ISSN 2313-0768

Volume 12 Issue 1, June 2025

MJHS

MALTA JOURNAL OF HEALTH SCIENCES

L-Università ta' Malta
Faculty of Health Sciences

www.um.edu.mt/healthsciences/mjhs

Malta Journal of Health Sciences
https://www.um.edu.mt/healthsciences/mjhs/

CONTENTS

Malta Journal of Health Sciences Volume 12 – Issue 1 (2025)

Malta Journal of Health Sciences – Journal of the Faculty of Health Sciences, University of Malta
https://www.um.edu.mt/healthsciences/mjhs/
mjhs@um.edu.mt

Editorial Board

Aim and scope

The Malta Journal of Health Sciences is a peer–reviewed, open access publication that promotes the sharing and exchange of knowledge in Health Sciences. It provides a platform for novice and established researchers to share their findings, insights and views within an inter–professional context. The Journal originates within the Faculty of Health Sciences, University of Malta.

The Malta Journal of Health Sciences disseminates research on a broad range of allied health disciplines. It publishes original research papers, review articles, short communications, commentaries, letters to the editor and book reviews. The readership of the journal consists of academics, practitioners and trainee health professionals across the disciplines of Applied Biomedical Science, Audiology, Community Nursing, Dietetics, Environmental Health, Food Science, Health Services Management, Medical Physics, Mental Health, Midwifery, Nursing, Nutrition, Occupational Therapy, Physiotherapy, Podiatry, Radiography and Speech Language Pathology.

Submitted manuscripts undergo independent blind peer review, typically by two reviewers with relevant expertise. All manuscripts are reviewed as rapidly as possible and an editorial decision is generally reached within approximately two months of submission. Authors of manuscripts that require revisions will have two weeks to submit their revised manuscripts. No manuscript that has already been published or is under consideration for publication elsewhere will be considered.

Contents

Malta Journal of Health Sciences Volume 12 – Issue 1 (2025)

Malta Journal of Health Sciences
https://doi.org/10.14614/GENETICTEST/12/25

Guest Editorial

Genetic testing for inherited diseases – it is not just about diagnosis

Prof Rosienne Farrugia
Head of Department, Applied Biomedical Science, Faculty of Health Sciences, University of Malta

Just over two decades ago, in October 2004, the first complete sequence of the Human Genome was published (International Human Genome Sequencing Consortium, 2004). It was the culmination of a massive and expensive collaboration between 20 sequencing centres across the world, spanning more than a decade of work and at an estimated cost of over €2.5 billion. This first complete genome build made available the sequence of 92% of the 3 billion nucleotides that make up a human genome.

Fast forward 20 years, and advancements in technology – the development of high throughput sequencing and bioinformatics – have enabled us to fill in the missing 8% (Amaral *et al.*, 2023). More importantly, the advent of high throughput sequencing has made it possible to sequence an entire genome in a few hours and generate a list of genetic variants within a couple of days. All this for less than €1000 per genome. As more and more genomes are sequenced, we have also come to the realisation that there is huge variation between individuals, with any two unrelated people having around 5 million differences in their DNA sequence (The 1000 Genomes Project Consortium, 2015). These differences make us unique. These differences also underpin a vast number of genetic disorders. Unexpectedly this knowledge has ushered in a new era in medicine – the era of precision medicine.

Traditionally, testing for genetic mutations that caused disease was a targeted, sometimes lengthy, process that required knowledge of the mutation to be tested for. This was problematic for rare diseases, since the required knowledge was often not available. Modern diagnostics, including the search for new and unknown novel variants, takes a more holistic approach – sequencing the entire human genome and using bioinformatics to sift through the 5 million variants to get to the causative ones. This makes possible a definitive diagnosis for more patients, even those with the rarest of conditions (Ng *et al.*, 2010).

The principle behind second generation genome sequencing technologies is beautifully simple compared to the massive amount of data that it generates. DNA is obtained from any nucleated cell and broken down into small pieces. The DNA fragments are then ligated to short synthetic nucleotide sequences and bound to a solid support. Each fragment is read starting from the known synthetic nucleotide sequences. Since all fragments are immobilised, millions of them can be sequenced in parallel with fluorescence data collected in real time and stored in individual computer files. The sequence is then put together again, like a jigsaw puzzle, giving the individual's entire, continuous DNA sequence. A bioinformatic comparison to the standard reference sequence is finally used to identify DNA variations within the individual. The majority of the 5 million identified variants have no effect on health. Some may have subtle effects, where a variant influences risk for complex diseases such as diabetes or cardiovascular disease. A few variants will directly cause disease, mostly rare diseases.

'The right drug, at the right dose, at the right time' is the final aim of precision medicine, and knowledge of the exact DNA mutation causing the disease can at times determine treatment – thus personalising

Correspondence to Rosienne Farrugia
rosienne.farrugia@um.edu.mt
Received: 27.2.2025
Revised: 19.5.2025
Accepted: 20.5.2025
Published: 30.6.2025

medicine. In this way, genomic data makes possible diagnostics, prognostics as well as targeted treatment. Cancer therapy is probably one of the clearest and best-known examples where patients with the same tumour may be given different targeted therapy based on the genetic mutation driving their tumour (Shuel, 2022). But the use of precision medicine extends beyond targeted cancer therapy. Locally we have used genetic analysis to determine that the type of phenylketonuria prevalent in the Maltese, is the rarer, atypical forms (Neville *et al.*, 2005) which do not respond well to solely a phenylalanine restricted diet but require additional neurotransmitter supplementation treatment with carbidopa and 5-hydroxytryptophan. Additionally, these conditions are far more prevalent locally than elsewhere (Farrugia *et al.*, 2007) and warrant their own screening programme which has been recently launched. As we continue to study the genetic basis of disease within the Maltese, we and others (Axiak *et al.*, 2023, Ciantar *et al.*, 2024) are finding that there are other conditions locally where genetic tests will not only offer a diagnostic result, but may also offer prognostic information and the possibility of targeted treatment with a concomitant improvement to quality of life, a reduced financial strain on medical services and better outcomes for our patients.

Moving forward, genomic testing will become an even more integral part of pathology, applicable not only to the study of inherited genetic diseases, but also infectious diseases as clearly illustrated by the expansion of molecular COVID-19 tests during the pandemic. Though it brings many benefits, genomics also brings with it several ethical and GDPR considerations. DNA sequences are shared with parents and siblings (and to a lesser extent with members of the extended family), and a genetic finding does not only belong to the individual, but to the family. When analysing a genome, other information not pertinent to the disease being studied may be uncovered. These incidental findings may impact a person's life in multiple ways. Thus, genomic medicine has huge potential, it is changing the way we carry out diagnostic tests and prescribe treatment, however, policies need to be put in place to safeguard both patients and clinicians such that the benefits may continue to outweigh the risks.

References

International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* 431, 931–945 (2004). https://doi.org/10.1038/nature03001

Amaral, P., Carbonell-Sala, S., De La Vega, F.M. *et al.* The status of the human gene catalogue. *Nature* 622, 41–47 (2023). https://doi.org/10.1038/s41586-023-06490-x

The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* 526, 68–74 (2015). https://doi.org/10.1038/nature15393

Ng SB, Nickerson DA, Bamshad MJ, Shendure J. Massively parallel sequencing and rare disease. *Hum Mol Genet*. 15;19(R2):R119–24 (2010). https://doi: 10.1093/hmg/ddq390. Epub 2010 Sep 15

Shuel SL. Targeted cancer therapies: Clinical pearls for primary care. *Can Fam Physician*. 68(7):515–518 (2022). https://doi: 10.46747/cfp.6807515

Neville BG, Parascandalo R, Farrugia R, Felice A. Sepiapterin reductase deficiency: a congenital dopa-responsive motor and cognitive disorder. *Brain*. 128(Pt 10):2291–6 (2005). https://doi: 10.1093/brain/awh603

Farrugia R, Scerri CA, Montalto SA, Parascandolo R, Neville BG, Felice AE. Molecular genetics of tetrahydrobiopterin (BH4) deficiency in the Maltese population. *Mol Genet Metab*. 90(3):277–83 (2007). https://doi: 10.1016/j.ymgme.2006.10.013

Axiak CJ, Pleven A, Attard R, Borg Carbott F, Ebejer JP, Brincat I, Cassar K, Gruppetta M, Vassallo J, Bezzina Wettinger S, Farrugia R. High Population Frequency of *GNRHR* p.Q106R in Malta: An Evaluation of Fertility and Hormone Profiles in Heterozygotes. *J Endocr Soc*. 29;8(2):bvad172 (2023). https://doi: 10.1210/jendso/bvad172.

Ciantar N, Zahra G, Delicata J, Sammut F, Calleja-Agius J, Farrugia E, Said E. Genotype-phenotype of autosomal dominant polycystic kidney disease in Malta. *Eur J Med Genet*. 69:104934 (2024). https://doi: 10.1016/j.ejmg.2024.104934.

Malta Journal of Health Sciences
https://doi.org/10.14614/DANCERDIET/12/25

Research Article

Assessing Dietary Intake in Female Adult Dancers: A Cross-Sectional Study

Ms Heather Galea
Department of Food Sciences and Nutrition, Faculty of Health Sciences, University of Malta
Dr Petra Jones
Department of Food Sciences and Nutrition, Faculty of Health Sciences, University of Malta

Abstract

Optimal dietary intake supports training and performance and minimises risk of developing disordered eating habits for dancers. Nevertheless, dietary intake data is limited globally. This study aims to assess dietary intake, exercise energy expenditure (EEE) and physical activity level of female, adult dancers. Professional and recreational dancers were conveniently recruited from all dance schools in Gozo, a small Mediterranean island. Dietary intake during weekdays and the weekends was assessed through a 4-day food and beverage diary. Data on EEE was collected using a lifestyle questionnaire adapted from the 'International Physical Activity Questionnaire – Short Form'. One-sample t-tests compared participants' intakes to recommended dietary allowances (RDA), while paired-samples t-tests assessed differences in intakes between weekdays and weekends. A total of 14 dancers, 11 recreational and 3 professional, were recruited. The mean energy intake was 1306±348kcal/day with 34.7±5.7% derived from fat. The mean carbohydrate and protein intakes were 2.4±0.8g/kg and 1.1±0.3g/kg body weight respectively. The mean total EEE was 2034±190kcal/day. Daily mean energy, carbohydrate and protein intakes were lower than the RDA, while mean fat intake was higher. There was no significant evidence of dietary intake variations between weekdays and the weekend (p=0.309 carbohydrates, p=0.596 fat, p=0.956 protein). Recreational dancers were likely to consume sufficient dietary intake. Professionals consumed suboptimal energy, carbohydrate and protein intakes. Dietary fat recommendations were met by half of the participants and exceeded by the rest. This study is the first to assess the dietary intake of dancers in Malta, highlighting the need for further research with a larger cohort of local dancers recruiting diverse age groups and training intensities.

Keywords: female dancers, sports nutrition, Malta, Gozo, dietary intake, energy expenditure

Correspondence to Heather Galea
heather.galea.18@um.edu.mt
Received: 22.8.2024
Revised: 26.11.2024
Accepted: 19.12.2024
Published: 30.6.2025

1. Introduction

Aesthetic sports encompasses all athletes engaging within a discipline reinforcing a lean body composition. Dancers represent a primary example of aesthetic athletes due to the great emphasis placed on maintaining svelte figures with the aim of enhancing the aesthetics of performance. A greater lean muscle mass to fat ratio is perceived advantageous since this supports optimal agility, speed, power and strength (Nepocatych *et al.*, 2017). As a result, dietary intake suffers due to nutrition propaganda leading to a cycle of unordinary eating rituals resulting in clinical eating disorders (Struab, 2018).

The dietary intake of various athletic populations has been extensively researched globally. However, there is a paucity of evidence about the nutritional habits of individuals in the dance arts. This lack of data is especially noteworthy considering that ballerinas, particularly females, are consistently reported to be 10–12% below their ideal body weight (BW) (Doyle-Lucas *et al.*, 2010; Sousa *et al.*, 2013) and 7–38% possibly experiencing eating disorders (Dotti *et al.*, 2002). Satisfying nutritional requirements is fundamental for dancers to support training and performance and minimise risks of developing disordered eating habits.

Energy availability (EA) refers to the daily amount of energy available for the body's physiological functions other than physical activity. The average energy availability required for health and weight maintenance is 45kcal/kg fat-free mass (FFM)/day, with 30kcal/kg FFM/day recommended as the lowest EA (LEA) threshold. While dancers frequently strive for LEA levels to achieve and sustain reduced BWs, it is fundamental to recognize that LEAs do not invariably result in low BWs. Disruptions in energy balance (EB), caused by reduced dietary intake and/or elevated exercise energy expenditure (EEE), prompt the body to either mobilize energy reserves from adipose tissue and body proteins or conserve energy by downregulating non-essential physiological processes. These adaptive mechanisms, evolved to prioritize survival, may ultimately stabilize BW despite a prolonged negative EB. Consequently, even in the absence of significant reductions in BW or fat levels, dancers with sustained LEA, may experience physiological impairments due to compromised EA which may lead to the Female Athlete Triad or Relative Energy Deficiency in Sport (REDs) with possible career threatening consequences (Areta *et al.*, 2020; Burke *et al.*, 2018; Mountjoy *et al.*, 2023).

Both professional and recreational dancers were found to suffer from a negative EB (Hajian & Mohaghegh, 2024; Sousa *et al.*, 2013). EEE for female ballerinas when performing barre work is estimated as 0.08kcal/kg/min while 0.10kcal/kg/min are approximately expended during center floor training. Therefore, an hour-long ballet class consumes approximately 200kcal. Modern dance is estimated to expend 0.12kcal/kg/min (Cohen *et al.*, 1985; Koutedakis & Jamurtas, 2004).

Carbohydrates are fundamental to an athlete's diet as the optimal energy substrate. The evidence suggests that prolonged sustained activities, such as dance, are improved through the maintenance of high carbohydrate availability and most moderate intensity exercise is characterised as being carbohydrate dependent (Thomas *et al.*, 2016). As dancers seek to maintain low BWs, the carbohydrate recommendation is set at 3–5g/kg BW/day (Sousa *et al.*, 2013). This range is recommended for low-intensity or skill-based activities and should increase proportionally with intensity levels, with 5–7g/kg BW/day suggested for moderate exercise programmes, 6–10g/kg BW/day for endurance programmes and 8–12g/kg BW/day for extreme exercise (Thomas *et al.*, 2016). Nevertheless, general carbohydrate guidelines need to be adjusted according to the athlete's body size and training characteristics.

Fat is also a vital source of energy with the acceptable macronutrient distribution range set at 20–35% of total energy intake (E%). Recommended intake should be met by consuming essential fatty acids while maintaining saturated fat intake below 10E%. For endurance training, fats are supplied to muscles as plasma free fatty acids, intramuscular triglycerides and adipose tissue. Intakes below 20E% impair performance. Therefore, dancers should avoid restricting fat intake to maintain low BWs or improve body composition. If necessary, such practices should be limited to pre-event diets or carbohydrate-loading (Sousa *et al.*, 2013; Thomas *et al.*, 2016).

Athletes frequently opt for high protein diets to maximise adaptation and recovery following training as the macronutrient provides a substrate for both contractile and metabolic protein synthesis. Protein intakes between 1.2–2.0g/kg BW/day support metabolic adaptation, repair and protein turnover. Such targets should be met evenly throughout the day. Higher intakes of 2.0g/kg BW/day and above may be beneficial during inactivity due to physical injury and energy restriction periods maintaining lower BWs while preventing FFM loss. An individualised approach based on the athlete's activity level, training volume and energy and macronutrient targets should be adopted (Egan, 2016; Thomas *et al.*, 2016).

The principal aim of this study is to assess the actual dietary intake of female, Gozitan, adult dancers, determining whether they meet established recommendations for energy and macronutrient intake.

2. Methods

All dance schools active during the research period in the Maltese island of Gozo were recruited using purposive sampling through email outreach. Heads of schools were provided with information about the study including what their participation would entail; acting as intermediaries by inviting all eligible dancers within their dance school to participate. The inclusion criteria included dancers training any dance genre, students or teachers and professional or recreational dancers.

Questionnaires and 4-day food and beverage diaries (FD), including guidelines for accurately recording detailed FDs with household measures for estimating portion sizes and a template, were distributed to the heads of schools who provided consent. Previous studies using similar methodologies have demonstrated the potential to reveal nutritional deficits in comparable populations (Brown *et al.*, 2017a; Brown *et al.*, 2020; Challis *et al.*, 2020; Civil *et al.*, 2019; Hajian & Mohaghegh, 2024; Mangion, 2010). They were instructed to disseminate a copy of each of the documents to all eligible students and teachers. Complete FDs and questionnaires were to be returned to the researcher by post through the self-addressed envelopes provided to the head of school and were considered as consent, as detailed in the information sheet provided to respondents.

The 4-day FD was recorded between Thursday and Sunday to account for dietary intake variations within weekdays and the weekend (Brown *et al.*, 2017a; Civil *et al.*, 2019) and training and non-training days. A protocol was developed to ensure a standardised approach, particularly when dealing with missing or incomplete information during coding of the FD (Guan *et al.*, 2019; Puš *et al.*, 2012). Dietary intake was analysed using the commercially available dietary analysis software, Nutritics (Nutritics, 2021). Traditional meals reported were assumed to follow recipes from traditional cookbooks, while modern recipes were based on online sources. Missing portion sizes were estimated using standard portion sizes (Patel *et al.*, 2012). Nutrition labels for local products were used to obtain precise nutritional profiles when not available in the dietary analysis software.

The questionnaire aimed to gather general demographic data and enable an estimation of EEE. Physical activity levels (PALs) were estimated based on the 'International Physical Activity Questionnaire–Short Form' (IPAQ-SF). The IPAQ-SF automatic report was used to automatically convert self-reported EEE into kcal/week and enable classification according to the three levels of PA; low, moderate or high (IPAQ, 2005). All data was analysed using the Statistical Package for the Social Sciences (SPSS) Version 27. Ethical clearance was secured for the study from the Faculty of Health Sciences Research Ethics Committee (FREC) and the University Research Ethics Committee (UREC) at the University of Malta (UREC FORM V_15062020 5549).

Descriptive statistics were applied to analyse the nature of the data. The one sample t-test was used to investigate whether the participants' mean energy and macronutrient intakes and BMI significantly differ from the respective guidelines. The paired sample t-test was used to investigate any significant differences in dietary intakes between weekdays and the weekend. The One sample Wilcoxon Signed Ranks test was used to compare mean waist circumference (WC) with recommended range. The Spearman Rank Correlation Coefficient was used to test associations between macronutrient intakes, mean daily EI, TEE and EB, BMI and WC, weekly PAL and full-time job activity and studio locality and EEE. For all statistical tests, an alpha of 0.05 was employed as a significance cut-off point unless otherwise stated.

3. Results

From the Gozitan dance schools initially invited to participate (n=7), approval was granted for six schools. Out of the potential participants eligible for recruitment (n=25), 56% engaged in the study. Table 1 describes baseline characteristics of participants (n=14). Ballet was the genre most practiced (n=13) followed by modern jazz (n=11), tap (n=4) and contemporary (n=3) with some dancers practicing more than one genre. The minority had reached a professional level (n=3) with only one participant reporting dance as her full-time job. The rest were recreational dancers (n=11) who were either still academic students (n=10), practiced dance as a hobby (n=2) or had another full-time occupation (n=1). The majority had a sedentary full-time job (n=9) and traveled to their studio by car (n=12), whilst only 2 participants traveled to their studio on foot (n=2). Participants' mean BMI of 23.0±2.4kg/m2 was within a normal range while the mean WC of 72.1±9.6cm was below the recommended cut-off points (World Health Organization Expert Consultation [WHO], (2011).

Table 1: Descriptive characteristics, dietary intake and estimated energy balance of study participants (n=14)			
	Mean ± S.D	**Range**	**Variance**
Age (*years*)	19.9±2.2	18.0–25.0	4.8
Weight (*kg*)	59.6±6.3	52–70	39.8
Body mass index (*kg/m²*)	23.0±2.4	19.6–27.3	5.8
Waist circumference (*cm*)	72.1±9.6	45.0–84.0	92.0
Dancing years	15.4±2.7	8.0–20.0	7.3
Daily exercise energy expenditure (*kcal*)	296±221	29–779	48887
Daily energy intake (*kcal*)	1306±348	879–2009	121140
Daily energy balance (*kcal*)	-707±368	-1085–194	135359
Basal metabolic rate (*kcal*)	1343±95	1188–1564	9088
Physical activity level	1.5±0.1	1.4–1.7	0.01
Daily total energy expenditure (*kcal*)	2034±190	1663–2361	36252
Carbohydrates (*g/kg BW*)	2.4±0.8	1.2–3.9	0.7
Total carbohydrates (*g*)	143.4±47.8	78.1–232.9	2280.4
Carbohydrates (*%E*)	43.6±7.0	30.8–55.6	49.5
Sugars (*g/kg BW*)	0.9±0.3	0.5–1.5	0.1
Total sugars (*g*)	51.5±17.0	27.9–92.5	289.6
Fat (*g/kg BW*)	0.8±0.3	0.4–1.4	0.1
Total fat (*g*)	50.7±17.1	24.7–81.2	292.0
Fat (*%E*)	34.7±5.7	21.5–44.8	32.9
Total saturated fat (*g*)	15.2±7.4	6.2–31.0	55.1
Saturated fat (*%E*)	10.2±3.4	5.0–14.8	11.3
Protein (*g/kg BW*)	1.1±0.3	0.5–1.6	0.1
Total protein (*g*)	61.8±16.9	25.3–92.0	287.0
Protein (*%E*)	19.5±5.8	11.5–31.3	34.0

S.D = standard deviation
kcal = kilocalories
BW = bodyweight
%E = percentage of total energy

Table 2: Comparison of participants' mean energy and macronutrient intakes with recommendations derived from a one sample t-test.

