Automated Identification of Yellow Flags and Their Signal Terms in Physiotherapeutic Consultation Transcripts

Joep Wegstapel[1], Thymen den Hartog[1], Mick Sneekes[1], Bart Staal[2][a], Ellis van der Scheer-Horst[2][b], Sandra van Dulmen[3][c] and Sjaak Brinkkemper[1][d]

[1]*Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands*

[2]*Department of Physiotherapy, HAN University of Applied Sciences, Nijmegen, The Netherlands*

[3]*Nivel (Netherlands institute for health services research), Utrecht, The Netherlands*

Keywords: Physiotherapy, Yellow Flags, Signal Terms, Automated Medical Reporting, Automated Text Identification.

Abstract: This paper investigates the possibility of automating the process of identifying yellow flags and their signal terms in physiotherapeutic consultation transcripts from patients with low back pain, using Automated Text Identification. It is part of the Automated Medical Reporting research domain. In physiotherapy focused on low back pain, yellow flags are considered psycho-social predictors of poor recovery and risk factors for chronic disability development. This paper uses a 6-step mixed method approach. Consultation transcripts and yellow flag assessment guidelines were collected, an automated identification tool was built and the OSPRO assessment guideline was used to test the tool for accuracy. It was found that it is possible to identify Yellow Flags and their Signal Terms automatically with the tool developed in this experiment. However, this is just the beginning, and much more research must be done in the future to further enhance the tool, mainly to improve precision.

1 INTRODUCTION

In healthcare in the broader sense, much time is spent on administrative tasks, of which identifying yellow flags is an example (Stearns et al., 2021). This requires a lot of budget and takes the (para)medic's attention away from providing a patient with care (Bukman, 2019; Maas et al., 2020). To solve this issue to a big extent, research is spent on developing the discipline of Automated Medical Reporting (AMR) (Maas et al., 2020). This paper concerns Automated Medical Reporting in the field of physiotherapy, where the aim is to reduce the administrative burden by automatically turning a consultation recording into a medical report. Many people suffer from low back pain (LBP) (Hoy et al., 2010). In physiotherapy focused on LBP, yellow flags are used to denote psycho-social predictors of chronicity or poor recovery after physiotherapeutic treatment (Moffett and McLean, 2006; Barron et al., 2007); an example could be a patient stating that he experiences pain when working, making him avoid this work, or a patient complaining about low back pain while suffering from depressive feelings at the same time. These yellow flags and their corresponding signal terms are usually identified at the start of treatment to anticipate the potential risks and address them adequately from the start of the treatment. This paper investigates the feasibility of identifying yellow flags and their signal terms automatically, using Automated Text Identification. Based on this, the main research question this paper answers is formulated as: **"How can the process of identifying yellow flags and their signal terms in physiotherapeutic consultation transcripts be automated?"**

First, more information on the context and background of this research is given, then the 6 step research method is discussed, the data analysis is described based on the same 6 steps, and finally the discussion is presented, leading to the conclusion at the end of the paper.

[a] https://orcid.org/0000-0002-0083-6380
[b] https://orcid.org/0000-0001-5494-2050
[c] https://orcid.org/0000-0002-1651-7544
[d] https://orcid.org/0000-0002-2977-8911

Wegstapel, J., den Hartog, T., Sneekes, M., Staal, B., van der Scheer-Horst, E., van Dulmen, S. and Brinkkemper, S.
Automated Identification of Yellow Flags and Their Signal Terms in Physiotherapeutic Consultation Transcripts.
DOI: 10.5220/0011793800003414
In *Proceedings of the 16th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2023) - Volume 5: HEALTHINF*, pages 530-537
ISBN: 978-989-758-631-6; ISSN: 2184-4305

2 THEORETICAL SETTING & EXPLANATION

Currently, a physiotherapeutic consultation can be recorded and transcribed, but the search for Yellow Flags and Signal Words is not yet automated (Parker, 2007; Stewart et al., 2011). If this identification process could be automated, it would drastically improve efficiency by decreasing the administrative burden, and lead to more accurate treatment designs (Watson, 1999).

