in title or abstract, as well as at least one of eight terms indicating that the article had in part a theoretical focus (see Appendix A). On the one hand, the list of terms we used is not conclusive and it is possible that there are other terms with which we could have detected additional articles. On the other hand, it is possible that there are articles that did not use any of the above-described terms, for instance articles with a strong empirical focus using a minimal, but still relevant amount of theory. However, it can be expected that our cross-reference search would have detected any overlooked conceptualizations also in purely empirical articles. Regarding the search terms itself, it can be said that in probatory searches no additional term resulted in a surplus of eligible articles.

Another primary limitation is a risk of bias because only one researcher conducted the extraction of verbatim quotations from the text corpus. Possibly, the perspective of the review
was narrowed by this, insofar as the quotations may have been selected based on a subjective pre-defined notion of suffering stemming from specific articles or authors. However, at least for the extraction of definitions, it can be argued that their identification within an article is a very unambiguous task, insofar as definitions are usually emphatically expressed in sentences such as “suffering is…”, or “suffering can be seen as”. Furthermore, the fact that we successfully cross-validated all our results, i.e., both our exploration of the dimensions and of the definitions of pain-related suffering yielded very similar results, independently of the method used (machine learning or conventional qualitative methods), suggests that the bias in selecting relevant quotations was minimal.

**Conclusion**

There is currently no consensus on a definition of pain-related suffering. Important aspects and dimensions of suffering that are stressed in some parts of the literature, are ignored in others. Our integrative definition brings all important contributions together in a unified definition that does justice to the multifaceted character of the experience of suffering. Furthermore, our multilevel and multi-dimensional model indicates concretely, on which dimensions suffering is experienced and lays the ground not only for an operationalization of suffering as a whole, but also for operationalizing specific aspects of suffering depending on the context.
References


Tate, T. (2020). What we talk about when we talk about pediatric suffering. Theoretical medicine and bioethics, 41(4), 143-163. doi:10.1007/s11017-020-09535-8


**Table legends**

**Table 1: Study characteristics**

The table lists all articles from our text corpus specifying general study characteristics, such as country and discipline, but also the central outcome – the respective definitions of suffering. For readability reasons, only the most important part of each definition is shown in this table, to allow the reader to gain a quick overview.
Table 2: Integrative definition of suffering

The table shows an integrative definition of pain-related suffering based on the examined literature. It includes only those aspects of suffering that were mentioned by at least two articles (core elements).

Table 3: Similarity metrics comparing our definition to those from ChatGPT/GPT-3

The table displays the similarity values obtained from comparing the definitions generated by Chat-GPT and GPT-3 to the integrative definition obtained with conventional qualitative methods.

Table 4: The Topic Model

Each row shows the ten top words found by the search algorithm for each topic, and – in the last column – the label(s) given to these top words by N.N.S. and J.T. The respective top words that were considered indicative of the given label are written in bold and italics, or underlined, in those cases where two interpretations seemed possible.
Figure legends

Figure 1: The multilevel multidimensional model

The model visualizes the multilevel multidimensional model of pain-related suffering. The eight segments of the innermost circle symbolize the first-level dimensions of pain-related suffering according to our model, the second-level dimensions are shown in the medial circle, and the third-level dimensions in the outermost circle. The size of the segments is determined by the number of articles explicitly mentioning the third-level dimensions (or synonyms). The second-level dimensions that could not be confirmed by our machine learning approach are and the corresponding third-level dimensions are bracketed. These dimensions are not part of our final model of pain-related suffering.

Figure 2: The disciplines within the text corpus and their frequency

The figure shows which disciplines the articles in the text corpus come from. The outer circle shows the exact disciplines, the inner circle summarizes them into discipline clusters. The size of the segments indicates the number of articles from each discipline (clustered). The absolute (relative) number of articles is also written behind each discipline label in the figure.

Figure 3: Frequencies of the defining attributes of suffering in our text corpus

The figure illustrates how many articles mention the attributes from the different categories. The categories are listed outside the biggest circle. The number belonging to each circle indicates the number of articles symbolized by this circle.