Intake	Mean ± SD	Recommendation	*p*-value
Energy (*kcal*)	1306±348	2000	0.000
Carbohydrates (*g/kg BW*)	2.4±0.8	3.0–5.0	0.001
Fat (*%E*)	34.7±5.7	20.0–35.0	0.004
Protein (*g/kg BW*)	1.1±0.3	1.2–2.0	0.001

SD = standard deviation
kcal = kilocalories
BW = bodyweight
%E = percentage of total energy

Participants' mean daily energy, carbohydrate and protein intakes were below that recommended for dancers. Contrastingly, fat intake was towards the upper limit of the recommended range since participants either satisfied or exceeded the recommendation. This is indicated in Table 2. No evidence of a significant difference in dietary intake between weekdays and the weekend and training and non-training days was found (p=0.309 carbohydrates, p=0.596 fat, p=0.956 protein). The Spearman Rank Correlation Coefficient only identified a significant positive association between EB and EI (p=0.021).

4. Discussion

In this study, participants were found to consume energy, carbohydrate and protein intakes below that recommended for dancers while fat intake was either adequate or exceeded the recommendation. However, since the majority were recreational dancers, it is likely that dietary intakes, except for fat, met the general RDAs. This emphasises the importance for dietary intake to reflect variations in training regimens to prevent unfavourable changes in body composition and performance. Dietary requirements vary across participants, depending on whether they classify as recreational or professional dancers. Table 3 gives an overview of energy and macronutrient intakes of dancers, comparing findings reported in this study with those from previous investigations. Variations in findings can be attributed to differences in dancers' age, genre, professional level and methodological limitations in accurately measuring EI.

Table 3: Mean±S.D of energy and macronutrient intakes of female aesthetic dance populations, including RDA for dancers for reference.

	Energy Intake	Carbohydrates	Fat	Protein	Athlete Type	Sample size
RDA (Brown *et al.*, **2020; Sousa** *et al.*, **2013; Thomas** *et al.*, **2016)**	30 kcal/kg/FFM 2000 kcal/day	3.0–5.0 g/kg BW/day 55.0–65.0 E% -	- 20.0–35.0E% -	1.2–1.7 g/kg BW/day 10.0–35.0 E% -	NA	NA
Findings from this study	- 1306±348 kcal/day	2.4±0.8 g/kg BW/day 43.6±7.0 E% 143.4±47.8 g/day	0.8±0.3 g/kg BW/day 34.7±5.7 E% 50.7±17.1 g/day	1.1±0.3 g/kg BW/day 19.5±5.8 E% 61.8±16.9 g/day	Recreational and professional dancers	14
Cohen *et al.*, **1985**	- 1,673±450 kcal/day	- 50.1±14.2 E% 206.8±70.9 g/day	- - 71.4±28.7 g/day	- 14.2±2.0 E% 59.4±18.9 g/day	Elite ballet dancers	10
Ziegler *et al.*, **2001**	- 1,416±442 kcal/day	3.4 g/kg BW/day - 183±60 g/day	- 38.0±9.5 E% 71.84±28.7 g/day	1.2 g/kg BW/day - 65±21g/day	Elite figure skating dancers	7
Dotti *et al.*, **2002**	- 1,524.5±385.9 kcal/day	- 55 E% 192.1±70.14 g/day	- 31 E% 53.32±15.54 g/day	- 14 E% 52.84±12.15 g/day	Young adult ballet dancers	30
Brown *et al.*, **2017a**	26±13 kcal/kg/FFM 2,428±458 kcal/day	5.0±1.0 g/kg FFM/day 52.0±7.0 E% 313.0±58.0 g/day	1.5±0.4 g/kg BW/day 34.0±5.0 E% 92.0±30.0 g/day	1.3±0.3 g/kg BW/day 13.0±2.0 E% 81.0±15.0 g/day	Pre-professional contemporary dancers	25

Table 3: Mean±S.D of energy and macronutrient intakes of female aesthetic dance populations, including RDA for dancers for reference.

	Energy Intake	Carbohydrates	Fat	Protein	Athlete Type	Sample size
Brown *et al.*, **2020**	- 2,040±710 kcal/day	3.7±1.6 g/kg BW/day - -	- >30 E% -	1.1±0.5 g/kg BW/day - -	Collegiate ballet and modern dancers	17
Doyle-Lucas *et al.*, **2010**	30.9±0.6 kcal/kg/FFM 1,577±89 kcal/day				Elite ballet dancers	15
Civil *et al.*, **2019**	37.0±8.0 kcal/kg/FFM 2,002±415 kcal/day	4.9±1.1 g/kg BW/day 53.4±4.0 E% 262.8±54.6 g/day	1.4±0.4 g/kg BW/day 32.7±5.0E% 74.7±21.8 g/day	1.2±0.2 g/kg BW/d 13.9±2.6 E% 65.8±12.3 g/day	Full-time vocational ballet dancers	20
Rossiou *et al.*, **2017**	- 1708 kcal/day	3.5 g/kg BW/day 46.8 E% 209.1 g/day	1.3 g/kg BW/day 39.2 E% 71.2 g/day	1.2 g/kg BW/day 14.8 E% 68.7 g/day	16 ballet, 14 modern jazz, 10 contemporary dancers	40 (32 adults)

SD = standard deviation
FFM = fat-free mass
E% = percentage of total energy
BW = bodyweight

The mean daily EI for the majority of participants was below the recommended ~2000kcal/day (p=0.000). While the reported daily EIs may be sufficient for recreational dancers, whose EEE is lower compared to professional dancers, the three professional participating dancers consumed suboptimal EIs, ranging between 1111–1623 kcal/day.

Compared to the EEE and TEE resulting from this study, previously reported values are greater and further explain why dancers of different professional levels have different dietary requirements. Full-time vocational and pre-professional dancers follow training programmes aiming to prepare them for the professional field whereas most of the participants recruited for this study danced recreationally. Moreover, the energy requirements of different genres are not clearly understood due to variations in EEE (Brown *et al.*, 2020). The relatively low resultant EEE values may also be due to decreased training schedules and performances as FDs were collected when participants were likely ending their summer break.

All participants, except for one, were found within a negative EB state. Given the participants' average normal BMI, such an observation is potentially due to under-reporting. The female gender, PA and dietary restraints present potential reasons increasing the chance of under-reporting (Macdiarmid & Blundell, 2007). Moreover, female ballerinas are known to under-report EI by 667kcal/day or 21% when compared to estimates derived from the doubly-labelled water technique (Hill & Davies, 1999). Reactivity bias may have led participants to genuinely under-eat during the 4-day timeframe. Under-recording simply due to forgetfulness presents another possible reason for resulting energy deficits (Macdiarmid & Blundell, 2007; Thompson & Subar, 2017).

Despite resulting carbohydrate intakes failing to meet recommendations, intakes are likely to be sufficient for recreational dancers. Keeping in mind professional level variety, resultant intakes were lower compared to those previously reported in literature. When training is escalated to higher intensities or durations, it is suggested that carbohydrate intakes increase to meet the elevated physiological requirements (Brown *et al.*, 2017a).

Resulting dietary fat intakes coincide with those previously reported with intakes either satisfying or exceeding the respective recommendation. Dietary fat recommendations for athletic populations should vary according to individual body composition goals and training levels and relevant to public health guidelines (Thomas *et al.*, 2016). The observed population appeared to follow a high-fat low-carbohydrate diet. Current evidence outlines that due to the resultant down-regulation of carbohydrate metabolism, regardless of glycogen availability, increased rates of fat oxidation may parallel exercise capacity or performance achieved by high carbohydrate diets only at moderate intensities. Exercise capacity or performance is impaired at higher intensities. Therefore, high-fat diets are generally concluded to diminish metabolic flexibility through the reduction of carbohydrate availability and efficiency to use the macronutrient as an exercise substrate (Burke *et al.*, 2017; Kerksick, 2019; Thomas *et al.*, 2016). Furthermore, recent evidence indicates that low carbohydrate availability, even in the absence of LEA, may pose detrimental effects on bone health and may accelerate the development of REDs (Mountjoy *et al.*, 2023).

Protein intakes also coincide with those previously reported with levels above the RDA but below that recommended for athletic populations (Brown *et al.*, 2020). Indeed, most female athletes consume protein intakes at the lower end or below the proposed recommendations (Hausswirth & Le Meur, 2011). Such targets are recommended to support metabolic adaptation and protein turnover and should be met evenly throughout the day. Guidelines should be based on training sessions, goals, dietary requirements and food choices. Emphasis is placed upon adequate EI, especially that from carbohydrates, so amino acids are spared for protein synthesis. Therefore, higher protein intakes of 1.6–2.4g/kg BW/day are suggested during periods of reduced EI (Brown *et al.*, 2017a; Brown *et al.*, 2017b; Thomas *et al.*, 2016). High protein intakes may potentially improve dancers' body composition (i.e. reduce fat mass while increasing lean soft tissue) in the absence of changes in overall BW (Brown *et al.*, 2017b; Brown *et al.*, 2019; Campbell *et al.*, 2018; Mettler *et al.*, 2010). Furthermore, low protein intakes may result in chronic negative muscle protein balance which potentially negatively influences performance more commonly in female athletes compared to the male athletic population (Brown *et al.*, 2017a; Hausswirth & Le Meur, 2011).

Similarly to participants from this study, full-time vocational ballerinas also reported similar macronutrient intakes between weekdays and weekend days (Civil *et al.*, 2019). Contrastingly, the eating behaviours of contemporary dancers differed as E% derived from carbohydrates and protein was lower during the weekend when no dance training was scheduled while fat E% was higher (Brown *et al.*, 2017a).

Resultant BMI levels may be higher compared to those available in literature due to participants being predominantly recreational dancers. Nevertheless, most (n=11) were within a normal BMI range and had lower WC than the 80cm cut-off point recommended for women. One participant reported a WC below the cut-off point despite a BMI of 24.97kg/m^2 and two participants classified as overweight and exceeded the WC cut-off point. The latter were students who trained in both ballet and modern jazz at a student level with no aspiration to train professionally.

One major strength of this study was that of being a pioneer when investigating the dietary intake of such a specific population in Malta offering a significant contribution to the local scientific knowledge base. Limitations of this study include its small sample size based in its majority on recreational dancers. Nevertheless, this should be taken within the context of the study being conducted in a small Mediterranean island with a relatively smaller general population. As participants trained in various dance genres, findings from this study were mainly compared to previous studies investigating similar genres as the energetic demands of different dance styles are poorly defined due to fluctuations in EEE (Brown *et al.*, 2017a; Brown *et al.*, 2020; Civil *et al.*, 2019). Dietary assessment in nutritional epidemiology is significantly challenged by issues of self-reported EIs, including under-reporting, under-consumption and reactivity bias, largely due to societal emphasis on aesthetic standards. Guidelines and a sample FD were provided to support coding accuracy along with the aforementioned coding protocol. However, reports with insufficient detail increase risk of misidentifying food items, portion sizes or cooking practices.

5. Conclusions

The fourteen participants were mostly recreational dancers highlighting the difference in nutritional requirements across professional level. Professional dancers failed to satisfy energy, carbohydrate and protein recommendations for athletic populations while such intakes were likely to be sufficient for recreational dancers due to enduring less exhaustive training regimens. Dietary fat recommendations were met by half of the participants and exceeded by the rest. No evidence of a significant difference in dietary intake between weekdays and the weekend and training and non-training days was found. The majority had a normal BMI and WC. Local comparison was denied since no previous local studies were available. Due to its novelty, this study offers a significant contribution to the local knowledge base. Further research is warranted, potentially expanded to the entire Maltese population, adopted to different age groups, stratified by dance genre or professional level and supported with modern technology such as GPS devices to facilitate a more objective accurate measurement of EEE, to improve understanding of the complex concept of dietary intake within dance populations. This would serve to better inform such populations about their dietary requirements.

Acknowledgments

Sincere acknowledgements towards participating dance schools and their students as without their participation this study would have not been possible.

Funding

This research has received no specific grant from any funding agency in the public, commercial or non-profit sectors.

Conflicts of Interest

The authors report no conflicts of interest.

References

Areta, J.L., Taylor, H.L. & Koehler, K. (2020). 'Low energy availability: History, definition and evidence of its endocrine, metabolic and physiological effects in prospective studies in females and males', *European Journal of Applied Physiology*, 121(1), pp. 1–21. doi:10.1007/s00421-020-04516-0.

Benson, J., Gillien, D. & Bourdet, K. *et al.* (1985). Inadequate nutrition and chronic calorie restriction in adolescent ballerinas. *The Physician and Sports Medicine*, 13(10), 79–90. https://doi.org/10.1080/00913847.1985.11708902

Brown, A., Bach, C. & De Almeida, G. *et al.* (2017b). Body composition and performance capabilities based on protein intake in collegiate dancers: A pilot study. *Journal of Sports Science*, 5(4). https://doi.org/10.17265/2332-7839/2017.04.001

Brown, A., Brooks, S. & Smith, S., *et al.* (2020). Female collegiate dancers body composition, macronutrient and micronutrient intake over two academic years: A longitudinal analysis. *Journal of Functional Morphology and Kinesiology*, 5(1), 17. https://doi.org/10.3390/jfmk5010017

Brown, A., Welsh, T. & Panton, L., *et al.* (2019). Higher-protein intake improves body composition index in female collegiate dancers. Applied Physiology, Nutrition, and Metabolism, 45(5), 547–554. https://doi.org/10.1139/apnm-2019–0517

Brown, M., Howatson, G & Qiun, E. *et al*. (2017a). Energy intake and energy expenditure of pre-professional female contemporary dancers. *PLOS ONE*, 12(2), e0171998. https://doi.org/10.1371/journal.pone.0171998

Burke, L., Lundy, B. & Fahrenholtz, I., *et al.* (2018). Pitfalls of conducting and interpreting estimates of energy availability in free-living athletes. *International Journal of Sport Nutrition and Exercise Metabolism*, 28(4), 350–363. https://doi.org/10.1123/ijsnem.2018–0142

Burke, L., Ross, M. & Garvican-Lewis, L., *et al.* (2017). Low carbohydrate, high fat diet impairs exercise economy and negates the performance benefit from intensified training in elite race walkers. *The Journal of Physiology*, 595(9), 2785–2807. https://doi.org/10.1113/jp273230

Campbell, B., Aguilar, D. & Conlin, L., *et al.* (2018). Effects of high versus low protein intake on body composition and maximal strength in aspiring female physique athletes engaging in an 8-week resistance training program. *International Journal of Sport Nutrition and Exercise Metabolism*, 28(6), 580–585. https://doi.org/10.1123/ijsnem.2017–0389

Challis, J. *et al.* (2020). 'Dietary intake, body composition, and nutrition knowledge of Irish dancers', *Journal of Dance Medicine &; Science*, 24(3), pp. 105–112. doi:10.12678/1089–313x.24.3.105.

Civil, R., Lamb, A. & Loosmore, D., *et al.* (2019). Assessment of dietary intake, energy status, and factors associated with RED-S in vocational female ballet students. *Frontiers in Nutrition*, 5(136). https://doi.org/10.3389/fnut.2018.00136

Cohen, J., Potosnak, L. & Frank, O., *et al.* (1985). A Nutritional and Hematologic Assessment of Elite Ballet Dancers. *The Physician and Sports Medicine*, 13(5), 43–54. https://doi.org/10.1080/00913847.1985.11708788

Dotti, A., Fioravanti, M. & Balotta, M., *et al.* (2002). Eating behaviour of ballet dancers. Eating and Weight *Disorders – Studies on Anorexia, Bulimia and Obesity*, 7(1), 60–67. https://doi.org/10.1007/bf03354431

Doyle-Lucas, A., Akers, J. & Davy, B. (2010). Energetic efficiency, menstrual irregularity, and bone mineral density in elite professional female ballet dancers. *Journal of Dance Medicine*, 14(4), 147–154.

Egan, B. (2016). Protein intake for athletes and active adults: Current concepts and controversies. *Nutrition Bulletin*, 41(3), 202–213. https://doi.org/10.1111/nbu.12215

Guan, V., Probst, Y. & Neale, E., *et al.* (2019). Evaluation of the dietary intake data coding process in a clinical setting: Implications for research practice. *PLOS ONE*, 14(8), e0221047. https://doi.org/10.1371/journal.pone.0221047

Hajian, M. & Mohaghegh, S. (2024). 'Do Iranian female dancers face a risk of chronic negative energy balance? A study on energy intake and expenditure among non-professional female dancers', *Asian Journal of Sports Medicine*, 15(2). doi:10.5812/asjsm-139119.

Hausswirth, C. & Le Meur, Y. (2011). Physiological and nutritional aspects of post-exercise recovery. *Sports Medicine*, 41(10), 861–882. https://doi.org/10.2165/11593180–000000000–00000

Hill, R. & Davies, P. (1999). The validity of a four-day weighed food record for measuring energy intake in female classical ballet dancers. *European Journal of Clinical Nutrition*, 53(9), 752–753. https://doi.org/10.1038/sj.ejcn.1600836

IPAQ (2002). International Physical Activity Questionnaire (August 2002) short last 7 days self-administered format. Available at: https://journals.plos.org/plosone/article/file?type=supplementary&id=info:doi/10.1371/journal.pone.0219193.s010 (accessed 11 January 2021).

IPAQ (2005). Guidelines for data processing and analysis of the International Physical Activity Questionnaire (IPAQ) – Short and long forms. Available at: https://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbnxoaGVpcGFxfGd4OjE0NDgxMDk3NDU1YWRlZTM (accessed 20 January, 2021).

Kerksick, C. (2019). Requirements of proteins, carbohydrates, and fats for athletes. *Nutrition and Enhanced Sports Performance*, 443–459. https://doi.org/10.1016/b978–0–12–813922–6.00038–2

Koutedakis, Y. & Jamurtas, A. (2004). The dancer as a performing athlete. *Sports Medicine*, 34(10), 651–661. https://doi.org/10.2165/00007256–200434100–00003

Macdiarmid, J. & Blundell, J. (2007). Assessing dietary intake: Who, what and why of under-reporting. *Nutrition Research Reviews*, 11(2), 231–253. https://doi.org/10.1079/nrr19980017

Mangion, D. (2010). *Athletes' nutritional practices: A local study*. Degree of Bachelor's in Education Thesis, University of Malta, Malta.

Mettler, S., Mitchell, N. & Tipton, K. (2010). Increased protein intake reduces lean body mass loss during weight loss in athletes. *Medicine & Science in Sports & Exercise*, 42(2), 326–337. https://doi.org/10.1249/mss.0b013e3181b2ef8e

Mountjoy, M. et al. (2023) '2023 International Olympic Committee's (IOC) consensus statement on Relative Energy Deficiency in sport (reds)', *British Journal of Sports Medicine*, 57(17), pp. 1073–1098. doi:10.1136/bjsports-2023–106994.

Nepocatych, S., Balilionis, G. & O'Neal, E. (2017). Analysis of dietary intake and body composition of female athletes over a competitive season. *Montenegrin Journal of Sports Science and Medicine*, 6(2), 57–65. https://doi.org/10.26773/mjssm.2017.09.008

Nutritics (2021). Research Edition (v5.095) [Computer software]. Dublin. Retrieved from https://www.nutritics.com/p/home.

Nutritics (2021). Libro (v.09) [Mobile application software]. Dublin. Retrieved from https://www.nutritics.com/p/home

Patel, S., Mills, A. & Crawley, H., et al. (2012). Food portion sizes (3rd ed.). London: H.M.S.O.

Puš, T., Podgrajšek, K. & Simčič, M. (2012). Nutrient intake variability induced by processing of food diary data: a pilot study. *Acta Agriculturae Slovenica*, 2(100), 117–121. Retrieved 20 March 2021, from http://aas.bf.uni-lj.si/zootehnika/100–2012/PDF/100–2012–2-117–121.pdf.

Rossiou, D., Papadopoulou, S & Pagkalos, I., et al. (2017). Energy expenditure and nutrition status of ballet, jazz and contemporary dance students. *Progress in Health Sciences*, 7(1), 31–38. https://doi.org/10.5604/01.3001.0010.1771

Scientific Advisory Committee on Nutrition (2011). Dietary reference values for energy (pp. 12–22). London: Crown.

Sousa, M., Carvalho, P. & Moreira, P. et al., (2013). Nutrition and nutritional issues for dancers. *Medical Problems of Performing Artists*, 28(3), 119–123. https://doi.org/10.21091/mppa.2013.3025

Struab, C. (2018). *A study of dance wellness for females age 8–13*. Degree of Bachelor's of Arts Thesis, University of Malta, Malta.

Thomas, D., Erdman, K. & Burke, L. (2016). Position of the Academy of Nutrition and Dietetics, Dietitians of Canada, and the American College of Sports Medicine: Nutrition and athletic performance. *Journal of the Academy of Nutrition and Dietetics*, 116(3), 501–528. https://doi.org/10.1016/j.jand.2015.12.006

Thompson, F. & Subar, A. (2017). Dietary assessment methodology. *Nutrition in the Prevention and Treatment of Disease*, 5–48.https://doi.org/http://dx.doi.org/10.1016/B978–0–12–802928–2.00001–1

Twitchett, E., Koutedakis, Y. & Wyon, M. (2009). Physiological fitness and professional classical ballet performance: A brief review. *Journal of Strength and Conditioning Research*, 23(9), 2732–2740. https://doi.org/10.1519/jsc.0b013e3181bc1749

World Health Organization Expert Consultation (2011). Waist circumference and waist–hip ratio: Report of a WHO Expert Consultation Geneva, 8–11 December 2008. Geneva: WHO Document Production Services, Geneva, Switzerland. Retrieved from https://apps.who.int/iris/bitstream/handle/10665/44583/9789241501491_eng.pdf?sequence=1&isAllowed=y

World Health Organization. Body mass index – BMI. Euro.who.int. Retrieved 8 March 2021, from https://www.euro.who.int/en/health-topics/disease-prevention/nutrition/a-healthy-lifestyle/body-mass-index-bmi.

Ziegler, P., Jonnalagadda, S. & Lawrence, C. (2001). Dietary intake of elite figure skating dancers. *Nutrition Research*, 21(7), 983–992. https://doi.org/10.1016/s0271–5317(01)00320–7.