2.1 Background

To properly understand the important concepts of yellow flags and signal terms discussed in this paper, extensive literature research has been done. A collection of the most relevant information is discussed below. First, short definitions of the concepts are given and after that these concepts are discussed in more detail.

Definitions. Yellow Flags are "psycho-social factors associated with the risk of the development of chronicity in patients/employees that may also predict a poor recovery after treatment" (Moffett and McLean, 2006; Barron et al., 2007; Main, 2013).

A Signal Terms is "a word or phrase that gives an idea about what we might expect to come next" (Demetros, 2021). This is a more general definition, but in the context of this research signal terms are focused on signal terms signaling potential yellow flags.

Yellow Flags. Psycho-social flags in healthcare, referred to as yellow flags, look at factors that identify risk of chronic disability development, as well as potential incomplete recovery from the current condition different to the expectation (Kendall, 1997; Moffett and McLean, 2006; Barron et al., 2007). Yellow flags in physiotherapy are an example of a psychosocial flag, and they can be divided into three natures (Leerar et al., 2007):

- Beliefs, appraisals, judgements, and negative beliefs about pain.
- Emotional responses, distress for meeting criteria of mental disorder diagnosis, worries, fears, anxiety in general.
- Pain behavior (including coping strategies), avoidance of activities due to expectations of pain and possible re-injury, over-reliance on passive treatments, catastrophizing pain sensations.

This paper focuses on yellow flags in all these different natures.

Signal Terms. When a certain pattern of words can be found in similar positions in a text, these could be signal terms (Carballo-Costa et al., 2022). In 2011, Milojevic et al stated that "The higher the occurrence of signal words in the included articles, the more relevant it is to the topic to which it refers" (Milojević et al., 2011). This could help with finding signal terms for a specific context. In case of physiotherapeutic consults, where the definition of signal terms has not yet been clearly defined, scanning texts for patterns of words can help with the creation of a list of signal terms for this specific context. An example would be a patient catastrophizing the experienced pain by adding strong adjectives to the pain statement.

2.2 Yellow Flag Assessment Guidelines

In this research, 4 yellow flag assessment guidelines have been taken into account to assess yellow flags and their signal terms. These are the OSPRO, MPQ, PCS and HADS assessment tools. The yellow flag queries corresponding to these assessment guidelines can be found in the Appendix.

The OSPRO (Optimal Screening for Prediction of Referral and Outcome) is meant as a concise, multidimensional yellow flag assessment guideline for application in orthopedic physiotherapy. It assesses 11 psychological constructs measuring pain-associated psychological distress, based on validated psychological questionnaires (Lentz et al., 2016).

Before the development of the MPQ (McGill Pain Questionnaire), pain was mainly described and measured in terms of intensity, whereas MPQ also investigates the qualitative aspect of pain, like pain in terms of time, space, temperature, tenseness, fear and the general subjective experiencing of pain. MPQ can be used for standard registration and evaluation of pain as well as for diagnosing and monitoring the effects of therapy (Oerlemans et al., 1999; Van der Kloot and Vertommen, 1989).

PCS (Pain Catastrophizing Scale) focuses on the catastrophizing of pain. People over exaggerate their pain experience. With PCS, researchers constructed a scale that incorporated also the non redundant dimensions of pain catastrophizing, including the tendency to increase attentional focus on pain-related thoughts, making people exaggerate (Sullivan et al., 2009).

To understand the experience of suffering in the setting of medical practice, the contribution of mood disorders like anxiety and depression must also be assessed. The HADS (Hospital Anxiety and Despression Scale) was designed to provide a simple yet reliable tool for use in medical practice. It is useful for initial diagnosis and to track progression (or resolution) of psychological symptoms (Snaith, 2003; Stern, 2014).