Malta Journal of Health Sciences
https://doi.org/10.14614/EXEROLDER/12/25

Research Article

The Perceptions of Older Persons on Exercise

Ms Eliza Grech Debattista
Department of Physiotherapy, St Vincent de Paul, Malta
Dr Karl Spiteri
St Vincent de Paul Residence, St Vincent de Paul, Malta
Dr John Xerri de Caro
Department of Physiotherapy, Faculty of Health Sciences, University of Malta
Mr Emanuel Schembri
Department of Physiotherapy, Mater Dei Hospital, Malta
Dr Maria Aurora Fenech
Department of Gerontology and Dementia Studies, University of Malta

Abstract

Older people are encouraged to remain physically active and engage in exercise to remain independent. An understanding of exercise for older people is needed when promoting exercise in this age group. The study aimed to explore older persons' perceptions of exercise in a Maltese community. It included exploring the incentives to commence/or continue to exercise, and barriers to exercising. An Interpretative Phenomenological Analysis (IPA) was used for data analysis to obtain an in-depth perspective from the participants about exercise. Participants were recruited from an activity day centre and were all residents from one village. Data was collected between July and August 2020 through semi-structured interviews, which were audio-recorded and transcribed. These were used to obtain an in-depth perspective about their exercise perceptions. Purposeful quota sampling was employed, with a sample size of nine participants determined based on data saturation. Data coding followed transcript review, leading to the identification and discussion of emergent themes. The study included nine participants between the ages of 64 and 88; six females and three males. Five main themes emerged i) Perceptions of One's Own Ability, ii) Knowledge on Exercise, Exercising Preference, iii) Adjusting to New Circumstances, iv) Accessibility and v) Opportunities to Exercising. This study showed that older persons had different views and opinions about exercise. Reasons why they wanted to initiate exercising and what kept them doing so varied based on their perceptions. These preferences need to be considered when promoting exercise with older persons.

Keywords: Physical activity, Exercise, Older people, Aged

Correspondence to Karl Spiteri
karl.b.spiteri@gov.mt
Received: 9.5.2024
Revised: 7.10.2024
Accepted: 12.10.2024
Published: 30.6.2025

1. Introduction

As people get older, there is a decrease in physical activity and an increase in sedentary behaviour (Harvey, Chastin & Skelton, 2013). Concurrently as people age, there is an increased need to carry out different types of physical activity, and not limit oneself to aerobic type of activities, as evidenced by the World Health Organisation (WHO) guidelines (Bull *et al.,* 2020). Due to the onset of sarcopenia and a decrease in balance reactions with ageing, it is recommended that older persons engage in strength and balance exercises three times per week (Physical Activity

Guidelines Advisory Committee, 2018). In Malta, only 25% of older adults (60 years +) match the WHO aerobic guidelines (WHO, 2018). There are various barriers and motivators to initiating or maintaining physical activity which were identified in this age group (Spiteri et al., 2019).

The prevalence of chronic disease in those aged over 65 years of age was found to be over 80% in the United States (Garin *et al.,* 2016). Living well with chronic diseases was defined as achieving the best health status possible that besets social, cognitive and physical well-being (Wallace *et al.,* 2012). Engaging in health-promoting behaviours such as regular exercise contributes to achieving this state (Talley *et al.,* 2019). Benefits of exercise in older persons include improved cognitive functioning, enhanced mood and improvements in mental health, preventing falls and improved balance, delayed disability, and possibly even reversing the effects of metabolic diseases (Parra *et al.,* 2019).

An important aspect of promoting exercise among older persons is understanding their perceptions and tailoring health promotion interventions to increase the uptake of exercise. The study aimed to explore the perceptions of exercise amongst older persons living in a Maltese community. The objectives were to: (a) investigate what older persons perceived exercise to be, (b) explore motivational factors that encourage older persons to exercise, and (c) analyse barriers that deter older persons from exercising.

2. Method

Interpretative Phenomenological Analysis (IPA) was used to obtain an in-depth perspective about exercise. Eligibility criteria for participating in the project included older persons living in the community, attending a day centre and mobilising independently. Ethics approval for the study was obtained from the University of Malta Faculty Research Ethics Committee (5594_01062020). Participants were contacted via an intermediary, who explained the nature of the study and distributed information letters in the language of the participant's choice (Maltese or English). Older persons interested in partaking in the project contacted the primary researcher (EG) themselves through the contact details provided in the information letter. The study was explained in further depth to the prospective participants by the primary researcher; older person participants then signed the consent form. Dates and times convenient for the participants were agreed upon with the primary researcher. Interviews were carried out within a private area of the Day Centre attended by the participants. The interview consisted of open-ended questions (Appendix 1), which were audio-recorded and then transcribed by EG. The interview allowed the researcher to build a rapport with the individuals and was used to develop trust with the participants (Smith, Flowers & Larkin, 2021). This helped participants feel comfortable, thus becoming more open, and willing to share their personal experiences about their engagement in exercises. The researcher explored barriers and motivators participants experienced when trying to engage with exercises. A reflective diary was kept, and any comments were noted by the researcher.

Purposeful quota sampling was used with the sample size set at nine participants, with a mix of genders (Frechette *et al.,* 2020; Smith, Flowers & Larkin, 2021). If data saturation was not reached further recruitment of participants was considered. Data saturation was considered if no further new themes were identified from the last interview (Braun & Clarke, 2021). Data was collected between July and August 2020.

Familiarisation with the layout of the scheduled interview was ensured by conducting a pilot study with two older person participants (Mcgrath, Palmgren and Liljedahl, 2018). The pilot studies were included and integrated with the other main-study participants as they were deemed of good quality by the research team.

3. Data analysis

Interviews were transcribed verbatim by the primary researcher. To allow for immersion into the data (Frechette *et al.,* 2020; Smith, Flowers & Larkin, 2021), the research team went through the transcripts to ensure consistency. Transcripts were re-read with the audio, and salient experiences recorded in the notes taken during the interviews were also included. Data coding was carried out after re-reading the transcripts. In analysing the codes, emergent themes were developed and then discussed. Themes represented the participants' primary and original thoughts, as well as the researcher's comprehension of such thoughts and perceptions (Frechette *et al.,* 2020; Smith, Flowers & Larkin, 2021). A participant's profile was also developed based on the interview to provide a thick description and the idiographic aspect of the IPA. NVIVO 13® software assisted in the analysis.

4. Findings

A total of 15 participants were approached, and nine older persons (six older persons identified as female, three older persons identified as male), accepted to participate in the study. Ages ranged from 66 to 88 years. Six of the participants lived with their spouses, three lived alone, and eight of the older participants had children. The older persons attended the village Day Centre twice weekly. Demographic information is presented in Table 1.

Table 1: Demographic information on participants

Pseudonyms given to participants	Sex	Age	Marital status	Previous Employment	Mobility Status	Living situation
Helena	F	66	Living with spouse	Food factory line worker	Independent	Lives with husband
Joyce	F	88	Widowed, living alone	Housekeeper	Independent with a walking stick	Lives alone
Julia	F	68	Living with spouse	Housekeeper	Independent	Lives with husband and daughter
Rose	F	76	Living with spouse	Housekeeper	Independent	Lives with husband
Chris	M	79	Living with spouse	Ex-Army	Independent	Lives with wife
Gorg	M	84	Living with spouse	Shipwright	Independent with a walking stick	Lives with wife
Bertu	M	75	Widowed, living alone	Manual labourer	Independent	Lives alone
Miriam	F	88	Widowed, living alone	Housekeeper	Independent with a walking stick	Lives alone
Lydia	F	71	Living with spouse	Housekeeper	Independent	Lives with husband

The results present an idiographic representation of how two participants perceived exercise and the thematic aspect. Two participants' stories were chosen for their contrasting exercise experiences throughout life, and their story is presented in Table 2. One participant was never able to exercise, and her perception of exercise is bound to her daily walk-in during errands. The other participant used to exercise all his life since he was in his teenage years. This convinced him that exercise was an essential part of his life, and he tried to engage in it regularly.

Table 2. Exercise experience from two participants.

Joyce, 88 years old and lives alone since her husband passed away. She has groceries delivered from the grocer beneath her house. Meals delivered are delivered through the 'Meals-on-Wheels Service'. She never had an opportunity to practice sports. She did, however, enjoy taking part in social activities. Joyce walks frequently in summer accompanied by her friend; both Joyce and the friend used a walking aid. At times, Joyce requires manual help to go up/down a pavement. Her outdoor recreation activities depended on whether others could accompany her or not. She was afraid to go out alone, especially at night. Joyce expressed her preference to exercise with people she felt comfortable with, as different opinions bothered her. She preferred one-to-one exercise and in her opinion going up and down her stairs at home was enough exercise. With respect to TV exercise programmes, she indicated that she preferred some of the exercises, as the rest were deemed to be too energetic for her taste. At the Day Centre, Joyce sporadically participated in the exercise classes that were organised. She claimed exercises were beneficial for pain relief and re-energising oneself.	Bertu, 75 years old and a returning migrant. He preferred living in Malta due to having more social engagement and a better social life. He worked until the age of 65 years in a leisure/sports centre. At the age of 15 years, he started training in a gym and, since then continued to maintain an active lifestyle. Bertu was currently awaiting doctor's instructions and certification to exercise in a gym, since he had recently been operated upon. He was on medications for hypertension and renal problems. He mentioned the experiences he had in weight training by mentioning how muscle gain was achieved, when to best take protein, muscle tears and the negative effects of steroids. He enjoyed being guided by professionals and felt comfortable knowing the exercises beforehand. He followed social media daily and subscribed to exercise and gym channels. Older persons older than himself inspired him to be active at the gym. He was an avid swimmer and went to the beach daily. He led some exercise classes at the Day Centre and planned them ahead with adequate music and constant instructions and encouragement. He walked daily, 30 minutes in the morning and 30 minutes in the evening. He preferred summer to winter because he could enjoy walking and swimming, although in winter, he tended to walk for longer distances.

The second part of the results presents the five identified themes: (1) Self-efficacy and perceptions of one's own ability, (2) Personal biography, Health literacy, Knowledge and information sources, (3) Exercising within a group versus Individualised settings, (4) Adjusting to new circumstances, and (5) Accessibility and opportunities to carry through and/or initiate exercise. Table 3 below represents the main themes and their subthemes.

Table 3. Study themes.

Theme	Subtheme	Interview abstracts
Perceptions of One's Own Ability	Comparison of present self to past self-experiences, shaping current choices	I tell myself is it because I'm lazy or it's in my brain? I think I decreased because I chose to decrease... I've been decreased bit by bit
	Setting limits-age discrimination from others and from one-self	you go whom to ever you go to, most of them that's what they tell you. "Walk!" Maybe if the doctor tells you to do them, do ten, maybe I do some and stop
	Environmental, health and psychological issues	even my legs, they keep me back from doing more because I suffer from leg pain I start fearing that my leg will start hurting me and I won't have strength in it
Knowledge on Exercise	Accumulated past opportunities and experiences that influenced knowledge on exercise	If you've ever lifted weights, it's in your blood, in your system. Speaking for myself, it's in my blood. I was fifteen years old when I had used to go...when you start these things, it's in your blood, in your system. But I kept going! Even in Australia I trained...
	Distinguishing between leisured and structured exercise with age-appropriate exercises	... I try to walk. For example, today, I mean, before I came here, this morning I woke up at quarter to six. I went to XXXX and back.
Exercising Preference	Specific goals and aims	Loose weight so that I become lighter. So that I go back to what I was before... skinny! Active! I find some exercises where I maybe can get more movements where my joints "melt" (melt; to become more flexible
	Social engagement; an incentive or a barrier	...I enjoy it more if it's one-on-one... because someone says one thing and someone says something else! And I don't like it! When they tell us to do something, I tell them, "yes!". I would be eagerly waiting for it!...I think, for myself apart from that, when there is someone who is sort of forces you, sort of...gets you into it, pushes you. All right, once that occurs then yes. You are more encouraged. On our own it's like...
	Choice of exercise dependent on personal preferences	I know what I have to do, but I leave it up to the instructor I want something, exercise, really good for legs, so they become light and I can be... flexible... so I can walk properly. Because I suffer to walk.

Table 3. Study themes.

Theme	Subtheme	Interview abstracts
Adjusting to New Circumstances	Different mediums to present and offer exercise opportunities – social media, group classes and therapeutic interventions	sometimes I see on the television with older persons. There's much older than myself and they're supple! I tell myself, "oh bless them/poor them look at them", but I tell myself that they stayed active On the internet I see those aged ninety going to the gym! *very excitedly said* Ninety years old they go to the gym
	Acquisitions of new skills and habits	you need to have willpower because, when you go, you start an activity at the gymnasium and so on, it's like it stays in your blood!
Accessibility and Opportunities to Exercising	Transport and facility access including convenience and environment.	on the sand you can always go down slowly slowly. And on the rocks I don't know how to jump in either! So...the smoother it is. It's better. Pavement...here we have a pavement along the shore. Because I start from near the boules club, I don't know if you know where it is? I go till restaurant and back uu!
	Autonomy and choice on where and how to exercise	In winter, we don't go to the sea, so I said maybe I join a swimming pool, if it's not expensive, they give me something (reduced price) Because I live across the road from the sea

5. Discussion

This study aimed to understand older people's perceptions of exercise in a seaside community. This was achieved by using an IPA methodology which integrated the narrative aspect of the individual participants' and their collective understanding. Findings from this study add to the body of literature on the topic, adding a local context to the findings. Study findings are similar to Franco *et al.* (2015) and Collado-Mateo *et al.* (2021), which support the need for awareness of the benefits of exercise for people to engage with it. It adds that what people perceive as exercise is related to their past experiences of exercise, like Rodrigues *et al.* (2019).

The perception of exercise was based on the person's personal preference. Walking was the exercise of choice of the participants because it was free, deemed safe, and practical. Older persons who grew up with a structured daily routine, leaving little time for participating in exercise might explain why most participants perceived walking as the main type of exercise (Witcher, 2017).

The locality understudy offered options where one could walk near the sea. The outdoor environment offered various walking opportunities due to its promenade. Participants identified the vicinity and accessibility to the sea as important for them to engage in swimming as an exercise. However, this was limited to specific seasons. All participants in this study construed exercise as primarily walking because it did not cause them pain or discomfort and was not considered to be high effort. Walking within the daily errand routine was the participants' idea of exercising. They used walking as a measure of their health status. Participants timed themselves or used landmarks as targets. Similar to other studies, they used the duration of their trips to experience how fit and healthy they were (Pae & Akar, 2020). Participants set personal limitations on how much they exercise. Setting limits was affected by: (a) one's own self-discrimination, (b) discrimination coming from others, (c) pre-conceived ideas about exercise, (d) the consideration of those exercises deemed appropriate for older persons, and (e) self-knowledge.

Participants mentioned technology as a way to engage with exercise mediums such as television (TV) and social media (Facebook, YouTube) in a positive manner. The use of technology by older people to exercise was on a steady increase, partially owing to the COVID-19 restrictions, which advised older persons to remain at home and exercise indoors (WHO, 2020). All participants in this study mentioned exercises being delivered daily via television programmes, specifically aimed towards their age group. Other participants mentioned social media such as Facebook and YouTube to gain access to exercises. Through this medium, the participants were able to choose which exercises they felt most appealed to them. Geraedts *et al.* (2017) mentioned the increasing feasibility of using internet-delivered exercises, addressed specifically towards older persons. In Malta, computer literacy in older persons was an issue. Formosa, (2013) found, via a qualitative study using semi-structured interviews on sixty-six older Maltese persons, that many older persons felt too old to learn to use technology. They felt anxious if they had to use a computer and, if the past, had not made use of such technologies.

The current study identified barriers and motivators towards exercise engagement based on participants' subjective perceptions of exercise. All the participants mentioned that they engage in exercise to avoid or prolong the process of becoming physically dependent. Another motivator was the social engagement aspect of meeting peers and people from their same age group, exercising. Being with persons of the same age was indicated as the preferred choice by older persons, as it was deemed less competitive, provided support and encouragement (Farrance *et al.*, 2016; Talley *et al.*, 2019). Exercising was linked to an improved quality of life and health, thus delaying the likelihood of being dependent on others (Svantesson *et al.,* 2015). This was similar to what the participants in this study regarded as encouraging them to exercise. Knowledge influenced exercising behaviour since being aware of the benefits of exercise positively altered the participants' perception of exercise. Health education about exercise can act as a motivator towards exercise (Gothe & Kendall, 2016).

Another motivator that influenced the participants was the environment. This included: (a) the proximity to the place where exercise was to take place, (b) accessible and safe locations, and (c) a choice of different mediums for exercise possibilities. Like Tong *et al.* (2020), older women preferred local exercise activities, since most did not drive and relied on others or public transport to get around. Having accessible locations aided in exercise uptake and social interaction, and lower socio-economical groups were the most frequent users (Anthun *et al.*, 2019).

A barrier to exercising identified by the study participants was fear of falling. This threatened the older person's autonomy (Denkinger *et al.*, 2015). Pae and Akar, (2020) explained the phenomenon of limiting what exercises an older person carried out as a reflection of how the individual perceived their health status. When older persons felt confident in their abilities, they exercised with increased positivity.

For interventions to be effective in promoting exercise with older people, perceptions of exercise must be considered. This can allow them to build onto their current knowledge. Room *et al.*, (2017) identified that interventions based on feedback and monitoring effectively promoted exercise in older people. The need for support and monitoring was identified among study participants. In their perception of exercise participants considered their perceived abilities, preferences, and their knowledge. They were able to adjust to new circumstance in using technology to engage with exercises. Accessibility influenced the type of exercises they engaged in within their locality. Being a seaside locality with a promenade and easy access to the sea might have influenced their perception of exercise.

6. Strengths and limitations

Using IPA, the study captured lived experiences of exercise, with a depth and richness that allowed the lived experiences of older persons to be brought to light (Smith, Flowers & Larkin, 2021). Semi-structured interviews allowed participants to elaborate further and express their perspectives on exercise (Patton and Patton, 2015; Smith, Flowers & Larkin, 2021). In using an intermediary to recruit participants, it was guaranteed that the researcher did not have any previous contact with the participants, thus reducing the possibility of bias (Yin, 2016). The participants were interviewed in a location and time of their choosing to help them feel at ease, allowing for a more comfortable and open interview process. A reflective diary was kept throughout the whole study (Frechette *et al.,* 2020; Smith, Flowers & Larkin, 2021). This gave more rigour to the study and allowed the reflection on knowledge and previous views on exercise (Horrigan-Kelly, Millar & Dowling, 2016).

A limitation of this study was that participants were all attendees at the day centre and had the opportunity to partake in exercises offered by the centre. The study was focused on one locality and, therefore, cannot be used to generalise other older people. Most of the participants were female. There might have been gender differences in how the participants perceived exercise, but due to the small sample size, it was impossible to analyse this further.

7. Conclusion

This study interviewed older people from different aspects of life who lived in the community. It highlights the heterogeneity of older persons and the challenge of addressing their beliefs. Participants had different beliefs about what exercise meant to them. These need to be considered when developing health promotion programs for older people. Using terms such as exercise, strength exercises, or physical activity might mean different things to older people.

8. Funding

This research has received no specific grant from any funding agency in the public, commercial or non-profit sectors.

9. Conflicts of interest

The authors declare that they have no conflict of interest.

References

Anthun, K., Maass, R., Hope, S., Espnes, G., Bell, R., Khan, M., & Lillefjell, M. (2019). Addressing Inequity: Evaluation of an Intervention to Improve Accessibility and Quality of a Green Space. *International Journal Of Environmental Research And Public Health*, 16(24), 5015. https://doi.org/10.3390/ijerph16245015

Braun, V. and Clarke, V. (2021). 'To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales', *Qualitative Research in Sport, Exercise and Health*, 13(2), pp. 201–216. doi: 10.1080/2159676X.2019.1704846.

Bull, F. C. *et al.* (2020). 'World Health Organization 2020 guidelines on physical activity and sedentary behaviour', *British Journal of Sports Medicine*, 54(24), pp. 1451–1462. doi: 10.1136/bjsports-2020–102955.

Collado-Mateo, D. *et al.* (2021). 'Key factors associated with adherence to physical exercise in patients with chronic diseases and older adults: An umbrella review', *International Journal of Environmental Research and Public Health*, 18(4), pp. 1–24. doi: 10.3390/ijerph18042023.

Denkinger, M. D., Lukas, A., Nikolaus, T., & Hauer, K. (2015). Factors Associated with Fear of Falling and Associated Activity Restriction in Community-Dwelling Older Adults: A Systematic Review. *The American Journal of Psychiatry*, *23*(1), 72–86. https://doi.org/https://doi.org/10.1016/j.jagp.2014.03.002

Farrance, C., Tsofliou, F., & Clark, C. (2016). Adherence to community based group exercise interventions for older people: A mixed-methods systematic review. *Preventive Medicine*, *87*, 155–166. https://doi.org/10.1016/j.ypmed.2016.02.037

Franco, M. R. *et al.* (2015). 'Older people's perspectives on participation in physical activity: A systematic review and thematic synthesis of qualitative literature', *British Journal of Sports Medicine*, 49(19), pp. 1268–1276. doi: http://dx.doi.org/10.1136/bjsports-2014–094015.

Frechette, J. *et al.* (2020). 'Capturing Lived Experience: Methodological Considerations for Interpretive Phenomenological Inquiry', *International Journal of Qualitative Methods*, 19, pp. 1–12. doi: https://doi.org/10.1177/1609406920907254.

Geraedts, H. A., Zijlstra, W., Zhang, W., Spoorenberg, S. L., Báez, M., Far, I. K., ... Stevens, M. (2017). A Home-Based Exercise Program Driven by Tablet Application and Mobility Monitoring for Frail Older Adults: Feasibility and Practical Implications. *Preventing Chronic Disease, 14*. https://doi.org/10.5888/pcd14.160227

Garin, N., Koyanagi, A., Chatterji, S., Tyrovolas, S., Olaya, B., Leonardi, M., Lara, E., Koskinen, S., Tobiasz-Adamczyk, B., Ayuso-Mateos, J. L., & Haro, J. M. (2016). Global Multimorbidity Patterns: A Cross-Sectional, Population-Based, Multi-Country Study. *Journals of Gerontology - Series A Biological Sciences and Medical Sciences*, 71(2), 205–214. https://doi.org/10.1093/gerona/glv128

Gothe, N. P., & Kendall, B. J. (2016). Barriers, Motivations, and Preferences for Physical Activity Among Female African American Older Adults. *Gerontology and Geriatric Medicine*, *2*, 233372141667739. https://doi.org/10.1177/2333721416677399

Harvey, J. A., Chastin, S. F. M. and Skelton, D. A. (2013). 'Prevalence of Sedentary Behavior in Older

Adults: A Systematic Review', *International Journal of Environmental Research and Public Health*, 10(12), pp. 6645–6661. Available at: http://lcproxy.shu.ac.uk/login?url=http://search.proquest.com/docview/1496659573?accountid=13827.