2.3 Automated Identification

Text mining, also known as text data mining or knowledge discovery from textual databases, refers to the process of extracting interesting and non-trivial patterns or knowledge from text documents (Tan et al., 1999). This article presents a text mining framework consisting of two components: Text refining, which transforms unstructured text documents into an intermediate form, and knowledge distillation, which deduces patterns or knowledge from this intermediate form. To apply automated identification to a text, these two components need to be executed consecutively, and then the knowledge distillation phase must be able to identify the yellow flags and their signal terms present in the transcripts (Tan et al., 1999). There are other cases in which text-mining is used in the healthcare domain.

2.3.1 State-of-the-Art

In 2008, Raya et al. (Raja et al., 2008) showed that text mining can be an effective tool in healthcare contexts. The shift to electronic clinical records opened a huge opportunity window for the text mining research area. The 2014 research by Pendyala et al. gave further insights into the use of text mining techniques for automating medical diagnosis (Pendyala et al., 2014). This led to an overall increased use of text and data mining in healthcare, strengthened by the information richness of the healthcare sector making the use of these techniques indispensable (Tăranu, 2016). This fact is supported by a 2019 review written by Luque et al. of over 90 research papers on text mining in healthcare. They stated that Text mining can be used in healthcare and is especially helpful in improving early disease diagnosis. It can help in developing novel and improved therapies that reduce risk and derived problems and for producing new medical hypothesis (Luque et al., 2019). More recently, van Dijk et al. proved that text-mining in electronic healthcare records can be used as an efficient tool for screening and data collection in cardiovascular trials (van Dijk et al., 2021).

3 RESEARCH METHOD

The main goal of this research is to test if it is possible to automatically identify yellow flags and signal terms in physiotherapeutic consultation transcripts. Based on this goal, the main research question introduced in the introduction was formulated. To answer this question, 2 sub-questions were formulated, being: *Which of the available assessment tools produces the most results in the automated identification of yellow flags and their signal terms in physiotherapeutic consultation transcripts?* and *How can the process of identifying Yellow Flags and their Signal Terms in physiotherapeutic consultation transcripts be automated?*

To answer the research questions, a mixed-method approach was used; interviews were held and an experiment was executed. The following 6 steps were followed.

(1) Collecting transcripts. First, 2 physiotherapists and the HAN Research Group were interviewed. The latter provided 50 real physiotherapeutic consultation transcripts concerning people suffering low back pain. They have been made anonymous to deal with any privacy infringements.

(2) Collecting yellow flag assessment guidelines. Literature research was done concerning yellow flag assessment guidelines, which led to the collection of four yellow flag assessment guidelines discussed in section 2.2. The yellow flag queries corresponding to these guidelines can be found in the Appendix.

(3) Building automated search tool for yellow flag queries. A tool was built to automatically identify specific yellow flag queries in a text. The queries and the text to be searched can be separately provided to the tool as input, allowing for multiple texts to be searched with multiple search queries.

(4) Comparing yellow flag assessment guidelines. The tool was used to search 10 randomly chosen transcripts with the yellow flag queries from the yellow flag assessment guidelines. These yellow flag queries can be found in the tables in the appendix. Statistics of each of these assessment guidelines were generated; they are summarized in table 2. Based on these statistics it was decided to continue with OSPRO (see step 4 in section 4 for more information on this decision).

(5) Filtering results of OSPRO guideline. Before determining the accuracy of the guideline, the results were filtered; how this was done is described in step 5 of section 4.

(6) Determining accuracy based on ground truth. Finally, the 10 transcripts marked based on OSPRO-YF were marked by hand (see figure 1). The hand-marked document, referred to as ground truth, was compared to the automatically marked document, which resulted in the True Positives (TP), False Positives (FP) and False Negatives (FN) summarized in table 3, leading to the recall and precision scores presented in the same table. These results are discussed in section 4.

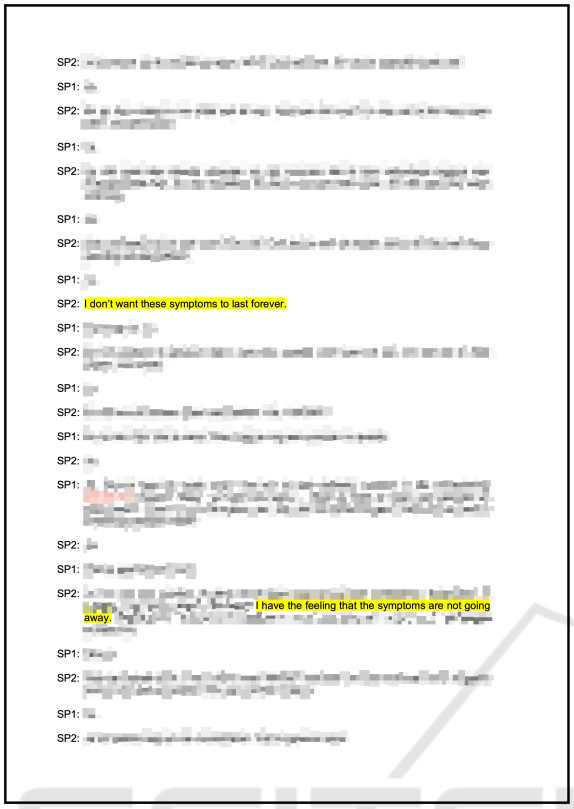

Figure 1: Excerpt of hand-marked consultation transcript with yellow flags and signal terms marked in yellow and the remaining text blurred.

4 DATA ANALYSIS

This section is based on the 6 steps described in the previous section (section 3).

(1) Collecting transcripts. The 50 consultation transcripts have an average of 4910 words, divided over 15 pages. The corresponding recordings are on average 32 minutes and 13 seconds long. Interesting to note is that the duration variation of the consultations is large, with the shortest consultation transcribed being 7 minutes long, whereas the longest consultation had a duration of 1 hour and 21 minutes. The shortest consultation turned out the be an outlier. Since the average consultation duration is 32 minutes, it can be said that physiotherapists take their time when assessing a patient for low back pain. Looking at the number of words per minute, the variation between consultations is smaller; on average a physiotherapeutic low back pain consultation contains 159 words per minute, with a standard deviation of 26 words. In 10 of the 50 consultations, a third speaker was present. Twice these third speakers were physiotherapist supervisors; the other 8 were patient relatives. Speaker turns by these persons are treated as patient information, as they often contain valuable information. See table 1 for more transcript metadata.

Table 1: Metadata Transcripts.

	Length H:M:S	Pages	Words	Words per minute
AVG	00:32:13	15	4910	159,1
Min	00:07:06	6	1402	89,8
Max	01:20:53	36	12746	205,0
St.Dev	00:19:54	8	2678	26,4

(2) Collecting yellow flag assessment guidelines. Here the yellow flag assessment guidelines were collected. In step 4, these guidelines are compared to each other. The OSPRO-YF and HADS were in English, so in this step they were translated to Dutch to match the transcripts.

(3) Building automated search tool for yellow flag queries. In this step the automated identification tool was built, to automatically identify yellow flags and signal terms in the physiotherapeutic consultation transcripts. The tool, created in Java 17, takes two input files: the consultation transcript and the search terms (yellow flag query) that the tool searches for within the transcript. Then the tool calculates a score for every speaker turn in the transcript, based on the sum of the frequencies of all the search terms found in the speaker turn. A high score indicates a high probability of a speaker turn containing a yellow flag. The output is presented in a .txt file, a .json file or in the console. People interested in the source code of this tool can contact us by email.