Horrigan-Kelly, M., Millar, M. and Dowling, M. (2016). 'Understanding the Key Tenets of Heidegger's Philosophy for Interpretive Phenomenological Research', *Journal of Qualitative Methods*, 15(1). doi: https://doi.org/10.1177/1609406916680634.

Mcgrath, C., Palmgren, P. J. and Liljedahl, M. (2018). 'Twelve tips for conducting qualitative research interviews', *Medical Teacher*, 41(9), pp. 1002–1006. doi: doi:10.1080/0142159x.2018.1497149.

Pae, G., & Akar, G. (2020). Effects of walking on self-assessed health status: Links between walking, trip purposes and health. *Journal of Transport & Health*, 18, 100901. https://doi.org/10.1016/j.jth.2020.100901

Parra, D. C. *et al.* (2019). 'A qualitative study of older adults' perspectives on initiating exercise and mindfulness practice', *BMC geriatrics*, 19(354), pp. 1–11. doi: https://doi.org/https://doi.org/10.1186/s12877–019–1375–9.

Patton, M. Q. and Patton, M. Q. (2015). *Qualitative research & evaluation methods: integrating theory and practice*. Sage.

Physical Activity Guidelines Advisory Committee (2018). *Physical activity guidelines advisory committee scientific report*. Washington DC. doi: 10.1111/j.1753–4887.2008.00136.x.

Rodrigues, F., Bento, T., Cid, L., Pereira Neiva, H., Teixeira, D., & Moutão, J. et al. (2018). Can Interpersonal Behavior Influence the Persistence and Adherence to Physical Exercise Practice in Adults? A Systematic Review. *Frontiers In Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.02141

Room, J., Hannink, E., Dawes, H., & Barker, K. (2017). What interventions are used to improve exercise adherence in older people and what behavioural techniques are they based on? A systematic review. *BMJ Open*, 7(12), 1–12. https://doi.org/10.1136/bmjopen-2017-019221

Smith, J. A., Flowers, P. and Larkin, M. (2021). *Interpretative phenomenological analysis: Theory, method and research*. 2nd edn. London: Sage.

Spiteri, K. *et al.* (2019). 'Barriers and motivators of physical activity participation in middle-aged and older adults—a systematic review', *Journal of Aging and Physical Activity*, 27(6), pp. 929–944. doi: 10.1123/japa.2018–0343.

Svantesson, U. *et al.* (2015). 'Impact of Physical Activity on the Self-Perceived Quality of Life in Non-Frail Older Adults', *Journal of Clinical Medicine Research*, 7(8), pp. 585–93. doi: 10.14740/jocmr2021w.

Talley, K. *et al.* (2019). 'Aging Adults' Preferences for Wellness Program Activities and Delivery Characteristics', *Topics In Geriatric Rehabilitation*, 35(4), pp. 289–299. doi: https://doi.org/10.1097/tgr.0000000000000247.

Tong, C. E., McKay, H. A., Martin-Matthews, A., Mahmood, A., & Sims-Gould, J. (2020). "These Few Blocks, These Are My Village": The Physical Activity and Mobility of Foreign-Born Older Adults. *The Gerontologist*, 60(4), 638–650. https://doi.org/https://doi.org/10.1093/geront/gnz005

Wallace, R. B. & *et al.* (2012). *Living well with chronic illness: a call for public health action*. The national academies press.

WHO (2018). *Malta physical activity fact sheet 2018*, *World Health Organisation*. Available at: http://www.euro.who.int/__data/assets/pdf_file/0008/382571/malta-eng.pdf?ua=1 (Accessed: 7 November 2019).

World Health Organisation. (2020, August 6). Coronavirus disease (COVID-19): Small public gatherings. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-covid-19-small-public-gatherings.

Witcher, C. S. G. (2017). Rural Older Adult Physical Activity Promotion. *Topics in Geriatric Rehabilitation*, *33*(3), 162–169. https://doi.org/10.1097/tgr.0000000000000152

Yin, R. (2016). *Qualitative research from start to finish*. 2nd edn. The Guilford Press.

Appendix 1

Semi-Structured Interview Schedule

1. Demographic data:

- Locality, Status, Age, Previous/current occupation, Gender

2. I am interested in your perceptions about exercise. What do you understand about "Exercise"?

Prompts

- Do you think there are different forms of exercise?
- Do you think you do enough exercise during the week?
- Do you think there is a limit to the amount of exercise one should do?
- Do you carry out any form of exercise?

If Yes:

Prompts

- What sort? Would you consider housework or going for errands as exercise? Why do you exercise?

Prompts

- Environment?
- Health benefits? Want to remain independent? Not wanting to become a burden on family?
- Part of routine?
- Wife/husband/family push you to exercise?
- Social/Making friends?
- Previous experience of exercise?

If No:

Prompts

- Why did you stop? What kind of exercises did you use to do? What hinders you from restarting?

Prompts

- Why not? Do you feel what you do during the day is enough?

3. Are there things which stop you from exercising as much as you wish you would?

Prompts

- Time, Money, Family, Environment, Feeling shy

4. If you had an opportunity to exercise, how would you do it?

Prompt

- Do you prefer group exercising or individual?
- Do you feel pressured?
- Don't like having people of different ages?
- Want to make friends in a group?
- Feeling shy?
- Feel the need for individual attention?

5. Do you think you are informed enough about exercise?

Prompts

- What is your current knowledge and education about both? From where did you get this knowledge? Do you want more information? In what format?

6. Do you know who to ask for advice or help on exercise?

Prompts

- Doctor? Physiotherapist? Neighbours? Internet?

7. Do you feel technology helps you keep physically active and exercising?

Prompts

Yes or No? If yes, what type of technology? If no, why not?

8. Would you consider being monitored and keeping track of your exercise and physical activities, via a computer or smartphone programme?

9. Has your exercising changed since COVID19?

Prompts

- Yes/no? In what way? Have you decreased/increased exercise time? Have you changed the way you exercise?

Thank you.

Malta Journal of Health Sciences
https://doi.org/10.14614/LARYNXVMAT/12/25

Research Paper

Target delineation error assessment for patients treated to the larynx with VMAT: a quantitative study performed at a local Maltese radiotherapy department.

Ms Kristina Caruana
Department of Radiography, Faculty of Health Sciences, University of Malta
Sir Anthony Mamo Oncology Centre Tal-Qroqq, Msida, Malta
Ms Jessica Falzon
Sir Anthony Mamo Oncology Centre Tal-Qroqq, Msida
Dr Jose Guilherme Couto
Department of Radiography, Faculty of Health Sciences, University of Malta
Prof Francis Zarb
Department of Radiography, Faculty of Health Sciences, University of Malta
Prof Paul Bezzina
Department of Radiography, Faculty of Health Sciences, University of Malta

Abstract

Introduction: Target volume delineation error (Σ_{delin}) affects the accuracy of radiation therapy, but it is often not incorporated in studies measuring the Planning Target Volume (PTV) margin. The aim of this study was to assess the impact of Σ_{delin} on the PTV margin for patients receiving treatment with Volumetric-Modulated Arc Therapy (VMAT) to the larynx in a radiotherapy department.

Methods: Six clinical oncologists were asked to delineate the Clinical Target Volume (CTV) of five patients receiving treatment to the larynx with VMAT. Van Herk's formula was used to measure the PTV margin based on target volume delineation error. All data were collected and analysed to measure the Σ_{delin} and the PTV margin.

Correspondence to Kristina Caruana
kristina.caruana@gov.mt
Received: 30.3.2024
Revised: 30.9.2024
Accepted: 12.5.2024
Published: 30.6.2025

Results: The PTV margin based on Σ_{delin} was 5.6 mm, 11.1 mm, and 4.5 mm, in the left-right (X), superior-inferior (Y) and anterior-posterior (Z) directions respectively. The standard deviation (SD) in the X and Z direction was 3.47 mm, and the SD in the Y direction was 6.92 mm.

Conclusion: The PTV margin calculated based on the evaluation of Σ_{delin} was considerably larger than the margin currently used (5mm). If the resulting margin was implemented it would increase radiation induced side-effects.

Keywords: Σ_{delin}, Target volume delineation error, VMAT, Volumetric-Modulated Arc Therapy, larynx

1. Introduction

Delivery of radiotherapy within the head and neck region requires high levels of accuracy due to the large number of organs at risk within this region (Malicki, 2012). With the introduction of Volumetric-Modulated Arc Therapy (VMAT) in radiotherapy, higher conformal doses may be applied that require higher accuracy as slight deviations from the original plan may result in mistreatment (Nyarambi, Chamunyonga & Pearce, 2015).

There are many errors and variabilities that can impact the accuracy of treatment delivery, one of the most common sources of errors is target volume delineation variability (Bernstein *et al.*, 2021). Target volume delineation error (Σ_{delin}) is the variability in interpreting medical images and contouring protocols resulting in variation in the contouring of the radiotherapy targets (van Herk, Osorio, & Troost, 2019). As such, there is variation in the interpretation of what is considered target resulting in different patients receiving therapeutic doses to different regions depending on this interpretation. Incorrect Σ_{delin} can have a significant consequence, either lowering the probability of tumour control if part of the tumour is not contoured or unnecessary dose to normal tissue if additional areas of normal tissues are included within the target (Kristensen *et al.*, 2017).

According to The Royal College of Radiologists (2021), there are two types of variabilities: interobserver variation – variation among clinicians in delineating the target volume, and intra-observer variation – the mean of the margin outline drawn repeatedly by the same clinician.

Several studies (Bernstein *et al.*, 2021; Dewas *et al.*, 2011; Franco *et al.*, 2018) investigated differences in the delineation of tumour and normal tissues. However, the direct comparison of published data is difficult due to the use of a variety of methods to quantify the Σ_{delin}.

The only metric for delineation uncertainty that can be used to calculate the PTV margin is to quantify the Σ_{delin} (Bernstein *et al.*, 2021). Σ_{delin} occurs when the tumour is not properly contoured and this error is then carried over to the treatment phase. This error is considered a "systematic error" because it is created during the treatment planning and repeated in the same direction and magnitude in every treatment (The Royal College of Radiologists, 2021).

Although international studies have evaluated delineation variation for head and neck cancers, including the larynx, there is limited research specifically focusing on local departments. Each radiotherapy centre has unique procedures, protocols, and levels of experience, which can contribute to site-specific variations in target delineation (Lowther, Marsh, & Louwe, 2020). Therefore, it is essential to assess delineation accuracy and variability locally to determine if improvements can be made. By comparing local data to published international studies, this study aims to evaluate the interobserver variation in the delineation of laryngeal cancer at the local oncology department, where such data have not previously been collected. This comparison can help assess whether there is room for improvement in local practices and whether VMAT's dosimetric benefits are being fully utilised.

The aim of this study was to calculate Σ_{delin} in patients receiving VMAT treatment to the larynx in a local radiotherapy department and assess its impact on the PTV margin size. This research will also help identify potential areas for standardisation and improvement in delineation practices, both locally and in comparison to global findings.

2. Methods

2.1. Research design

This study had a quantitative and prospective research design. A quantitative approach was necessary since the aim was to quantify the variation in target contouring. The study had to be done prospectively since the hospital required the patient's consent to use their data, thus making it inappropriate to ask patients to provide their images once they finished treatment.

2.2. Patient selection

Patients treated for laryngeal cancer at the local oncology center between June 2021 and May 2022 were eligible for inclusion in the study. Five cases were randomly selected by the intermediary radiographer from patients' that were diagnosed with laryngeal cancer, regardless of the specific subsite (e.g., epiglottis, supraglottis, glottis, or subglottis) or lymph node involvement. Study commenced once ethical approval was obtained (UREC FORM V_15062020 8219)

2.3. Image acquisition and post-processing

All the patients had a contrast-enhanced computed tomography (CT) scan fused with a non-contrast CT scan during CT planning. Table 1 presents the specifications of the CT scanner that was used in the local department. Magnetic Resonance Imaging (MRI) diagnostic scans were also available to aid clinicians in target delineation. However, the MR images were not acquired in the radiotherapy treatment position, therefore the images were not geometrically comparable and could not be co-registered with the planning CT scan.

Table 1. CT Scanner Specifications of the Local Department						
Manufacturer	**Model**	**kV**	**mA**	**Scan rotation**	**Bore aperture**	**Slice thickness**
Canon	Aquilion LB	120	Auto mA (10–600 mA)	Helical	90 cm	2 mm

2.4. Selection of participants

All three oncology higher specialists' trainees (HSTs) and three oncologists specialised in head and neck radiotherapy at the local oncology hospital were invited to participate in this study. The participating clinicians were all considered competent in head and neck target contouring as the target delineation was done routinely as part of their normal work procedures, however all contours done by HSTs have to be peer-reviewed by an oncologist in clinical practice."

2.5. Data collection process

All six participants were asked to manually delineate the CTVs of the five selected patients. These contours were done on a Monaco® HD Treatment Planning System (TPS) (version 5.51) with which the clinicians were familiar to use in the clinical setting. In addition to the CT images, the oncologists and HSTs were also provided with a summary of the medical history to facilitate the delineation process. Since in the local oncology hospital where the study took place no specific delineation guidelines were used to delineate the target the clinicians were instructed to follow any guidelines depending on their preference.

2.6. Data analysis

The approach to evaluating CTV delineation variability was modified to account for outliers in the calculation (Lowther, Marsh, Louwe, 2020). This adjustment was implemented to ensure a more precise reflection of the delineation results achieved by the clinicians within the local department.

2.7. Ethical considerations

This study was approved by the local research ethics committee and written informed consent was obtained from all participants. The selected cases were retrieved and anonymised by an intermediary person at the radiotherapy department where the study was conducted as required by the ethical procedure established for this study

2.8. Data analysis.

All contours for each patient case were superimposed as a single structure set using the Monaco® HD TPS (version 5.51). Six perpendicular measurements were taken at specific points chosen to be visually representative of the variation around the contour on every alternating CT slice when all the clinicians' six delineated contours were visible. The contouring range was measured as the distance taken from the outer to the innermost superimposed contours in the left-right (X), superior-inferior (Y), and anterior-posterior (Z) directions. These measurements were taken on each alternating CT slice. The mean value of the measures was calculated in each direction to represent the data range of contour variation. This data was recorded and analysed using a Microsoft Excel datasheet that was previously validated and tested for reliability.

Data were analysed quantitatively using both descriptive and inferential statistics, with the results of the study being generalised to the target population.

Since the number of clinicians for each case was less than 15, the standard deviation (S) was calculated using the following equation as proposed by Tudor *et al.* (2020):

$$S = \frac{R}{d_2(N)}$$

Where R represents the data range and is calculated by measuring the distance between the inner and outermost contours in each image plane along each axis of interest at a representative point with 'average' observer variation. N represents the sample size, and d_2 is a tabled value that depends on the number of samples in the range. For a sample of six observers, the d_2 value is 2.53 (Tudor *et al.*, 2020).

The SD values from each case were combined by taking the mean value in each direction, representing the systematic error of target volume delineation.

The van Herk formula was used to calculate the PTV margin size based on $\sum_{delin}$. This formula is expressed as:

$M=2.5\sum + 0.7\sigma$.

"$\sum$" represents the SD for the population systematic errors, and "σ" represents the SD for the population random errors. Since $\sum_{delin}$ is purely a systematic error, there is no random error component.

The van Herk formula was considered adequate to achieve this study's goals based on a previous systematic literature review performed by this research team, evaluating the different methods to calculate PTV margins in head and neck cancer patients undergoing VMAT published by this research team (Caruana *et al.*, 2021).

3. Results

3.1. Characteristics of the patients

The characteristics of the patients are summarised in Table 2.

Table 2. Patients' Demographics

Patient' number	Sex	Age	Diagnosis	Tumour Location	Staging and/or Grading	Prescription
1	M	62	SCC	Left glottis	T1a N0 M0	5500cGy @275cGy in 20#
2	M	65	SCC	Right vocal cords and anterior commissure	T3 N0 M0	6600cGy @275cGy in 30#
3	F	82	SCC	Right vocal cord	T3 N0 M0	5500cGy @275cGy in 20#
4	F	45	SCC	Right vocal cord	T1a N0 M0	5500cGy @275cGy in 20#
5	M	81	SCC	Glottis and Subglottic involvement	T2 N0 M0	5500cGy @275cGy in 20#

SCC = Squamous Cell Carcinoma

F = Female
M = Male

3.2. Delineation variability

Figure 1 shows an example of measurements that were taken for patient study 2 in the coronal plane, respectively. The coronal plane was used to measure the superior and inferior distances, whilst the axial plane was used to measure the anterior, posterior, left and right distances.

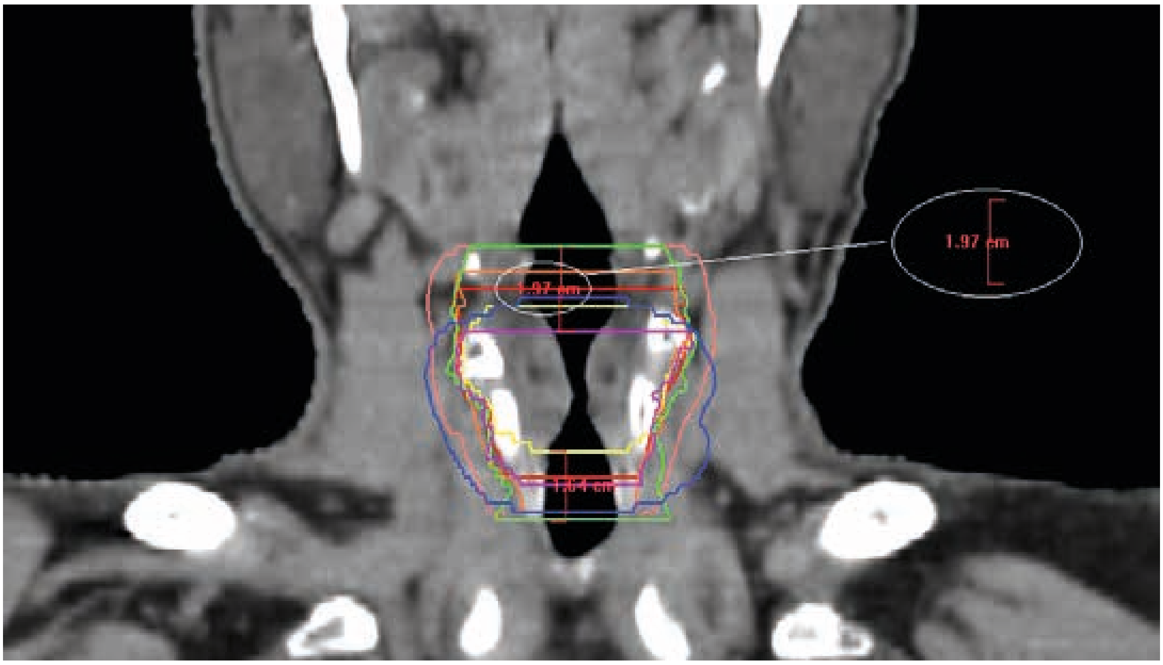
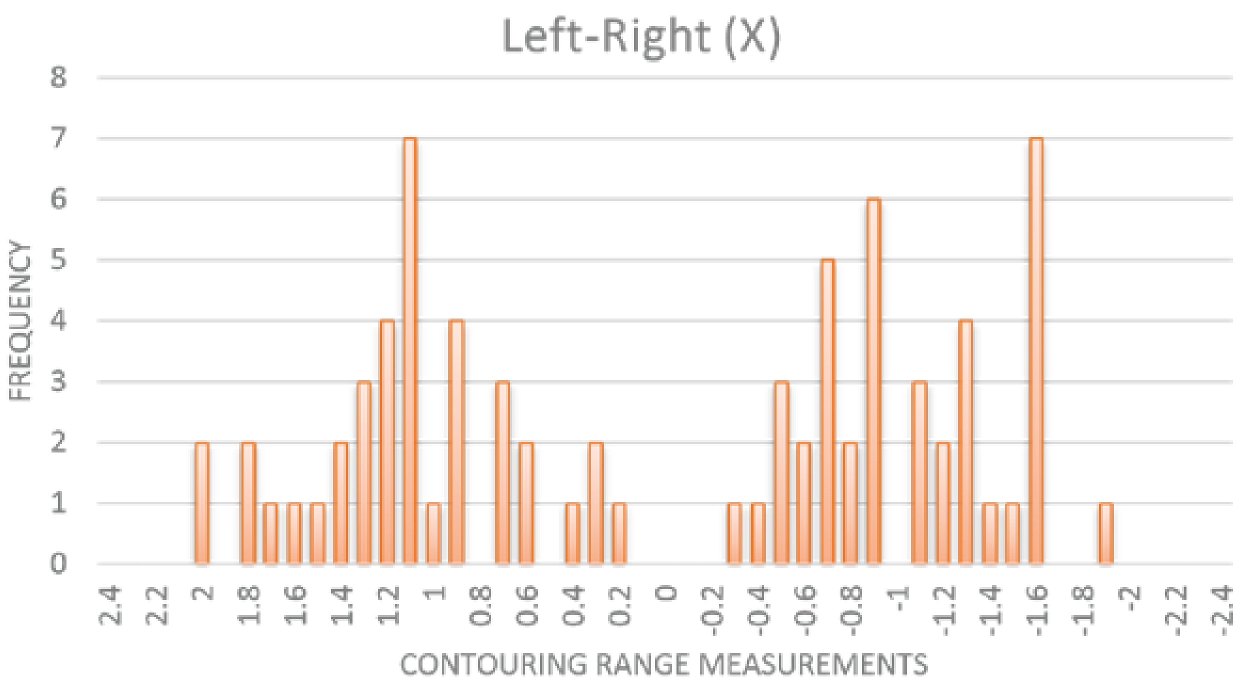
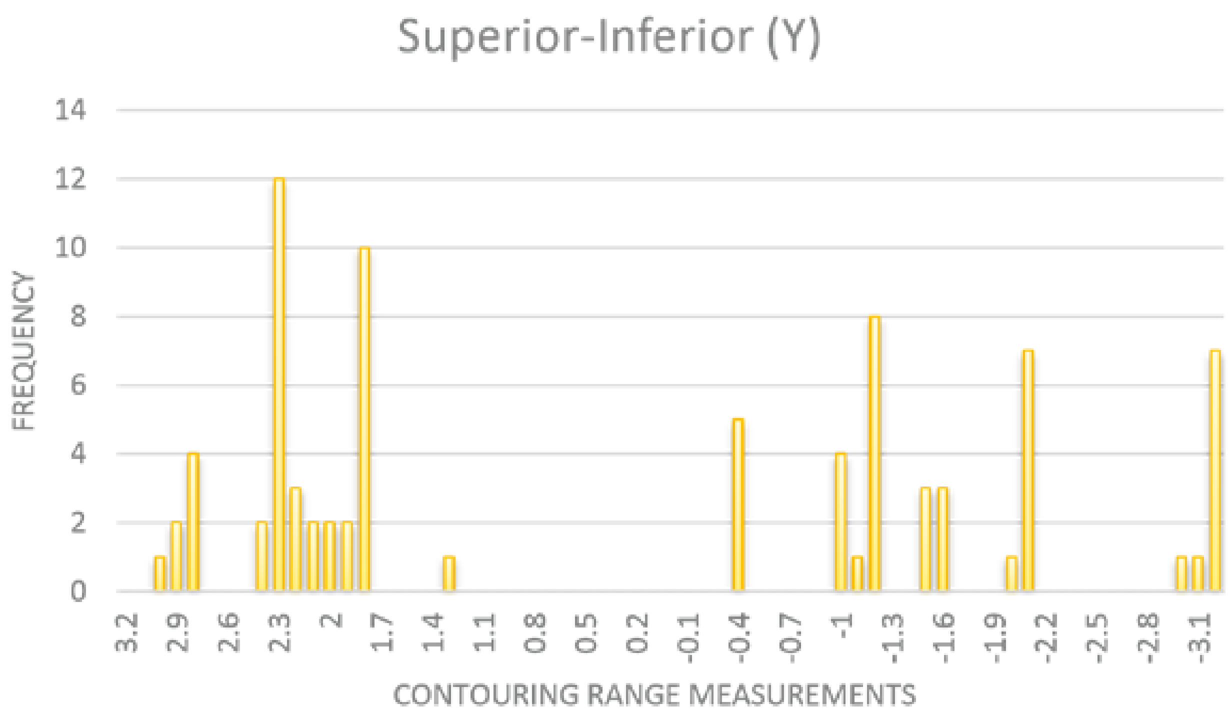


Figure 1: A Demonstration of the Superior and Inferior Measurements of Distances taken in the Coronal Plane.

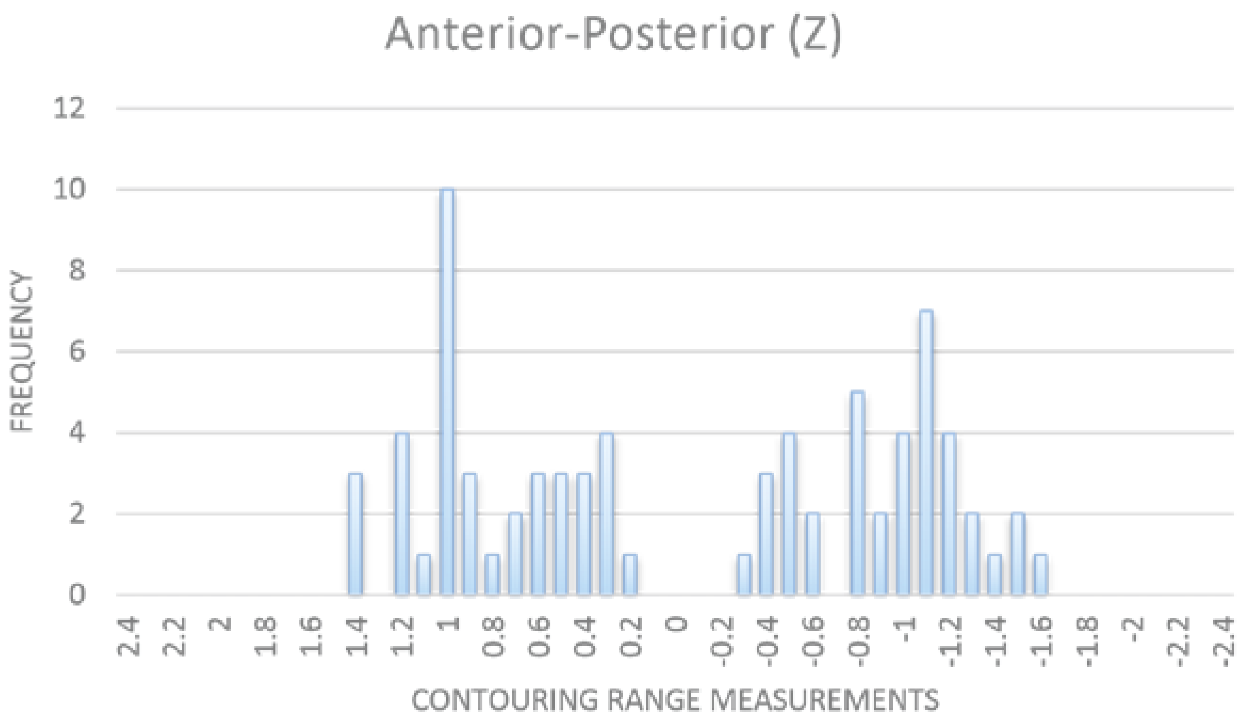
Frequency histograms were plotted to demonstrate the contouring range measurements, presented in Figure 2. a-c.



(a)


Superior-Inferior (Y)
FREQUENCY
14
12
10
8
6
4
2
0
3.2
2.9
2.6
2.3
2
1.7
1.4
1.1
0.8
0.5
0.2
-0.1
-0.4
-0.7
-1
-1.3
-1.6
-1.9
-2.2
-2.5
-2.8
-3.1
CONTOURING RANGE MEASUREMENTS

(b)


Anterior-Posterior (Z)
FREQUENCY
12
10
8
6
4
2
0
2.4
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1.2
-1.4
-1.6
-1.8
-2
-2.2
-2.4
CONTOURING RANGE MEASUREMENTS

(c)

Figure 2. Frequency histograms showing the distribution of contouring range measurements (cm) that were taken on each alternating CT slice for all patients in the sample. a) Left-Right (X) $\sum_{delin}$. b) Superior-Inferior (Y) $\sum_{delin}$. c) Anterior-Posterior (Z) ∑D.

The frequency histograms of the contouring range measurements in the X, Y, and Z axes indicated a bi-modal distribution. The left direction's mode value was 1.15 cm, whereas the right direction was -1.6 cm. The superior direction's mode value was 2.3 cm and the inferior direction's was -1.2 cm. The anterior direction's mode value was 1 cm, whereas the posterior value was -1.1 cm.

Table 3. shows the mean value of the contouring range measurements for each patient from the outer to the innermost superimposed contours, obtained at each translational axis.

Table 3. Systematic Individual Mean Values of the Contouring Range Measurements in each Translational Axis

	Left (mm)	Right (mm)	Superior (mm)	Inferior (mm)	Anterior (mm)	Posterior (mm)
Patient 1	11.9	8.3	28.2	9.6	4.5	12.4
Patient 2	8.2	10.0	19.8	16.3	8.1	5.9
Patient 3	7.6	9.9	22.3	11.6	9.4	10.6
Patient 4	4.6	6.0	21.8	4.4	3.5	5.4
Patient 5	13.9	14.6	17.0	24.0	9.5	11.4
Population means	9.2	9.8	21.8	13.2	7.0	9.1

The superior direction had the highest overall mean discrepancy variability of 21.8 mm, with patient 1 having the highest recorded average variability of 28.2 mm. The anterior direction had the smallest discrepancy overall, with a mean discrepancy variability of 7.0 mm, and with patient 4 having the smallest contouring range of 3.5 mm.

The CTVs for each of the six participating clinicians are shown in Figure 3.

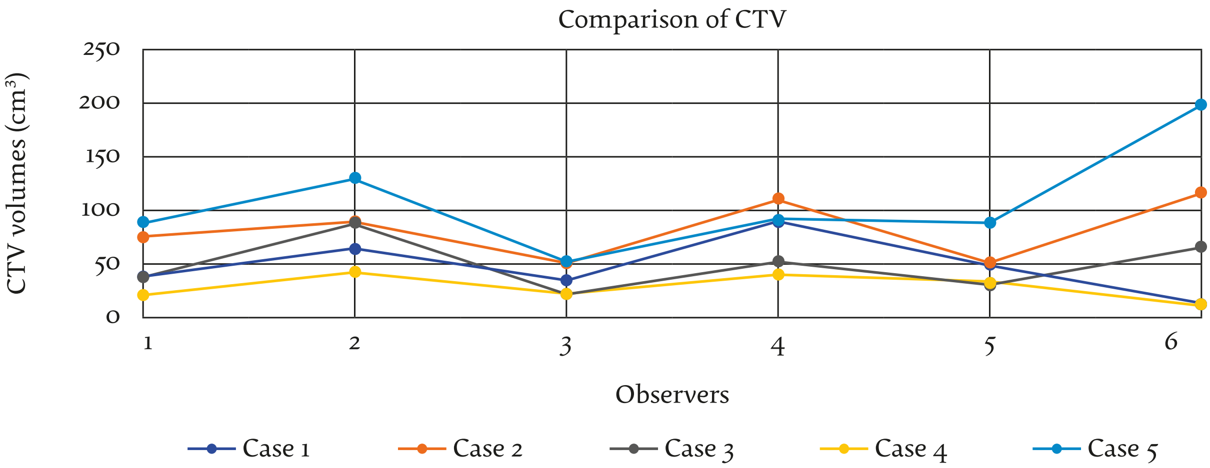


Figure 3. Graph showing the variation of the contoured clinical target volumes (cm3) obtained from the observers.

The mean distance between the outlines of six observers across all directions was 11.7 mm. However, there was an evident larger systematic variation in the Y direction when compared to the other directions. The population mean of the combined value of the Y direction was nearly twice as high as that obtained in the other directions, therefore as suggested by Tudor *et al.* (2020) the average range value was not considered as an isotropic Σ_{delin} and the variation for the Y direction was calculated separately from the X and Z directions.

In four out of five instances, among the group of six participants, observer number six was identified as the outlier by consistently generating either the largest or smallest CTVs when compared to the other clinicians.

The standard deviation for the X and Z directions was calculated as 8.8/2.53 = 3.47 mm. For the estimation of target delineation in the Y direction, an average value of the contouring range of both the superior and inferior directions resulted in a value of 17.5 mm. This value was divided by 2.53 and resulted in a target Σ_{delin} of 6.92 mm.

These values were then inserted into van Herk's formula for calculation of PTV. Table 4 displays the PTV margin result based only on Σ_{delin}.

Table 4. PTV Margin Result based on Σ_{delin}

PTV margin (mm)	X	Y	Z
Population Systematic Errors	3.47	6.92	3.47
PTV margin	5.6	11.1	4.5

X, left–right; Y, superior–inferior; Z, anterior–posterior; PTV, Planning Target Volume

4. Discussion

When compared to the other directions, the Σ_{delin} in the Y direction was more dispersed and less homogenous, indicating that the contouring range were the largest in these directions. This was also observed by Jager *et al.* (2015), where the greatest delineation volume discrepancies for the epiglottic region were observed in the Y direction (Jager *et al.*, 2015). This difficulty in defining the superior-inferior borders of the CTV is in part due to unclear boundaries in this direction in comparison with the lateral and anterior-posterior borders, which are also dependent on the tissue contrast and slice thickness of the CT. This, combined with the lack of local protocols, results in various oncologists adhering to different literature/guidelines (Freedman, 2015; Jager, 2017).

An unnecessary larger variation in delineation (in any direction) of the target would result in patients receiving unnecessary dose to normal tissues worsening radiotherapy side-effects. On the other hand, if the target is smaller than necessary, then the tumour control is compromised (Tudor *et al.*, 2020). The target should not be larger than what is necessary.

The PTV margin result, considering only the variability in CTV contouring, is of 5.6mm laterally, 11.1mm superior-inferiorly, and 4.5mm antero-posteriorly. This is a larger margin than the one currently used in the local department (Taliana *et al.*, 2020). Instead of recommendations of increasing the local CTV-PTV margin, it is recommended that the origin of this systematic error is addressed. Below are some of the recommendations to address this issue.

4.1. Imaging quality

The accuracy of delineation is hindered if imaging modalities have a low resolution (Kristensen *et al.*, 2017) and according to Mercieca, Belderbos and Van Herk (2018), observation variation is reduced when superimposing CT planning scan with MRI. Simple measures, such as intravenous and/or intracavitary contrast and reproducible imaging protocols could significantly improve imaging quality. When contouring, the use of zoom levels, simultaneous viewing in multiple planes (sagittal and coronal planes), and adequate window level and window width settings on the planning CT reduce inter-observer variability (Segedin and Petric, 2016). In this study, all patients had a contrast scan fused with a non-contrast scan during CT planning. MRI diagnostic scans were made available to assist clinicians in target delineation, though their use was at the discretion of the individual clinician. As previously mentioned, the MR images were not acquired in the radiotherapy treatment position, which prevented accurate superimposition onto the planning CT scan. Other available information was also provided to the clinicians, but its use for delineation was not formally assessed. Schmidt and Payne (2015) had reported on this, and they were of the opinion that this limited the benefits of delineating with MRI as different observers may rely on different images to perform their delineation.

4.2. Standardised protocols/guidelines

International guidelines, such as those of the Radiation Therapy Oncology Group, Danish Head and Neck Cancer Group, European Organisation for Research and Treatment of Cancer and guidelines written by Grégoire *et al.* (2018) may be used by clinicians for CTV delineation. Studies have shown that the introduction of site-specific anatomical atlases could reduce variability between observers in various tumour sites (Kim *et al.*, 2019). There was no specific guideline which clinicians could follow, when different or ambiguous guidelines are used for target volume delineation, this will have a significant impact on the consistency of delineated structures (Mercieca, Belderbos and Van Herk, 2018). According to Tudor *et al.* (2020), outliers should not be considered when measuring the range measurement of contouring because these contours would be inconsistent with clinical protocols, and one should not attempt to correct for major differences in opinion of the target volume. For this study, outliers were still considered as the department did not follow a specific clinical protocol regarding target delineation of the CTV volume for laryngeal cancer and all participants performed head and neck contouring in clinical practice.

Since one of the clinicians was an outlier and was still considered, the margin size may have been larger than necessary for most patients, because the other clinicians' delineations were closer to each other. This could have also been the reason for the large Σ_{delin} obtained in this study.

Another reason for the large variability could be attributed to the consideration of organ motion in the CTV delineation instead of considering organ motion as part of the PTV delineation error. According to *ICRU Reports 50 and 62*, set-up and organ motion errors should be integrated into the treatment planning procedure by taking a margin around the CTV, consequently defining the PTV. However, the Northern Association of Clinical Physics recommendation was to create a separate margin for set-up errors and organ motion (Van Herk, 2003). These different thoughts result in the lack of standardisation of margin calculation

4.3. Specialised training

Having a diverse group of clinicians with varying roles and experiences was an accurate representation of the local department, however during the clinical settings the contours performed by HSTs would have been checked by the clinical oncologists. The HSTs in the local department rotate according to different treatment areas and this may cause inconsistencies in target delineation (Tudor *et al.*, 2020).

Some publications have addressed the issue of training in target delineation, for example, Schimek-Jasch *et al.* (2015) reported that after a teaching session at a study group meeting, there was an improvement in overall inter-observer agreement, as evidenced by a reduction in target volumes. Khoo *et al.* (2012) obtained similar results and were of the opinion that a well-structured education programme reduced both inter – and intra-observer prostate contouring variations. In contrast, Dewas *et al.* (2011) reported no improvement among clinicians following a teaching course and believed that the reason for this could have been the high standard of the initial delineations. Furthermore, the authors had noted that several clinicians had discussed with each other to reach an agreement about the volumes that needed to be treated within their groups. This could have explained the homogeneity and high quality of the contours reported by Dewas *et al.* (2011), and shows the importance of having target delineation checked by other clinicians who are experts in the field.

This study has several limitations that must be acknowledged. The diversity of the cases, ranging from T1a to T3 tumor staging, may have contributed to some of the observed variations. Additionally, intra-observer variation was not assessed, as the study focused primarily on inter-observer variability among experienced clinicians, which we considered more relevant to our objectives. However, evaluating intra-observer consistency could offer valuable insights into the reproducibility of delineation practices and should be explored in future research. Another limitation is the small sample size, with only five cases included, which restricts the generalisability of the findings. This decision was made to minimise the risk of attrition among participating oncologists if the case load were increased. Future studies should aim to replicate this research with a larger sample. Furthermore, as previously mentioned, the impact of the oncologists' experience on contouring was not evaluated. It is recommended that future studies assess whether seniority or experience influences variability in contouring.

Another recommendation for future studies is to evaluate if training, local protocols, or improvements in imaging would improve the variability and reduce the

PTV margin. The results of such future studies could be compared with the current research findings.

5. Conclusion

The results obtained from this study clearly indicate that Σ_{delin} may have a considerable impact on the PTV margin, particularly in the superior direction. Therefore, it is crucial to reduce the variability associated with Σ_{delin} to achieve better treatment outcomes and radiotherapy-induced toxicities.

In this study all clinicians who contour the head and neck region, irrespective of experience, participated in the study, however in clinical practice the contours performed by HSTs would have been checked by the most experienced clinical oncologists. This could have been one of the reasons of the resulting large PTV margin.

Several recommendations can help minimise target Σ_{delin}. It is recommended that clinical contouring protocols are implemented locally as this will reduce the impact of target Σ_{delin}. It is also advised that clinicians receive contouring training and follow published contouring guidelines. Further studies to measure target Σ_{delin} for assessing PTV margin are also recommended as most studies that assessed Σ_{delin} were not done with the scope of measuring the delineation variability to calculate the PTV margin size.

References

Bernstein, D., Taylor, A., Nill, S., Imseeh, G., Kothari, G., Llewelyn, M., *et al.* (2021). 'An inter-observer study to determine radiotherapy planning target volumes for recurrent gynaecological cancer comparing magnetic resonance imaging only with computed tomography-magnetic resonance imaging', *Clinical Oncology*, 33(5), pp. 307–313. doi:10.1016/j.clon.2021.02.003.

Caruana, K., Refalo, N., Spiteri, D., Couto, J. G., Zarb, F. & Bezzina, P. (2021). 'PTV margin calculation for head and neck patients treated with VMAT: A systematic literature review', *Journal of Radiotherapy in Practice*, 21(4), pp. 586–593. doi:10.1017/S1460396921000546.

Choi, M., Refaat, T., Lester, M. S., Bacchus, I., Rademaker, A. W. & Mittal, B. B. (2014). 'Development of a standardized method for contouring the larynx and its substructures', *Radiation Oncology*, 9, p. 285. doi:10.1186/s13014-014-0285-4.

Dewas, S., Bibault, J., Blanchard, P., Vautravers-Dewas, C., Pointreau, Y., Denis, F., *et al.* (2011). 'Delineation in thoracic oncology: a prospective study of the effect of training on contour variability and dosimetric consequences', *Radiation Oncology*, 6(1), p. 118. doi:10.1186/1748-717X-6-118.

Franco, P., Arcadipane, F., Trino, E., Gallio, E., Martini, S., Iorio, G. C., Piva, C., Moretto, F., Ruo Redda, M. G. & Verna, R. (2018). 'Variability of clinical target volume delineation for rectal cancer patients planned for neoadjuvant radiotherapy with the aid of the platform Anatom-e', *Clinical and Translational Radiation Oncology*, 11, pp. 33–39. doi:10.1016/j.ctro.2018.06.002.

Freedman, A. (2015). 'A radiation oncologist's guide to contouring the larynx', *Practical Radiation Oncology*. doi:10.1016/j.prro.2015.10.007.

Grégoire, V., Evans, M., Le, Q.T., Bourhis, J., Budach, V., Chen, A., Eisbruch, A., Feng, M., Giralt, J., Gupta, T., Hamoir, M., Helito, J.K., Hu, C., et al., (2018). Delineation of the primary tumour Clinical Target Volumes (CTV-P) in laryngeal, hypopharyngeal, oropharyngeal and oral cavity squamous cell carcinoma: AIRO, CACA, DAHANCA, EORTC, GEORCC, GORTEC, HKNPCSG, HNCIG, IAG-KHT, LPRHHT, NCIC CTG, NCRI, NRG Oncology, PHNS, SBRT, SOMERA, SRO, SSHNO, TROG consensus guidelines. *Radiotherapy and Oncology*, 126(1), pp.3–24. Available at: https://doi.org/10.1016/j.radonc.2017.10.016 [Accessed 29 Sept. 2024

Jager, E. A., Kasperts, N., Caldas Magalhaes, J., Philippens, M. E. P., Pameijer, F. A., Terhaard, C. H. J., *et al.* (2015). 'GTV delineation in supraglottic laryngeal carcinoma: interobserver agreement of CT versus CT-MR delineation', *Radiation Oncology*, 10(1), p. 26. doi:10.1186/s13014-014-0321-4.

Jager, E. A. (2017). *Towards accurate target delineation for head and neck cancer*. UMC Repository (Dissertation). Supervisor(s): Terhaard, C.H.J., Raaijmakers, C.P.J., Philippens, Marielle.

Khoo, E. L. H., Schick, K., Plank, A. W., Poulsen, M., Wong, W. W. G., Middleton, M. & Mittal, B. B. (2012). 'Prostate contouring variation: Can it be fixed?', *International Journal of Radiation Oncology, Biology, Physics*, 82(5), pp. 1923–1929. doi:10.1016/j.ijrobp.2011.02.050.

Kim, S. H., Oh, S. A., Yea, J. W. & Park, J. W. (2019). 'Prospective assessment of inter – or intra-fractional variation according to body weight or volume change in patients with head and neck cancer undergoing radiotherapy', *PLOS ONE*, 14(5), p. e0216655. doi:10.1371/journal.pone.0216655.

Kristensen, I., Nilsson, K., Agrup, M., Belfrage, K., Embring, A. & Haugen, H., *et al.* (2017). 'A dose based approach for evaluation of inter-observer variations in target delineation', *Technical Innovations & Patient Support in Radiation Oncology*, 3–4, pp. 41–47. doi:10.1016/j.tipsro.2017.10.002.