(4) Comparing yellow flag assessment guidelines. As described in section 3, the OSPRO, MPQ-DLV, PCS and HADS assessment guidelines were compared to each other. They were compared based on 3 different factors: the number of turns with potential yellow flags, the number of found words which are potential yellow flags and the average number of words per marked turn. The corresponding statistics are presented in Table 2, 'Assessment guideline comparison' below.

Table 2: Assessment guideline statistics.

	MPQ	HADS	PCS	OSPRO-YF
Number of turns with potential yellow flag	62	130	1315	2287
Number of found words which are potential yellow flags	74	150	3232	14702
Avg. number of words per marked turn	1.19	1.15	2.46	6.43

The results presented above show that the OSPRO-YF assessment guideline returned the most potential yellow flags and signal terms, in absolute

numbers as well as on average. Since it is important to find as many relevant yellow flags as possible using the automated marking tool, and since OSPRO is the only guideline with queries consisting of full sentences, which is most similar to the way yellow flags are generally described, it was decided to choose this assessment guideline over the other 3, and use it to test the tool for accuracy.

(5) Filtering results based on OSPRO-YF guideline. Since the OSPRO-YF tool returned a lot of potential yellow flags and signal terms with low scores, these results were not taken into account in this research. After translating to Dutch, there was a total of 155 different search terms divided over 27 unique word combinations. This gives an average of 5.74 words per unique combination with a standard deviation of 3.98. The choice to filter out any score lower than 5.0 was based on the average number of words per search term and the relatively large standard deviation. It was decided not to round up to a score of 6.0, to not exclude even more potential yellow flags. Rounding off was necessary because the tool scores each speaker turn as a score without decimals.

(6) Determining accuracy based on ground truth. The 10 randomly selected transcripts had already been searched for the OSPRO yellow flag query. The same transcripts were also marked by hand (ground truth) to more critically analyse the accuracy of the returned results. The tool was able to mark 117 true positives, meaning the automatically marked flag was also marked as a Yellow Flag by hand. 782 of the total of 944 marked words turned out to be false positives, meaning the yellow flag marked by the tool was not marked by hand, and there were 45 yellow flags that the tool missed (false negatives). This led to an overall recall for this sample of 72.22% with a precision rate of 13.01%. The precision rate tells us that most of the marked words turned out to not be real yellow flags. However, the recall of 72% showed that, using the automated identification tool, almost 3 quarters of the relevant instances were actually retrieved, which is actually pretty good. All together, it can be said that the tool already does well on retrieving real yellow flags and signal terms, but it currently retrieves too many irrelevant words as well. An overview of all data per transcript can be found in Table 2, 'Recall and precision data based on OSPRO-YF', below.

Table 3: Recall and precision data based on OSPRO-YF.

Consult. file	TP	FP	FN	Recall	Precision
Cons. file 1	24	114	8	75.00%	17.39%
Cons. file 2	12	29	8	60.00%	29.27%
Cons. file 3	21	76	4	84.00%	21.65%
Cons. file 4	9	52	1	90.00%	14.75%
Cons. file 5	14	58	4	77.78%	19.44%
Cons. file 6	5	145	6	45.45%	3.33%
Cons. file 7	13	81	3	81.25%	13.83%
Cons. file 8	10	97	6	62.50%	9.35%
Cons. file 9	2	73	3	40.00%	2.67%
Cons. file 10	7	57	2	77.78%	10.94%
Total	117	782	45	72.22%	13.01%

5 DISCUSSION

The results from the previous section tell us more about how feasible it is to automatically identify yellow flags and their signal terms in physiotherapeutic consultation transcripts.

First of all it is important to mention that in this research, the transcripts marked by hand are assumed to be 100% correct. It is very likely that this does not reflect reality.

It is also not to be forgotten that the fact that OSPRO-YF returns the most *potential* yellow flags, does not mean that it returns the most *relevant* yellow flags. It is possible that the other 3 assessment guidelines, the MPQ, PCS and HADS, even though they return less *potential* yellow flags, in fact return more *relevant* yellow flags than the OSPRO-YF guideline does. This should be further investigated in future research.

Furthermore, it was decided to only focus on the speaker turns in which the patient was speaking. After analyzing the transcripts, it was found that many yellow flags and signal terms are actually pronounced by the physiotherapist, after which the patient only agrees or does not agree. To not miss these yellow flags, it was decided to take into account speaker turns of all involved speakers.

The transcripts contain natural language which is currently interpreted by a trained physiotherapist. Since natural language is ambiguous, it is hard for the tool to identify all yellow flags and signal terms, as they can be formulated in many different ways. This makes the use of Natural Language Processing (NLP) crucial for correctly automating the identification process (Dalpiaz et al., 2019). The tool must also be extended with synonyms to make sure no information is missed.

The two professionals interviewed were physiotherapists working for a private practise, who mentioned that in their experience, the administrative burden of adding yellow flags to a patients Electronic Health Record (EHR) is nonexistent. When a yellow flag is noticed, they act on it immediately by altering the treatment. This alteration of treatment might end up in the EHR, but mentioning the yellow flag as a reason of alteration is not necessary according to them. Since 'n' is only 2 in this case, it cannot be generalized across all physiotherapists. Further research

must be done to investigate if the claims made here are more broadly applicable.

6 CONCLUSION & FUTURE WORK

This paper seeks to answer the research question: "How can the process of identifying yellow flags and their signal terms in physiotherapeutic consultation transcripts be automated?" The answer to this question is that it is possible to automatically mark yellow flags and their signal terms in Physio Therapeutic consultation transcripts by using an automated identification tool. Even though the tool constructed for this paper was only able to mark yellow flags and signal terms with a total precision of 13%, it reaches a recall score of 72%. It can be said that the tool already does well on retrieving real yellow flags and signal terms, but it currently also retrieves too many irrelevant ones. Due to time constraints the scope of this research was narrow; more research needs to be conducted. Since natural language is very ambiguous, more research must be conducted to embed Natural Language Processing (NLP) in the tool, to up the recall and the precision scores. Also, more training data like examples of yellow flags and signal terms are needed, and a standardized format must be established, stating what components yellow flags consist of. The OSPRO, MPQ, PCS and HADS assessment guidelines might be able to contribute to achieve this.
To conclude, it is safe to assume that automated identification of yellow flags and signal terms is possible using the tool described in this research. However, this is just the beginning, and much more research must be done in the future to further enhance the tool, aiming to improve the accuracy metrics. The recall score of 72% can be improved, but the focus must be on improving the precision score of 13%.

ACKNOWLEDGEMENTS

We want to thank Psychology researcher Wim van Lankveld from the HAN University of Applied Sciences for the interview, guidance, and providing us with the transcripts and other relevant information.

Ethical approval for the data collection of this study was given by the Ethical Research Committee of the HAN University of Applied Sciences in Nijmegen, the Netherlands (EACO 145.04/19).

REFERENCES

Barron, C. J., Klaber Moffett, J. A., and Potter, M. (2007). Patient expectations of physiotherapy: definitions, concepts, and theories. *Physiotherapy theory and practice*, 23(1):37–46.

Bukman, B. (2019). 'zorgverleners besteden 40 procent van hun tijd aan administratie'.

Carballo-Costa, L., Michaleff, Z. A., Costas, R., Quintela-del Río, A., Vivas-Costa, J., and Moseley, A. M. (2022). Evolution of the thematic structure and main producers of physical therapy interventions research: A bibliometric analysis (1986 to 2017). *Brazilian Journal of Physical Therapy*, 26(4):100429.

Dalpiaz, F., Van Der Schalk, I., Brinkkemper, S., Aydemir, F. B., and Lucassen, G. (2019). Detecting terminological ambiguity in user stories: Tool and experimentation. *Information and Software Technology*, 110:3–16.