Lowther, N., Marsh, S. & Louwe, R. (2020). 'Dose accumulation to assess the validity of treatment plans with reduced margins in radiotherapy of head and neck cancer', *Physics and Imaging in Radiation Oncology*, 14, pp. 53–60. doi:10.1016/j.phro.2020.05.004.

Malicki, J. (2012). 'The importance of accurate treatment planning, delivery, and dose verification', *Reports of Practical Oncology and Radiotherapy*, 17(2), pp. 63–65. doi:10.1016/j.rpor.2012.02.001.

Mercieca, S., Belderbos, J. S. A. & van Herk, M. (2021). 'Challenges in the target volume definition of lung cancer radiotherapy', *Translational Lung Cancer Research*, 10(4), pp. 1983–1998. doi:10.21037/tlcr-20-627.

Muijs, D. (2011). *Doing quantitative research in education with SPSS.*

Nyarambi, I., Chamunyonga, C. & Pearce, A. (2015). 'CBCT image guidance in head and neck irradiation: the impact of daily and weekly imaging protocols', *Journal of Radiotherapy in Practice*, 14(4), pp. 362–369. doi:10.1017/S1460396915000266.

Taliana, K., Bezzina, M., Muscat, K. & Bartolo, C. (2020). *Clinical protocol: Radiotherapy for head and neck cancer*. Sir Anthony Mamo Oncology Centre.

The Royal College of Radiologists (2021). *On Target 2: Updated guidance for image-guided radiotherapy*.

Tudor, G., Bernstein, D., Riley, S., Rimmer, Y., Thomas, S., Herk, M., *et al.* (2020). 'Geometric uncertainties in daily online IGRT: Refining the CTV-PTV margin for contemporary radiotherapy', *Radiology and Oncology*, 50(3), pp. 254–262. doi:10.1515/raon-2016–0023.

Schimek-Jasch, T., Troost, E. G. C., Rücker, G., Prokic, V., Avlar, M., Duncker-Rohr, V., *et al.* (2015). 'A teaching intervention in a contouring dummy run improved target volume delineation in locally advanced non-small cell lung cancer', *Strahlentherapie Und Onkologie*, 191(6), pp. 525–533. doi:10.1007/s00066–015–0812–8.

Schmidt, M. A. & Payne, G. S. (2015). 'Radiotherapy planning using MRI', *Physics in Medicine & Biology*, 60(22), pp. R323-R361. doi:10.1088/0031–9155/60/22/R323.

van Herk, M. (2004). 'Errors and margins in radiotherapy', *Seminars in Radiation Oncology*, 14(1), pp. 52–64. doi:10.1053/j.semradonc.2003.10.003.

Van Herk, M., Osorio, E. V. & Troost, E. G. C. (2019). 'Is reducing irradiated margins key to improving outcomes for radiotherapy?', *The Lancet Oncology*, 20(9), pp. 1208–1210. doi:10.1016/S1470–2045(19)30539-X.

Malta Journal of Health Sciences
https://doi.org/10.14614/PROGNOSTIC/12/25

Prognostic models of non–surgical treatment outcomes for lumbar–related leg pain: a scoping review of systematic reviews.

Mr Emanuel Schembri
Department of Physiotherapy, Faculty of Health Sciences, University of Malta.
Dr Karl Spiteri
Department of Physiotherapy, St Vincent de Paul, Malta.
Dr Laura Jayne Bonnett
Department of Health Data Science, University of Liverpool.
Dr John Xerri de Caro
Department of Physiotherapy, Faculty of Health Sciences, University of Malta.

Abstract

Prognosis identifies a relationship between future outcomes and a given health state. Prognostic factors help estimate the likelihood of a particular outcome, regardless of any specific intervention. The prognosis of persons treated for lumbar–related leg pain (LRLP) varies considerably. To the authors' knowledge, no systematic review (SR) has evaluated any formal combination of multiple predictors (prognostic models) for non–surgical interventions in this population. This scoping review aimed to: (a) determine if SRs specifically evaluating prognostic models of non–surgical interventions for LRLP exist, and (b) identify prognostic models addressing non–surgical interventions in LRLP within the SRs. A systematic search of PubMed, Embase, CINAHL, and Epistemonikos was conducted in October 2024, including SRs published from the year 2000 onwards. References were hand–searched, and forward citation searches were performed. The search identified 9,398 records after deduplication. Following screening, the full texts of 18 SRs were evaluated against the eligibility criteria. The lack of SRs in this area highlights a critical gap in understanding the prognostic models for non–surgical interventions in LRLP. Three reviews reported two primary studies that had derived and/or evaluated prognostic models in cohorts that met the threshold set for this scoping review, with 80% of participants experiencing LRLP. This scoping review highlights a gap in the SR literature evaluating specific prognostic models for non–surgical interventions in individuals with LRLP even though various prognostic models for LRLP exist. This underlines the need for a dedicated SR to consolidate evidence and guide clinical practice in the non–surgical management of LRLP.

Keywords: Sciatica; Radiculopathy; Models, Statistical; Predictive model; Scoping review.

Correspondence to Emanuel Schembri
esche05@um.edu.mt
Received: 19.12.2024
Revised: 19.2.2025
Accepted: 8.3.2025
Published: 30.6.2025

1. Introduction

Acute episodes of referred leg pain that originate from the lumbar region normally resolve within a few weeks as a consequence of the natural course of healing. However, outcomes appear to be poorer than those episodes of low back pain presenting without leg pain (Konstantinou *et al.*, 2018). Studies have reported that following the conservative management of persons with lumbar radicular pain, only 35.7% had a good outcome at 12 months (Azharuddin *et al.*, 2022) and 42% needed surgery after one year (Boden *et al.*, 2018). 70% of the patients referred with symptomatic lumbar spinal stenosis reportedly remained the same or worsened or required surgical treatment after three years (Matsudaira *et al.*, 2016).

Episodes that present as referred pain along the leg are often labelled as sciatica, giving the impression that sciatica is more than a symptom. Definitions for sciatica vary with distinctions being reported, for example, between self-reported and clinically assessed symptoms, which leads to a poor interpretation of clinical outcomes (Konstantinou & Dunn, 2008). Sciatica is often confused with other sources of somatic referred leg pain. This confusion contributes to wide discrepancies in prevalence rates (1.6 to 43%) and complicates effective patient management (Konstantinou & Dunn, 2008). Consequently, the International Association for the Study of Pain (IASP) NeuPSIG recommends avoiding the term 'sciatica' in favour of 'spine-related leg pain' as a more precise umbrella term that addresses both somatic referred pain and radicular pain with or without radiculopathy (Schmid *et al.*, 2023). In this present article, we adopt the term lumbar-related leg pain (LRLP) to specifically focus on the lumbar spine.

Prognosis research investigates the relations between future outcomes (endpoints) among people with a given health state (start point) to improve health. The PROGnosis RESearch Strategy (PROGRESS) series is a framework of four distinct but inter-related prognosis research themes being: overall prognosis, prognostic factors, prognostic models and stratified medicine. The PROGRESS series aims to explain how each of these four themes provides evidence that can be used at multiple (translational) pathways toward improving clinical outcomes—from the discovery of new interventions, through to their evaluation and implementation in the clinical management of individual patients, and to examine the impact of interventions and healthcare policies on patient outcomes (Hemingway *et al.*, 2013; Riley *et al.*, 2013; Steyerberg *et al.*, 2013; Hingorani *et al.*, 2013).

Multivariable prediction models are categorised into diagnostic and prognostic models based on their purpose and temporal focus. Diagnostic models combine multiple predictors, often diagnostic test results, to estimate the probability of a disease or condition being present or absent at the time of prediction (Moons *et al.*, 2015). Prognostic models integrate multiple predictors to estimate the probability of a specific outcome or event occurring in the future. These models apply to individuals at risk of that outcome, whether healthy or ill. They can forecast events such as recurrence, complications, or mortality within a defined timeframe, including in individuals without a diagnosed disease (D'Agostino *et al.*, 2008). The term prognostic is used here in a broader sense, encompassing predicting future outcomes in at-risk populations rather than being limited to forecasting the progression of patients with a specific disease, regardless of treatment. The key distinction between diagnostic and prognostic models is time: diagnostic models assess the present and are developed through a cross-sectional research design. Prognostic models forecast future outcomes and are usually longitudinal (Moons *et al.*, 2015).

The STarT Back tool (SBT) is a clinical prediction tool in the form of a questionnaire that assesses the likelihood of disability and chronicity in six months' time in a predominantly acute cohort, and it consists of nine items that screen for physical and psychological predictors of persistent disabling low back pain (LBP). The binary scores are added, and the total score stratifies responses into three pre-defined subgroups: low-risk, medium-risk, and high-risk of persistent disabling LBP. Each risk subgroup has a pre-defined treatment matched to it. This is recommended for the stratification of care pathways in the UK primary care setting (Hill *et al.*, 2011). The SBT was a considerable step in the personalised care approach for patients with LBP.

The NICE guidelines on LBP and sciatica (NICE, 2020a) included a reference to the adoption of the SBT with a disclaimer stating that the quality and usability of the tool had not yet been judged (NICE, 2020b). The inclusion of exclusively modifiable prognostic factors in the SBT does not provide a 360° approach to making accurate predictions since non-modifiable factors also have a predictive role (Parreira *et al.*, 2018). The value of a screening instrument is also directly related to setting-specific conditions and optimal in the cohorts for which it was developed (Karran *et al.*, 2017a). Studies have reported that the final outcomes for managing LBP following the SBT approach were superior in the UK primary care settings in which the SBT was developed and validated, compared to other settings that adopted the SBT approach, such as physiotherapy and chiropractic settings, as well as for secondary levels of care in which the SBT was not validated (Morsø *et al.*, 2014; Karran *et al.*, 2017b). The predictive value of the SBT also did not prove advantageous in persons suffering from chronic low back pain (Kendell *et al.*, 2018), most likely because the tool was developed for a cohort largely composed of acute cases of LBP (72–74% of the participants) (Hill *et al.*, 2011).

To the authors' knowledge, no systematic review (SR) evaluated predictive models of non–surgical interventions specifically for individuals with LRLP. The primary aim of this scoping review was to: (a) determine if SRs specifically evaluating prognostic models of non–surgical interventions for LRLP exist, and (b) to identify prognostic models addressing non–surgical interventions in LRLP within the SRs.

Therefore, this scoping review aims to answer the primary research question: Are there SRs that evaluate prognostic models of non–surgical interventions specifically related to LRLP? The secondary research question was: Are there prognostic models addressing non–surgical interventions for individuals with LRLP?

2. Methods

2.1. Reason for conducting a scoping review

This scoping review aimed to determine if the literature already reported a SR on the evaluation of prognostic models of treatment outcomes in the non–surgical management of LRLP. The broader research aims, and the exploratory nature of the envisioned literature review necessitated a methodology that would map and summarise the evidence and extent of knowledge to better inform further research (Peters *et al.*, 2020). Ensuring that a SR has not already been conducted on the topic helps prevent unnecessary duplication of efforts. For this reason, a scoping review was considered particularly suitable for exploring this aim since such reviews are exploratory projects that systematically map the literature available on a topic, identifying the key concepts, theories, sources of evidence, and gaps in the research (Canadian Institutes of Health Research).

2.2. Eligibility criteria

This scoping review included SRs (with or without meta–analyses) that were published in the English language, which evaluated a population composed of adults over 18 years reporting LBP, specifically requiring at least 80% of the study cohort to have LRLP, and which may include sciatica, neurogenic claudication, radicular pain, or a combination thereof. Studies were excluded if less than 80% of participants had LRLP or if pain originated from other causes, such as tumours or fractures (Table 1). The 80% threshold was designed in a pragmatic manner to limit the contamination of the sample with other types of LBP.

The reviews had to include studies that evaluated prognostic models of non–surgical outcomes, including medication, physiotherapy, epidural steroid injections, and multidisciplinary rehabilitation. Studies addressing models predicting the outcomes following surgical interventions or models predicting pain chronicity or disability unrelated to a specific treatment were excluded.

This scoping review aimed to identify prognostic models that adhere to the definition provided within TRIPOD (Moons *et al.*, 2015, p. 1), where a prognostic model is defined as the "mathematical equation that estimates the probability of having a disease or condition in the present (diagnostic prediction model) or the probability of developing a particular disease or outcome in the future (prognostic prediction model)". Therefore, for continuous outcomes, a prognostic model predicts an individualised expected (mean) outcome value by a particular time point. For binary or time–to–event outcomes, a prognostic model predicts an individual outcome risk (probability) by a particular time point (or time points) (Steyerberg *et al.*, 2013).

Due to the studies' methodological heterogeneity, especially in the early years of prognostic medicine, clinical prediction rules (CPRs) and screening tools for stratifying patients into meaningful prognostic subgroups will also be eligible in this review (Fu *et al.*, 2024). CPRs are simple statistical prediction tools designed to be used with individual patients. They comprise a small number of clinical variables that have been identified to be independently predictive of a given diagnosis, outcome, or treatment effect. These can be both prognostic and prescriptive CPRs, where the former consists of prognostic variables that inform predictions of future outcomes while the latter is a special type of prognostic CPRs since they inform predictions regarding the relative treatment effect a patient may experience from an intervention (Foster *et al.*, 2013). The variables included in a prescriptive CPR are treatment effect modifiers (Kraemer, Frank & Kupfer, 2006); thus, they function to inform clinical decisions regarding treatment selection (Cook, 2008).

Any identified prognostic model's predictive ability, including accuracy, calibration, discrimination, net benefit, and R^2 values were extracted (PROGRESS 3). If any of the prognostic models were evaluated within a

Table 1. PIOS criteria for inclusion in the review.

	Inclusion	Exclusion
Population	Adults (>18 years) with low back pain† disorders, having at least 80% of the participants with lumbar–related leg pain (including but not limited to sciatica or neurogenic claudication or radicular pain or a mixture of the three).	Less than 80% of the cohort having lumbar–related leg pain. Back pain or lumbar–related leg pain due to other conditions, e.g., tumour or fracture.
Intervention	Prognostic models of treatment outcomes for lumbar–related leg pain following non–surgical treatments (for example, but not limited to, medications, physiotherapy, epidural steroid injections, and multidisciplinary rehabilitation).	Models predicting outcomes following spinal surgeries or other conditions. Models/factors predicting disability or pain chronicity but not associated with any formal treatment/ intervention.
Outcome	Predictive abilities of the prognostic models, including calibration, discrimination, net benefit, and R^2 values. Measures of benefit in case a model was evaluated within an RCT (PROGRESS level 4).	The study solely focuses on treatment outcomes without evaluating the predictive ability of the models.
Study design	**Research question 1:** Systematic reviews with or without meta–analyses of prognostic model studies. **Research question 2:** Primary studies included within the systematic reviews screened by full text, adopting either an observational (cohort, registry or retrospective studies) or RCT design that also fulfils the first three aspects of the PIO criteria.	All other study designs, reviews that do not have an English language version.

†Low back pain was included since studies frequently evaluate back pain and lumbar–related leg pain together. However, to be eligible for inclusion, a prognostic model developed or validated in a mixed population must have at least 80% of the cohort composed of persons with lumbar–related leg pain.

randomised controlled trial (RCT) (PROGRESS 4), the outcomes were briefly summarised. Studies without predictive assessments were excluded.

2.3. Critical appraisal

Scoping reviews usually do not necessitate critically appraising the included articles (Peters *et al.*, 2020). However, this review incorporated an appraisal process for specific cases. If a SR examining prognostic models of treatment outcomes for LRLP was identified, it was critically appraised using the AMSTAR–II tool (Shea *et al.*, 2017). Similarly, if a prognostic model of treatment outcomes for persons with LRLP was identified, it was appraised using the PROBAST tool (Wolff *et al.*, 2019). Two reviewers (ES, KS) independently conducted the critical appraisal, and any discrepancies were resolved through consensus.

2.4. Information sources and search strategy

The scoping review was carried out following the PRISMA for scoping review guidelines (PRISMA–ScR) (Tricco *et al.*, 2018). The search strategy was developed by ES and checked by an academic librarian at the University using the PRESS checklist (McGowan *et al.*, 2016). An electronic literature search was conducted on the 7th of October 2024 within PubMed, Embase (Elsevier), CINAHL (EBSCO), and Epistemonikos (Goossen *et al.*, 2020). References of the included reviews were hand-searched, and forward citation searching using the software CitationChaser (Haddaway, Grainger & Gray, 2022) was conducted. Due to the inconsistent reporting standards, reliability, and data adequacy often associated with grey literature, this was not searched.

The focus of the search was on LRLP. However, terms related to LBP were included in the search since both conditions are frequently studied together. Three main facets were incorporated into the search strategy: LRLP, LBP, and a sensitive search filter for prognostic/ prediction models (Appendix A).

A combination of free-text keywords, their synonyms and, where appropriate, word truncation was employed in the search strategy. Furthermore, any relevant medical subject headings (MeSH terms) or Emtree terms were incorporated into the search. The first two facets were combined using the Boolean operator "OR" and subsequently joined with the search filter using the Boolean operator "AND". The year of publication was limited to the period 2000–2024, reflecting the publication timeline of the Orebro Musculoskeletal Questionnaire, a prognostic tool for back pain disorders, which was published in 2003 (Linton & Boersma, 2003).

2.5. Search Filters

To retrieve prognostic model studies within the three databases, validated search filters that were specifically designed to retrieve prognostic studies within CINAHL (Walker-Dilks, Wilczynski & Haynes, 2008), Embase (Holland, Wilczynski & Haynes, 2005) and PubMed (Geersing *et al.*, 2012) were used within the respective search strings (Appendix A). PubMed was preferred over other platforms, such as Medline, due to the unique aspects of the PubMed search engine. The Geersing *et al.*, (2012) search filter yielded higher sensitivity values [(0.97; 95% CI 0.83 to 0.99) (0.94; 95% CI 0.74 to 0.99)], and the lowest number needed to read (NNR = 68 to 125) values to justify it.

2.6. Screening and data charting

Two reviewers (ES, KS) collaboratively designed a data-charting form to identify the variables for extraction. The same reviewers independently screened the retrieved titles against the eligibility criteria. Liaison with a third reviewer (JXDC) was sought in case of disagreements. The same reviewers independently extracted data from the identified articles using a standardised data extraction sheet. In case a SR identified a model that was specifically developed, validated or tested in a cohort of persons with LRLP, the authors also sought related papers on its development, validation and/or updating to aid data extraction and critical appraisal. In such cases, the CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS) checklist (Moons *et al.*, 2014) was used to extract data. In cases of disagreement or conflict during the data extraction phase, a third reviewer (JXDC) was consulted to achieve consensus through discussion.

3. Results

3.1. Selection of sources of evidence

Epistemonikos (n=907), CINAHL (EBSCO) (n=1,522), PubMed (n=1,720), and Embase (Elsevier) (n=7,762) were searched on October 7, 2024, for a total of 11,911 records. After deduplication, the number of records was 9,398, which were screened by their title and abstract (Figure 1). The full text of 18 SRs was retrieved and screened against the eligibility criteria for both research questions 1 and 2.

3.2. Research Question 1: Are there systematic reviews explicitly evaluating prognostic models of non-surgical treatment outcomes for lumbar-related leg pain?

None of the SRs screened by their full text explicitly evaluated prognostic models of treatment outcomes for persons with LRLP. Table 2 presents the reasons for exclusion for each of these reviews. Further elaboration on the reasons for exclusion is presented in Appendix B. Most of the reviews included persons with LRLP but either failed to provide the percentage of this cohort within the total number of participants (Silva *et al.*, 2022; Karran *et al.*, 2017a), included less than 80% of the participants with LRLP (May & Rosedale, 2009; Kent & Keating, 2008; Haskins, Osmotherly & Rivett, 2015; Tagliaferri *et al.*, 2022; Ogbeivor & Elsabbagh, 2021), or both (Haskins, Rivett & Osmotherly, 2012; Stanton *et al.*, 2010), or did not include persons with LRLP (Fu *et al.*, 2024; Lheureux & Berquin, 2019; McIntosh *et al.*, 2018; Patel *et al.*, 2013; Almas, Parsons & Whalen, 2018). Therefore, this scoping review has determined that there are no SRs that evaluate prognostic models of non-surgical interventions, specifically in persons suffering from LRLP.

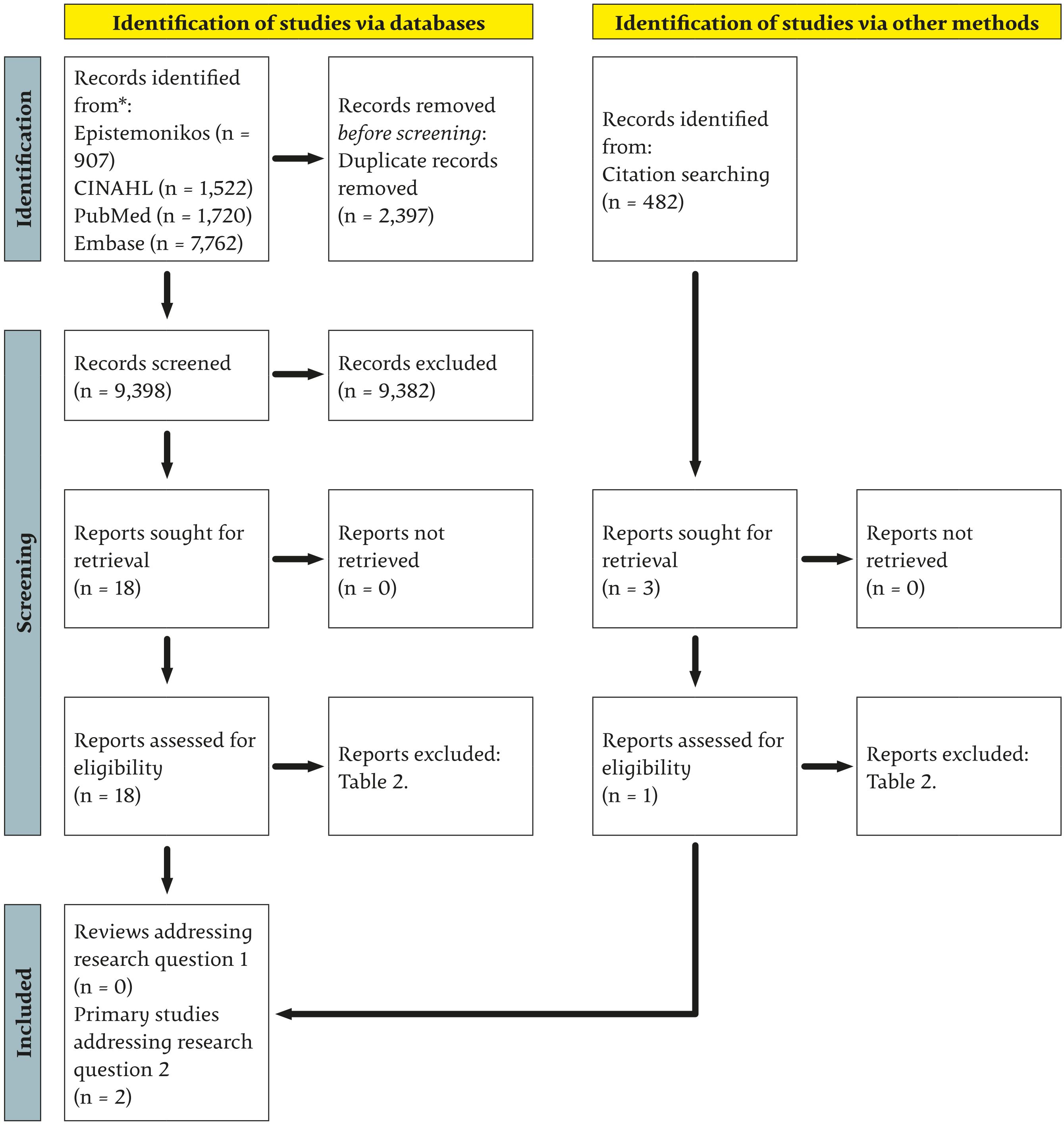


Figure 1. The PRISMA flow diagram for the scoping review

Table 2. Reasons for exclusion.

Systematic review	**Reason**
Fu *et al.*, (2024)	The review excluded studies evaluating persons with low back pain caused by disc herniations.
Feller *et al.*, (2024)	Evaluated prognostic models related to the surgical outcomes of lumbar–related leg pain (Fritzell, Mesterton and Hagg, 2022; Staartjes *et al.,* 2019) or diagnostic models (Stynes *et al.,* 2018).
Silva *et al.*, (2022) Karran *et al.*, (2017a)	The included primary studies either did not include persons with lumbar–related leg pain or, if they did, did not mention the proportion of such patients in respect to the total cohort; or less than 80% of the participants had lumbar–related leg pain.
Tagliaferri *et al.*, (2020)	The included studies evaluated prediction models of treatment outcomes for low back pain patients only or evaluated a model to predict who would experience recurrent lumbar disc herniations, yet this was not associated with any treatment.
Lheureux & Berquin, (2019) McIntosh *et al.*, (2018) Patel *et al.*, (2013)	None of the models included in these reviews evaluated persons with lumbar–related leg pain.
Haskins, Rivett & Osmotherly, (2012)	This review included two prognostic models; however, in one of the models the proportion of persons with lumbar–related leg pain was less than 80% of the cohort (n=28), and in the other study, the proportion of patients with lumbar–related leg pain was not provided.
Stanton *et al.*, (2010)	The included studies either excluded persons with lumbar–related leg pain or included studies with less than 80% of the participants having lumbar–related leg pain, or they did not mention the proportion of such patients in respect to the total cohort.
May & Rosedale, (2009) Kent & Keating, (2008)	The included studies had less than 80% of the participants having lumbar–related leg pain.
Haskins, Osmotherly & Rivett, (2015) Tagliaferri *et al.*, (2022) Ogbeivor & Elsabbagh, (2021)	The main aim of the reviews was to evaluate prediction models for low back pain disorders, and most of the studies included a cohort with less than 80% of the participants having lumbar–related leg pain.
King *et al.*, (2015)	No full text is available (presented only as a poster).
Chiodo & Haley, (2024)	Included only studies evaluating low back pain.
Almas, Parsons & Whalen, (2018)	The systematic review did not include participants with lumbar–related leg pain.

3.3. Research Question 2: Are there prognostic models addressing non–surgical treatment outcomes for individuals with lumbar–related leg pain?

Within the SRs which were excluded at the full–text stage for the first research question, there were three SRs (Haskins, Osmotherly & Rivett, 2015; Tagliaferri *et al.*, 2022; Ogbeivor & Elsabbagh, 2021) which included a primary study that derived or evaluated a prognostic model in a cohort composed of at least 80% of the participants having LRLP. Therefore, these primary studies (Kovacs *et al.*, 2012; Konstantinou *et al.*, 2020) were eligible for evaluation for our second research question. Since the SCOPIC trial evaluated the effectiveness of an algorithm, hence it was at the PROGRESS 4 level, we sought the paper on the development of the algorithm (Konstantinou *et al.*, 2019). Two of the SRs (Haskins, Osmotherly & Rivett, 2015; Tagliaferri *et al.*, 2022) included other prognostic models; however since these were evaluated in a cohort of less than 80% of persons with LRLP, they did not fulfil the inclusion criteria.

3.4. Characteristics of the systematic reviews

Table 3 presents an overview of the three SRs, which included the primary studies evaluating a prognostic model or CPR within a cohort of participants with LRLP. The review by Haskins, Osmotherly & Rivett, (2015) identified 30 CPRs, one of which was the multivariate predictive logistic regression model developed by Kovacs *et al.*, (2012). The review by Tagliaferri *et al.*, (2022) included 24 trials that evaluated the efficacy of classification approaches for managing LBP. Four of these trials used the SBT, and one trial (SCOPIC) included a cohort solely composed of persons with LRLP (Konstantinou *et al.*, 2020). Ogbeivor & Elsabbagh (2021) sought to determine the effectiveness of stratified care using the SBT compared to standard physiotherapy for LBP. The SR included seven trials and two health economic analyses for two of the seven included trials. One of the trials was also the SCOPIC trial (Konstantinou *et al.*, 2020).

3.5. Characteristics of the primary studies and their prognostic models

Kovacs *et al.*, (2012) aimed to develop three separate multivariate predictive logistic regression models to quantify the likelihood for a given patient to experience a clinically relevant improvement in LBP, leg pain and disability, respectively. The data was extracted from a post–marketing surveillance register for neuroreflexotherpay (NRT), which was defined as the implantation of surgical material in specific areas of the skin for up to 90 days (Urrútia *et al.*, 2004). The dependent variable for the model evaluating leg pain was the 0–10cm VAS scale. The improvement in pain was defined as any reduction in the VAS score that was higher than the minimal clinically important change (≥1.5 VAS points). Discrimination was evaluated using the areas under the receiver operating characteristic curve (AUC), while calibration was assessed using the Hosmer–Lemeshow test. Internal validation of the model was tested via 1,000 bootstrapping samples. Multiple imputation was done for missing data, resulting in five imputed data sets. Variable selection was performed for each dataset, but a variable was included in the combined model if it appeared as a predictor in at least two of the five imputed datasets. The regression coefficients were averaged using Rubin rules. A total of 4,477 participants were registered on the registry, with only 4.8% missing data for one or more variables, and only 0.07% of the patients were lost to follow–up. The data from 3,359 participants was used to develop the predictive model for LRLP. 75.2% of the participants with referred leg pain showed a clinically relevant improvement at discharge, while the rest did not. The model was composed of the following variables: had been treated with NRT (OR = 1.47; $p<0.0001$), previous surgery (–0.49; $p<0.0001$), baseline degree of disability (RMQ) (–0.05; $p<0.0001$), EMG (–0.52; $p=0.021$), baseline severity of LBP (VAS) (–0.07; $p=0.013$), baseline severity of referred pain (VAS) (0.21; $p<0.0001$). It obtained poor discrimination with an AUC of 0.655, and calibration was not reported.

The SCOPIC trial (Konstantinou *et al.*, 2020) was included in two SRs (Tagliaferri *et al.*, 2022; Ogbeivor & Elsabbagh, 2021) and it evaluated whether the SBT could provide any benefit compared to usual care in UK primary care for persons with sciatica. The SCOPIC trial was a two–parallel arm, pragmatic RCT (PROGRESS 4) within three centres in the UK, enrolling adults with a clinical diagnosis of sciatica. Patients were randomly allocated to either stratified care or usual care. Stratified care consisted of 3 risk–based groups: group 1 received brief advice and two physiotherapy sessions, group 2 received up to six physiotherapy sessions, and Group 3 was fast–tracked to MRI and spinal specialist assessment within 4 weeks of randomisation. The primary outcome was time to first resolution of the sciatic symptoms, defined as "completely recovered" or "much better" on a 6-point Likert scale. It enrolled 238 patients in each treatment arm. There was not a statistically significant difference in the median time to symptom resolution between the two treatment arms at 4 months (p=0.66) and 12 months (p=0.22), and stratified care was deemed not cost-effective compared to usual care.

The algorithm used within the SOCPIC trial was dependent on the SBT, but included other variables being current leg pain intensity (NRS 0–10 with a cut-off >6), pain radiating below the knee (yes/no), pain interference with work or home activities (NRS 0–10 with a cut-off >6) and objective sensory loss (yes/no) subgroup (Konstantinou *et al.*, 2019). The algorithm was derived by first conducting a logistic regression analysis with the dependent variable being the patient being referred to specialist services (yes/no). The model achieved a calibration of 1.0 (95%CI 0.57 to 1.43), and an AUC of 0.695 (95%CI 0.622 to 0.768), and after conducting bootstrapping, the AUC was 0.678 (95%CI 0.674 to 0.681). At this stage, three variables obtained statistical significance within the model, being: pain interference with work or home activities (aOR 2.17; 95% CI 1.13 to 4.17), current leg pain intensity (aOR 1.17; 95%CI 1.05 to 1.31) and sensory loss (aOR 2.41; 95% CI 1.05 to 5.53). The predictive factors were discussed with various experts who added an extra variable: pain below the knee (yes/no) since this is considered the best proxy indicator of leg pain due to nerve root involvement. Furthermore, the experts added cut-off points for the impact and pain intensity scales since these thresholds for pain and functional limitations are considered reasonable and have face validity for considering early referral to specialists. Afterwards, multiple combinations, including the SBT cut-off scores for the three risk groups and the four new variables, were computed, aiming to achieve the optimal combination for onward referral to specialist services based on sensitivity, specificity, and negative and positive predictive values. The authors and experts identified the optimal criteria for prioritizing sciatica patients for spinal specialist services as either: (1) a high-risk (≥4 on the psychological subscale score) on the SBT combined with at least three of the four specified clinical characteristics, or (2) a medium-risk (3≤ on the psychological subscale score) classification on the SBT alongside the presence of all four clinical characteristics. This algorithm achieved a sensitivity of 51% (95% CI 37 to 64), a specificity of 73% (95% CI 68 to 78), a positive predictive value of 22% (95% CI 16 to 31), a negative predictive value of 91% (95% CI 87 to 94) and 30% (129 out of 429 participants) of the total sample being referred for specialist services. The positive predictive value indicated that 22% of the patients being referred to specialist services would be appropriately referred. This algorithm was based on the sensitivity value and feasibility of the spinal specialist services to handle the number of referrals.

A direct comparison between the two models is challenging due to several factors, including differences in the healthcare settings (UK vs. Spanish primary care), the dependent variables predicted by each model (dichotomized pain intensity vs. onward referral to specialist services), the PROGRESS levels used (PROGRESS 3 vs. 4), and the methodological approaches adopted (logistic regression vs. mixed methods). The model developed by Kovacs *et al.*, (2012) could be subject to bias due to potential conflicts of interest stemming from pharmaceutical company involvement, which could have influenced the reported positive effect of NRT. However, despite these findings, NRT is not recommended in any major clinical practice guidelines for LBP with or without LRLP (NICE, 2020a). Additionally, Kovacs *et al.*, (2012) dichotomized a continuous outcome (VAS 0–10), thereby reducing the statistical power of their prognostic model. In contrast, the algorithm used in the SCOPIC trial (Konstantinou *et al.*, 2020) demonstrated slightly better AUC values in comparison to the Kovacs *et al.* (2012) model (0.678 vs. 0.655) and was well calibrated, with a calibration slope of 1.0 (95% CI 0.57 to 1.43). Notably, the calibration value for the model developed by Kovacs *et al.* (2012) was not reported.

3.6. Critical appraisal

The Kovacs *et al.*, (2012) model was at high risk of bias (Table 4) since the data was obtained from a routine care registry being a marketing surveillance study on NRT, explaining why 94.8% of the participants within the study had received such therapy. The final model included the variable "treated with neuroreflexology". This variable would not be available at baseline since the patient would not have yet received treatment, making the model unusable at baseline. The bias derived from this variable in the model is reflected by the high association with the outcome (coefficient = 1.47) compared to the magnitude of the other coefficients. The information on the model's calibration was not provided. Instead, the authors provided the p-value (0.156) of the Hosmer-Lemeshow test, which does not indicate the presence nor the magnitude of any miscalibration. All these factors lead to downgrading the risk of bias grade and applicability for this model.

The algorithm used in the SCOPIC trial is not strictly a prediction model based on statistical regression analysis, but it utilises a mixed methodology (Konstantinou *et al.*, 2019). After conducting the initial logistic regression analysis, the authors sought an iterative process involving several combinations to inflate the algorithm's performance, therefore downgrading the risk of bias grade for the outcome domain. The initial logistic regression model had an event per variable (EPV) ratio of 2.6, which would be even smaller when adding the three risk groups of the SBT. Ideally, an EPV of at least 20 is recommended in model development studies (Wolff *et al.*, 2019). The EPV ratio refers to the number of outcome events (e.g., patients experiencing an event) available per predictor variable in a prognostic model. A low EPV increases the risk of overfitting, making the model unreliable. Although a few participants had missing baseline data, the authors did not provide information on how this was tackled, for example, whether multiple imputation was used. These lead to downgrading the risk of bias grade. Figure 2 provides an overall summary of the risk of bias assessment using PROBAST and further detail is provided in Appendix C.

Table 3. Characteristics of the systematic reviews and the eligible primary studies.

Review	**Search span / Databases**	**Aims of the review**	**Primary study / Prognostic model/tool**	**Eligibility criteria pertaining to lumbar–related leg pain within the primary study**
Haskins, Osmotherly & Rivett, (2015)	From inception to July 2013. MEDLINE, EMBASE, CENTRAL, PsychINFO, CINAHL, AMED, and Index to Chiropractic Literature.	To identify prognostic forms of clinical prediction rules related to the nonsurgical management of adults with LBP and to evaluate their current stage of development.	Kovacs *et al.*, (2012) Multivariate predictive logistic regression model.	Inclusion criteria encompassed patients with low back pain, with or without leg pain, without trauma or systemic disease causes, and could read Spanish. Patients with prior unsuccessful spine surgery were eligible.

Table 3. Characteristics of the systematic reviews and the eligible primary studies.

Tagliaferri *et al.*, (2022)	From inception to June 2021. MEDLINE, EMBASE, CINAHL, the Web of Science Core Collection, CENTRAL.	To determine the efficacy of nonsurgical classification systems for treating LBP, compared to general comparators, for LBP, leg pain intensity, and disability. The secondary aim was to determine the effectiveness of treating subclasses of individuals in classification systems.	Konstantinou, *et al.*, (2020) Algorithm based on the STarT Back tool (SCOPIC trial)	Eligible patients were 18 years or older, with a clinical diagnosis of sciatica of any severity and duration, following confirmed by assessment, and had not received treatment for this condition within the last three months. Sciatica was defined by symptoms such as leg pain approximating following a dermatomal pattern, leg pain equal to or worse than back pain, leg pain aggravated by coughing or straining, dermatomal sensory changes, neurological deficits indicating nerve root compression, a positive neural tension test, or leg pain exacerbated by weight-bearing and relieved by sitting (specifically spinal stenosis).
Ogbeivor & Elsabbagh, (2021)	From 1 January 2000 to 5 July 2020. CINAHL, MEDLINE, Pedro, EMBASE, PsycINFO, CENTRAL and Web of Science.	The aim was to investigate stratified care's long-term clinical, and cost-effectiveness compared with non-stratified care to determine which approach is better for the long-term management of patients with LBP.		

CENTRAL – Cochrane Central Register of Controlled Trials

Table 4. Tabular presentation of PROBAST results.

Study	ROB				Applicability			Overall	
	Participants	**Predictors**	**Outcome**	**Analysis**	**Participants**	**Predictors**	**Outcome**	**ROB**	**Applicability**
Kovacs *et al.*, (2012)	–	–	+	–	+	–	–	–	–
Konstantinou *et al.*, (2020)	+	+	–	?	+	+	+	–	+

PROBAST = Prediction model Risk Of Bias ASsessment Tool; ROB = risk of bias.

+ indicates low ROB/low concern regarding applicability;
– indicates high ROB/high concern regarding applicability;
? indicates unclear ROB/unclear concern regarding applicability.

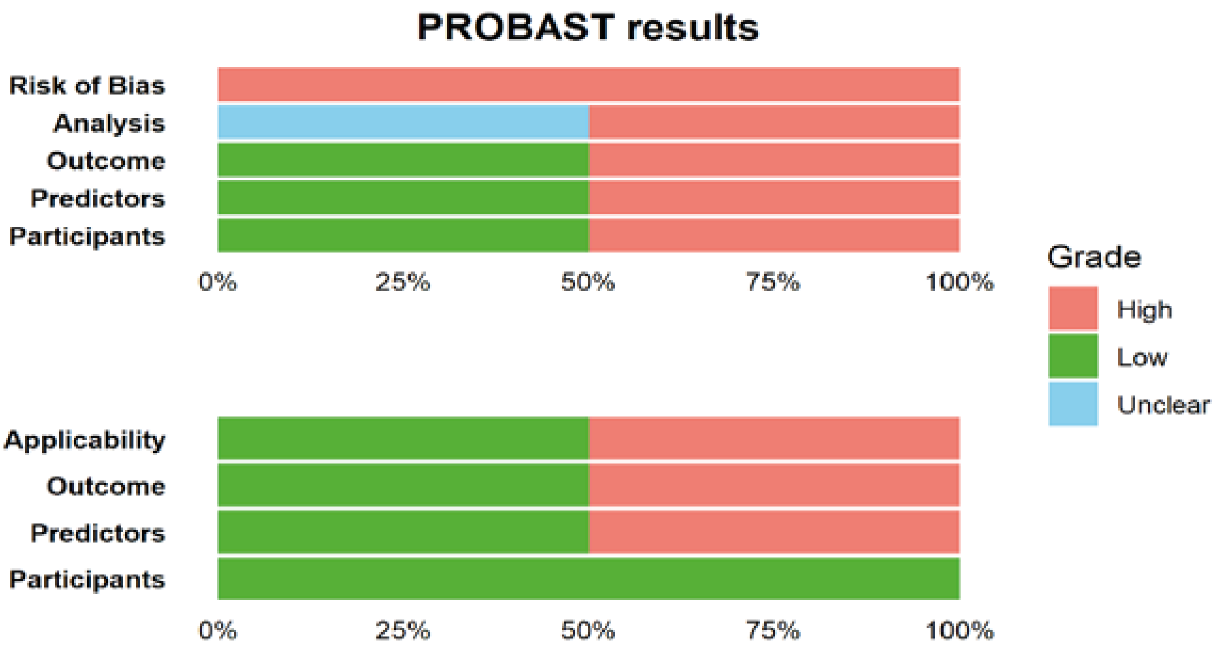


Figure 2. Summary plot of the PROBAST results.

4. Strengths and limitations of this scoping review

This scoping review demonstrates several methodological strengths. First, we conducted a comprehensive search across four databases, supplemented by forward citation searching, ensuring a broad and inclusive retrieval of relevant literature. A validated and sensitive search filter was employed for each database, except Epistemonikos, for which no validated filter currently exists. The search spanned the past 24 years, and since the review focused on SRs, their search periods potentially extended further back in time, allowing for the inclusion of a larger body of relevant literature. Moreover, the search strategy was peer–reviewed by a university librarian using the PRESS methodology, ensuring rigor and comprehensiveness. The review adhered to established reporting guidelines, with screening, data extraction, and critical appraisal of the two included prognostic models conducted independently and in duplicate to enhance reliability. Another strength lies in the investigation of the primary studies on the proportions of patients with LRLP to determine the value of the papers reviewed (as reported in Appendix B). The main limitations of this scoping review are that the search was limited to articles published in the English language, it did not search an interdisciplinary database for example, SCOPUS. Finally, this review pointed to research that needs to be conducted rather than contributing to original research.

5. Conclusion

The strengths of this scoping review provide a high level of confidence in its findings. We conclude that no SR to date has specifically evaluated prognostic models for treatment outcomes in individuals with LRLP. However, evidence indicates that apart from the two identified models (Kovacs *et al.*, 2012; Konstantinou *et al.*, 2020), various prognostic models for related leg pain have been developed (Matsudaira *et al.*, 2016; Sun *et al.*, 2023; Azharuddin *et al.*, 2022). Consequently, a SR of these models is warranted to consolidate evidence and inform clinical practice.

Funding: The research work disclosed in this publication is funded by the Tertiary Education Scholarships Scheme (Malta).

Conflict of Interest: The authors declare no conflict of interest.

Acknowledgements: The authors would like to thank Ms Agata Scicluna Derkowska, Senior Assistant Librarian at the University of Malta, for her expert assistance in developing the search strategy and reviewing the PRESS checklist.

References

Almas, M., Parsons, B. & Whalen, E. (2018). Prediction of therapeutic response to pregabalin in subjects with neuropathic pain. *Current Medical Research and Opinion*, 34(12), 2041–2052. https://doi.org/10.1080/03007995.2018.1520694.

Azharuddin, A., Aryandono, T., Magetsari, R. & Dwiprahasto, I. (2022). Predictors of the conservative management outcomes in patients with lumbar herniated nucleus pulposus: A prospective study in Indonesia. *Asian Journal of Surgery*, 45(1), 277–283. https://doi.org/10.1016/j.asjsur.2021.05.015.

Azimi, P., Mohammadi, H.R., Benzel, E.C., Shahzadi, S. & Azhari, S. (2015). Use of Artificial Neural Networks to Predict Recurrent Lumbar Disk Herniation. *Journal of Spinal Disorders & Techniques*, 28(3), E161–E165. https://doi.org/10.1097/BSD.0000000000000200.