Demetros, C. (2021). Signal words: 5 fun ways to explain these sentence superheroes!

Hoy, D., Brooks, P., Blyth, F., and Buchbinder, R. (2010). The epidemiology of low back pain. *Best practice & research Clinical rheumatology*, 24(6):769–781.

Kendall, N. (1997). Guide to assessing psychosocial yellow flags in acute low back pain. *Risk factors for long-term disability and work loss*.

Leerar, P. J., Boissonnault, W., Domholdt, E., and Roddey, T. (2007). Documentation of red flags by physical therapists for patients with low back pain. *Journal of Manual & Manipulative Therapy*, 15(1):42–49.

Lentz, T. A., Beneciuk, J. M., Bialosky, J. E., Zeppieri Jr, G., Dai, Y., Wu, S. S., and George, S. Z. (2016). Development of a yellow flag assessment tool for orthopaedic physical therapists: results from the optimal screening for prediction of referral and outcome (ospro) cohort. *journal of orthopaedic & sports physical therapy*, 46(5):327–343.

Luque, C., Luna, J. M., Luque, M., and Ventura, S. (2019). An advanced review on text mining in medicine. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(3):e1302.

Maas, L., Geurtsen, M., Nouwt, F., Schouten, S. F., Van De Water, R., Van Dulmen, S., Dalpiaz, F., Van Deemter, K., and Brinkkemper, S. (2020). The care2report system: automated medical reporting as an integrated solution to reduce administrative burden in healthcare. In *Information Technology in Healthcare: IT Architectures and Implementations in Healthcare Environments*. Hawaii International Conference on System Sciences (HICSS).

Main, C. J. (2013). Yellow flags.

Milojević, S., Sugimoto, C. R., Yan, E., and Ding, Y. (2011). The cognitive structure of library and information science: Analysis of article title words. *Journal of the American Society for Information Science and Technology*, 62(10):1933–1953.

Moffett, J. and McLean, S. (2006). The role of physiotherapy in the management of non-specific back pain and neck pain. *Rheumatology*, 45(4):371–378.

Oerlemans, H. M., Oostendorp, R. A., de Boo, T., and Goris, R. J. A. (1999). Pain and reduced mobility in complex regional pain syndrome i: outcome of a prospective randomised controlled clinical trial of adjuvant physical therapy versus occupational therapy. *Pain*, 83(1):77–83.

Parker, R. (2007). Physiotherapy students' assessment of psychological yellow flags in low back pain. *South African Journal of Physiotherapy*, 63(1):3.

Pendyala, V. S., Fang, Y., Holliday, J., and Zalzala, A. (2014). A text mining approach to automated healthcare for the masses. In *IEEE Global Humanitarian Technology Conference (GHTC 2014)*, pages 28–35. IEEE.

Raja, U., Mitchell, T., Day, T., and Hardin, J. M. (2008). Text mining in healthcare. applications and opportunities. *J Healthc Inf Manag*, 22(3):52–6.

Snaith, R. P. (2003). The hospital anxiety and depression scale. *Health and quality of life outcomes*, 1(1):1–4.

Stearns, Z. R., Carvalho, M. L., Beneciuk, J. M., and Lentz, T. A. (2021). Screening for yellow flags in orthopaedic physical therapy: a clinical framework. *Journal of Orthopaedic & Sports Physical Therapy*, 51(9):459–469.

Stern, A. F. (2014). The hospital anxiety and depression scale. *Occupational medicine*, 64(5):393–394.

Stewart, J., Kempenaar, L., and Lauchlan, D. (2011). Rethinking yellow flags. *Manual therapy*, 16(2):196–198.

Sullivan, M. J., Bishop, S. R., and Pivik, J. (2009). Pain catastrophizing scale. *The Journal of Pain*.

Tan, A.-H. et al. (1999). Text mining: The state of the art and the challenges. In *Proceedings of the pakdd 1999 workshop on knowledge disocovery from advanced databases*, volume 8, pages 65–70. Citeseer.

Tăranu, I. (2016). Data mining in healthcare: decision making and precision. *Database Systems Journal BOARD*, 33.

Van der Kloot, W. and Vertommen, H. (1989). Mpq-dlv: Een standaard nederlandstalige versie van de mcgill pain questionnaire. *Achtergronden en handleiding*.

van Dijk, W. B., Fiolet, A. T., Schuit, E., Sammani, A., Groenhof, T. K. J., van der Graaf, R., de Vries, M. C., Alings, M., Schaap, J., Asselbergs, F. W., et al. (2021). Text-mining in electronic healthcare records can be used as efficient tool for screening and data collection in cardiovascular trials: a multicenter validation study. *Journal of Clinical Epidemiology*, 132:97–105.

Watson, P. J. (1999). Psychosocial assessment: The emergence of a new fashion, or a new tool in physiotherapy for musculoskeletal pain? *Physiotherapy*, 85(10):530–535.

APPENDIX

Lists of Yellow Flags & Signal Terms

This section contains all yellow flags and their signal terms, retrieved from the HADS (Snaith, 2003; Stern, 2014), PCS (Sullivan et al., 2009), OSPRO-YF (Lentz et al., 2016) and MPQ (Oerlemans et al., 1999; Van der Kloot and Vertommen, 1989) assessment guidelines.

HADS:

Table 4: HADS search terms (Translated from dutch).

Search Terms HADS	
Tense	Enjoyment
Anxious	Laughing
Uncomfortable	Restless
Excited	Relaxed
Difficult	Uneasy
Interested	Restless
Rejoice	Anxiety
Panic	Feel
Sensation	Feeling

PCS:

Table 5: PCS search terms (Translated from dutch).

Search Terms PCS	
Pain	Stopping
Can't go on	Never get better
Overwhelming	Overwhelmed
Not enduring	Pain will get worse
Painful event	Going away
Can't get out of my mind	How much it hurts
Pain ceases	Intensity
Reduce intensity of pain	Serious thing happen

OSPRO-YF:

Table 6: OSPRO-YF search terms (Translated from dutch).

Search Terms OSPRO-YF	
Poor appetite	Overeating
Satisfied	Displeases me
Hot headed	Mad
Reacting irritated	Criticism from others makes me angry
Poor appetite or overeating problems	I am satisfied
Some unimportant thoughts go through my head and bother me	I am a hot headed person
When I get angry, I say nasty tings	It makes me furious when I am criticized in front of others
I can't keep it out of my mind	Physical activity can harm my painful body region
I can't do physical activities that (can) make my pain worse	My work is too hard for me
During painful episodes it's hard for me to think about anything other than the pain	I can lead a full life despite my chronic pain
Before I can make serious plans, I have to get my pain under control	My therapy doesn't care how I feel emotionally

MPQ:

Table 7: MPQ search terms (Translated from dutch).

Search Terms MPQ-DLV	
Throbbing	Thumping
Bursting	Flaring
Flashing	Shooting
Pungent	Stinging
Piercing	Sharp
Cutting	As sharp as a knife
Pressing	Squeezing
Stringing	Pulling
Splitting	Tearing
Burning	Burning
Flaming	Brooding
Glowing	Scorching
Cold	Ice-cold
Freezing	Tingling
Itching	Electric
Stiff	Tight
Cramping	Whining
Persistent	Tiring
Debilitating	Exhausting
Exhausting	Grumpy
Depressing	Sickening
Tense	Oppressing
Vestigating	Disturbing
Frightening	Terrifying
Harassing	Torturous
Mild	Moderate
Very	Enormous
Wearable	Obstructive
Dismaying	Unbearable
Annoying	Miserable
Dreadful	Terrible
Awful	