Barons, M.J., Parsons, N., Griffiths, F. & Thorogood, M. (2013). A comparison of artificial neural network, latent class analysis and logistic regression for determining which patients benefit from a cognitive behavioral approach to treatment for non–specific low back pain. *2013 IEEE Symposium on Computational Intelligence in Healthcare and E–Health (CICARE)*. https://doi.org/10.1109/cicare.2013.6583061

Beneciuk, J.M., Bishop, M.D., Fritz, J.M., Robinson, M.E., Asal, N.R., Nisenzon, A.N., *et al.* (2013). The STarT Back Screening Tool and Individual Psychological Measures: Evaluation of Prognostic Capabilities for Low Back Pain Clinical Outcomes in Outpatient Physical Therapy Settings. *Physical Therapy*, 93(3), 321–333. https://doi.org/10.2522/ptj.20120207.

Boden, L.M., Boden, S.A., Premkumar, A., Gottschalk, M.B. & Boden, S.D. (2018). Predicting Likelihood of Surgery Before First Visit in Patients With Back and Lower Extremity Symptoms: A Simple Mathematical Model Based on More Than 8,000 Patients. *Spine*, 43(18), 1296–1305. https://doi.org/10.1097/BRS.0000000000002603.

Chiodo, A.F. & Haley, M. (2024). Does risk stratification with a matched treatment pathway improve clinical outcomes for adults with acute back pain? A systematic review and meta–analysis. *Revista Brasileira de Fisioterapia*, 28(5), 101116. https://doi.org/10.1016/j.bjpt.2024.101116.

Congcong, C., Yong, H.P. & Kian, C.L. (2009). A clinical prediction rule for classifying patients with low back pain who demonstrate short–term improvement with mechanical lumbar traction. *European Spine Journal*, 18(4), 554–561. https://doi.org/10.1007/s00586-009-0909-9.

Cook, C. (2008). Potential Pitfalls of Clinical Prediction Rules. *The Journal of Manual & Manipulative Therapy*, 16(2), 69–71. https://doi.org/10.1179/106698108790818477.

D'Agostino, R. B., Sr, Vasan, R. S., Pencina, M. J., Wolf, P. A., Cobain, M., Massaro, J. M., *et al.* (2008). General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation*, 117(6), 743–753. https://doi.org/10.1161/CIRCULATIONAHA.107.699579.

Feller, D., Wingbermuhle, R., Berg, B., Vigdal, ØN., Innocenti, T., Grotle, M., *et al.* (2024). Improvements Are Needed in the Adherence to the TRIPOD Statement for Clinical Prediction Models for Patients With Spinal Pain or Osteoarthritis: A Metaresearch Study. *The Journal of Pain*, 25(11), 104624. https://doi.org/10.1016/j.jpain.2024.104624.

Field, J. & Newell, D. (2012). Relationship between STarT Back Screening Tool and prognosis for low back pain patients receiving spinal manipulative therapy. *Chiropractic & Manual Therapies*, 20(1), 17. https://doi.org/10.1186/2045-709X-20-17.

Flynn, T., Fritz, J., Whitman, J., Wainner, R., Magel, J., Rendeiro, D., *et al.* (2002). A Clinical Prediction Rule for Classifying Patients with Low Back Pain Who Demonstrate Short–Term Improvement With Spinal Manipulation. *Spine*, 27(24), 2835–2843. https://doi.org/10.1097/00007632-200212150-00021.

Foster, N.E., Hill, J.C., O'Sullivan, P. & Hancock, M. (2013). Stratified models of care. *Best Practice & Research Clinical Rheumatology*, 27(5), 649–661. https://doi.org/10.1016/j.berh.2013.10.005.

Fritz, J.M., Childs, J.D. & Flynn, T.W. (2005). Pragmatic application of a clinical prediction rule in primary care to identify patients with low back pain with a good prognosis following a brief spinal manipulation intervention. *BMC Family Practice*, 6(1), 29. https://doi.org/10.1186/1471-2296-6-29.

Fritz, J.M., Lindsay, W., Matheson, J.W., Brennan, G.P., Hunter, S.J., Moffit, S.D., *et al.* (2007). Is there a subgroup of patients with low back pain likely to benefit from mechanical traction? Results of a randomized clinical trial and subgrouping analysis. Spine, 32(26), E793–E800. https://doi.org/10.1097/BRS.0b013e31815d001a.

Fritzell, P., Mesterton, J. & Hagg, O. (2022). Prediction of outcome after spinal surgery—using The Dialogue Support based on the Swedish national quality register. *European Spine Journal*, 31(4), 889–900. https://doi.org/10.1007/s00586-021-07065-y.

Fu, Y., Feller, D., Koes, B. & Chiarotto, A. (2024). Prognostic Models for Chronic Low Back Pain Outcomes in Primary Care Are at High Risk of Bias and Lack Validation–High–Quality Studies Are Needed: A Systematic Review. *The Journal of Orthopaedic and Sports Physical Therapy*, 54(5), 1–314. https://doi.org/10.2519/jospt.2024.12081.

Gabel, C.P., Melloh, M., Yelland, M., Burkett, B. & Roiko, A. (2011). Predictive ability of a modified Örebro Musculoskeletal Pain Questionnaire in an acute/subacute low back pain working population. *European Spine Journal*, 20(3), 449–457. https://doi.org/10.1007/s00586-010-1509-4.

Gal, N., Stoicu–Tivadar, V., Andrei, D., Nemeş, D.I. & Nădăşan, E. (2014). Computer assisted treatment prediction of low back pain pathologies. *Studies in Health Technology and Informatics*, 197, 47–51. https://doi.org/10.3233/978-1-61499-389-6-47.

Geersing, G., Bouwmeester, W., Zuithoff, P., Spijker, R., Leeflang, M., Moons, K.G.M., *et al.* (2012). Search Filters for Finding Prognostic and Diagnostic Prediction Studies in Medline to Enhance Systematic Reviews. *PLoS ONE*, 7(2), e32844. https://doi.org/10.1371/journal.pone.0032844.

George, S.Z., Bialosky, J.E. & Donald, D.A. (2005). The Centralization Phenomenon and Fear–Avoidance Beliefs as Prognostic Factors for Acute Low Back Pain: A Preliminary Investigation Involving Patients Classified for Specific Exercise. *The Journal of Orthopaedic and Sports Physical Therapy*, 35(9), 580–588. https://doi.org/10.2519/jospt.2005.35.9.580.

Goossen, K., Hess, S., Lunny, C. & Pieper, D. (2020). Database combinations to retrieve systematic reviews in overviews of reviews: a methodological study. *BMC Medical Research Methodology*, 20(1), 1–138. https://doi.org/10.1186/s12874-020-00983-3.

Grimshaw, J. (2010). *A Guide to Knowledge Synthesis A Knowledge Synthesis Chapter.* Available at: https://cihr-irsc.gc.ca/e/41382.html

Grotle, M., Vøllestad, N.K. & Brox, J.I. (2006). Screening for Yellow Flags in First–time Acute Low Back Pain: Reliability and Validity of a Norwegian Version of the Acute Low Back Pain Screening Questionnaire. The *Clinical journal of Pain*, 22(5), 458–467. https://doi.org/10.1097/01.ajp.0000208243.33498.cb.

Haddaway, N.R., Grainger, M.J. & Gray, C.T. (2022). Citationchaser: A tool for transparent and efficient forward and backward citation chasing in systematic searching. *Research Synthesis Methods*, 13(4), 533–545. https://doi.org/10.1002/jrsm.1563.

Hancock, M.J., Maher, C.G., Latimer, J., Herbert, R.D. & McAuley, J.H. (2009). Can rate of recovery be predicted in patients with acute low back pain? Development of a clinical prediction rule. *European Journal of Pain*, 13(1), 51–55. https://doi.org/10.1016/j.ejpain.2008.03.007.

Haskins, R., Osmotherly, P.G. & Rivett, D.A. (2015). Validation and impact analysis of prognostic clinical prediction rules for low back pain is needed: a systematic review. *Journal of Clinical Epidemiology*, 68(7), 821–832. https://doi.org/10.1016/j.jclinepi.2015.02.003.

Haskins, R., Rivett, D.A. & Osmotherly, P.G. (2012). Clinical prediction rules in the physiotherapy management of low back pain: A systematic review. *Manual therapy*, 17(1), 9–21. https://doi.org/10.1016/j.math.2011.05.001.

Hazard, R.G., Haugh, L.D., Reid, S., McFarlane, G. & MacDonald, L. (1997). Early physician notification of patient disability risk and clinical guidelines after low back injury: A randomized, controlled trial. Spine, 22(24), 2951–2958. https://doi.org/10.1097/00007632-199712150-00019.

Hazard, R.G., Haugh, L.D., Reid, S., Preble, J.B. & MacDonald, L. (1996). Early Prediction of Chronic Disability After Occupational Low Back Injury. *Spine*, 21(8), 945–951. https://doi.org/10.1097/00007632-199604150-00008.

Hemingway, H., Croft, P., Perel, P., Hayden, J.A., Abrams, K., Timmis, A., *et al.* (2013). Prognosis research strategy (PROGRESS) 1: A framework for researching clinical outcomes. *BMJ*, 346(feb05 1), 5595. https://doi.org/10.1136/bmj.e5595.

Heneweer, H., Aufdemkampe, G., Van Tulder, M.W., Kiers, H., Stappaerts, K.H. & Vanhees, L. (2007). Psychosocial variables in patients with (sub)acute low back pain: an inception cohort in primary care physical therapy in The Netherlands. *Spine*, 32(5), 586–592. https://doi.org/10.1097/01.brs.0000256447.72623.56.

Hicks, G.E., Fritz, J.M., Delitto, A. & McGill, S.M. (2005). Preliminary Development of a Clinical Prediction Rule for Determining Which Patients With Low Back Pain Will Respond to a Stabilization Exercise Program. *Archives of Physical Medicine and Rehabilitation*, 86(9), 1753–1762. https://doi.org/10.1016/j.apmr.2005.03.033.

Hill, J.C., Whitehurst, D.G., Lewis, M., Bryan, S., Dunn, K.M., Foster, N.E., *et al.* (2011). Comparison of stratified primary care management for low back pain with current best practice (STarT Back): a randomised controlled trial. *Lancet*, 378(9802), 1560–1571. https://doi.org/10.1016/S0140-6736(11)60937-9.

Hingorani, A.D., Windt, D.A.v.d., Riley, R.D., Abrams, K., Moons, K.G.M., Steyerberg, E.W., *et al.* (2013). Prognosis research strategy (PROGRESS) 4: Stratified medicine research. *BMJ*, 346(feb05 1), e5793. https://doi.org/10.1136/bmj.e5793.

Holland, J.L., Wilczynski, N.L. & Haynes, R.B. (2005). Optimal search strategies for identifying sound clinical prediction studies in EMBASE. *BMC Medical Informatics and Decision Making*, 5(1), 11. https://doi.org/10.1186/1472-6947-5-11.

Jellema, P., van der Windt, Daniëlle A W M, van der Horst, H.E., Stalman, W.A.B. & Bouter, L.M. (2007). Prediction of an unfavourable course of low back pain in general practice: comparison of four instruments. *British Journal of General Practice*, 57(534), 15–22.

Karran, E.L., McAuley, J.H., Traeger, A.C., Hillier, S.L., Grabherr, L., Russek, L.N., *et al.* (2017). Can screening instruments accurately determine poor outcome risk in adults with recent onset low back pain? A systematic review and meta–analysis. *BMC Medicine*, 15(1) https://doi.org/10.1186/s12916-016-0774-4.

Kavanagh, P.L., Frater, F., Navarro, T., LaVita, P., Parrish, R. & Iorio, A. (2021). Optimizing a literature surveillance strategy to retrieve sound overall prognosis and risk assessment model papers. *Journal of the American Medical Informatics Association,* 28(4), 766–771. https://doi.org/10.1093/jamia/ocaa232.

Kendell, M., Beales, D., O'Sullivan, P., Rabey, M., Hill, J. & Smith, A. (2018). The predictive ability of the STarT Back Tool was limited in people with chronic low back pain: a prospective cohort study. *Journal of Physiotherapy*, 64(2), 107–113. https://doi.org/10.1016/j.jphys.2018.02.009.

Kent, P.M. & Keating, J.L. (2008). Can we predict poor recovery from recent–onset nonspecific low back pain? A systematic review. *Manual Therapy*, 13(1), 12–28. https://doi.org/10.1016/j.math.2007.05.009.

King, P., Galvin, R., O'Sullivan & Bennett, Z. (2015). A systematic review of clinical prediction rules to aid treatment selection in musculoskeletal physiotherapy practice. *BMC Proceedings*, 9(S1), A25. https://doi.org/10.1186/1753-6561-9-S1-A25.

Kongsted, A., Andersen, C.H., Hansen, M.M. & Hestbaek, L. (2016). Prediction of outcome in patients with low back pain – A prospective cohort study comparing clinicians' predictions with those of the Start Back Tool. *Manual Therapy*, 21, 120–127. https://doi.org/10.1016/j.math.2015.06.008.

Konstantinou, K., Dunn, K.M., Van Der Windt, D., Ogollah, R., Jasani, V., Foster, N.E., Artus, M., Hay, E.M., Lewis, M., Hill, J.C. & Mallen, C. (2019). Subgrouping patients with sciatica in primary care for matched care pathways: Development of a subgrouping algorithm. *BMC Musculoskeletal Disorders*, 20(1), 313. https://doi.org/10.1186/s12891-019-2686-x

Konstantinou, K., Dunn, K.M., Ogollah, R., Lewis, M., van der Windt, D. & Hay, E.M. (2018). Prognosis of sciatica and back–related leg pain in primary care: the ATLAS cohort. *The Spine Journal*, 18(6), 1030–1040. https://doi.org/10.1016/j.spinee.2017.10.071.

Konstantinou, K., Lewis, M., Dunn, K.M., Ogollah, R., Artus, M., Hill, J.C., *et al.* (2020). Stratified care versus usual care for management of patients presenting with sciatica in primary care (SCOPiC): a randomised controlled trial. *The Lancet Rheumatology*, 2(7), e401–e411. https://doi.org/10.1016/S2665-9913(20)30099-0.

Kovacs, F.M., Seco, J., Royuela, A., Corcoll Reixach, J., & Abraira, V. (2012). Predicting the evolution of low back pain patients in routine clinical practice: results from a registry within the Spanish National Health Service. *The Spine Journal*, 12(11), 1008–1020. https://doi.org/10.1016/j.spinee.2012.10.007.

Kraemer, H.C., Frank, E. & Kupfer, D.J. (2006). Moderators of Treatment Outcomes. *JAMA*, 296(10), 1286. https://doi.org/10.1001/jama.296.10.1286.

Law, R.K.Y., Lee, E.W.C., Law, S., Chan, B.K.B., Chen, P. & Szeto, G.P.Y. (2013). The Predictive Validity of OMPQ on the Rehabilitation Outcomes for Patients with Acute and Subacute Non–Specific LBP in a Chinese Population. *Journal of Occupational Rehabilitation*, 23(3), 361–370. https://doi.org/10.1007/s10926-012-9404-y.

Lheureux, A. & Berquin, A. (2019). Comparison between the STarT Back Screening Tool and the Örebro Musculoskeletal Pain Screening Questionnaire: Which tool for what purpose? A semi–systematic review. *Annals of Physical and Rehabilitation Medicine*, 62(3), 178–188. https://doi.org/10.1016/j.rehab.2018.09.007.

Linton, S.J. & Boersma, K. (2003). Early Identification of Patients at Risk of Developing a Persistent Back Problem: The Predictive Validity of The Örebro Musculoskeletal Pain Questionnaire. The Clinical Journal of Pain, 19(2), 80–86. https://doi.org/10.1097/00002508-200303000-00002.

Matsudaira, K., Hara, N., Oka, H., Kunogi, J., Yamazaki, T., Takeshita, K., *et al.* (2016). Predictive Factors for Subjective Improvement in Lumbar Spinal Stenosis Patients with Nonsurgical Treatment: A 3–Year Prospective Cohort Study. *PLOS ONE*, 11(2), e0. https://doi.org/10.1371/journal.pone.0148584.

May, S. & Rosedale, R. (2009). Prescriptive Clinical Prediction Rules in Back Pain Research: A Systematic Review. *The Journal of Manual & Manipulative Therapy*, 17(1), 36–45. https://doi.org/10.1179/106698109790818214.

McGowan, J., Sampson, M., Salzwedel, D.M., Cogo, E., Foerster, V. & Lefebvre, C. (2016). PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. *Journal of Clinical Epidemiology*, 75, 40–46. https://doi.org/10.1016/j.jclinepi.2016.01.021.

McIntosh, G., Steenstra, I., Hogg–Johnson, S., Carter, T. & Hall, H. (2018). Lack of Prognostic Model Validation in Low Back Pain Prediction Studies: A Systematic Review. *The Clinical Journal of Pain*, 34(8), 748–754. https://doi.org/10.1097/AJP.0000000000000591.

Mehling, W.E., Avins, A.L., Acree, M.C., Carey, T.S. & Hecht, F.M. (2015). Can a back pain screening tool help classify patients with acute pain into risk levels for chronic pain? *European Journal of Pain*, 19(3), 439–446. https://doi.org/10.1002/ejp.615.

Mehling, W.E., Ebell, M.H., Avins, A.L., & Hecht, F.M. (2015). Clinical decision rule for primary care patient with acute low back pain at risk of developing chronic pain. *The Spine Journal*, 15(7), 1577–1586. https://doi.org/10.1016/j.spinee.2015.03.003.

Moons, K.G.M., de Groot, J.A.H., Bouwmeester, W., Vergouwe, Y., Mallett, S., Altman, D.G., *et al.* (2014). Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist. *PLoS Medicine*, 11(10), e1001744. https://doi.org/10.1371/journal.pmed.1001744 16.

Moons K.G.M., Altman, D.G., Reitsma, J.B., Ioannidis, J.P.A., Macaskill, P., Steyerberg, E.W., *et al.* (2015). Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. *Ann Intern Med*, 162(1–73). https://doi.org/10.7326/M14-0698.

Morsø, L., Kent, P., Manniche, C. & Albert, H.B. (2014). The predictive ability of the STarT Back Screening Tool in a Danish secondary care setting. *European Spine* Journal, 23(1), 120–128. https://doi.org/10.1007/s00586-013-2861-y.

Newell, D., Field, J. & Pollard, D. (2015). Using the STarT Back Tool: Does timing of stratification matter? *Manual Therapy*, 20(4), 533–539. https://doi.org/10.1016/j.math.2014.08.001.

NICE (2020a). *Low back pain and sciatica in over 16s: assessment and management*. National Institute for Health and Care Excellence (NICE).

NICE (2020b). *Endorsed resource – STarT Back Screening Tool with matched treatment options.* National Institute for Health and Care Excellence (NICE). Available at: https://www.nice.org.uk/guidance/ng59/resources/endorsed-resource-start-back-screening-tool-with-matched-treatment-options-4906309933.

Nonclercq, O. & Berquin, A. (2012). Predicting chronicity in acute back pain: Validation of a French translation of the Örebro Musculoskeletal Pain Screening Questionnaire. *Annals of Physical and Rehabilitation Medicine*, 55(4), 263–278. https://doi.org/10.1016/j.rehab.2012.03.002.

Ogbeivor, C. & Elsabbagh, L. (2021). Management approach combining prognostic screening and targeted treatment for patients with low back pain compared with standard physiotherapy: A systematic review & meta–analysis. *Musculoskeletal Care*, 19(4), 436–456. https://doi.org/10.1002/msc.1541.

Parreira, P., Maher, C.G., Steffens, D., Hancock, M.J. & Ferreira, M.L. (2018). Risk factors for low back pain and sciatica: an umbrella review. *The Spine Journal*, 18(9), 1715–1721. https://doi.org/10.1016/j.spinee.2018.05.018.

Patel, S., Friede, T., Froud, R., Evans, D.W. & Underwood, M. (2013). Systematic Review of Randomized Controlled Trials of Clinical Prediction Rules for Physical Therapy in Low Back Pain. *Spine*, 38(9), 762–769. https://doi.org/10.1097/BRS.0b013e31827b158f.

Peters, M.D.J., Marnie, C., Tricco, A.C., Pollock, D., Munn, Z., Alexander, L., *et al.* (2020). Updated methodological guidance for the conduct of scoping reviews. *JBI*

Evidence Synthesis, 18(10), 2119–2126. https://doi.org/10.11124/jbies-20-00167.

Riley, R.D., Hayden, J.A., Steyerberg, E.W., Moons, K.G.M., Abrams, K., Kyzas, P.A., *et al.* (2013). Prognosis Research Strategy (PROGRESS) 2: Prognostic Factor Research. *PLoS Medicine*, 10(2), e1–e1001380. https://doi.org/10.1371/journal.pmed.1001380.

Schmidt, C.O., Kohlmann, T., Pfingsten, M., Lindena, G., Marnitz, U., Pfeifer, K., *et al.* (2016). Construct and predictive validity of the German Örebro questionnaire short form for psychosocial risk factor screening of patients with low back pain. *European Spine Journal*, 25(1), 325–332. https://doi.org/10.1007/s00586-015-4196-3.

Shaw, W.S., Pransky, G. & Winters, T. (2009). The Back Disability Risk Questionnaire for Work–Related, Acute Back Pain: Prediction of Unresolved Problems at 3–Month Follow–Up. *Journal of Occupational and Environmental Medicine*, 51(2), 185–194. https://doi.org/10.1097/JOM.0b013e318192bcf8.

Shea, B.J., Reeves, B.C., Wells, G., Thuku, M., Hamel, C., Moran, J., *et al.* (2017). AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non–randomised studies of healthcare interventions, or both. *BMJ (Online)*, 358, j4008. https://doi.org/10.1136/bmj.j4008.

Silva, F.G., Costa, L.O., Hancock, M.J., Palomo, G.A., Costa, L.C. & da Silva, T. (2022). No prognostic model for people with recent–onset low back pain has yet been demonstrated to be suitable for use in clinical practice: a systematic review. *Journal of Physiotherapy*, 68(2), 99–109. https://doi.org/10.1016/j.jphys.2022.03.009.

Silva, T., Macaskill, P., Kongsted, A., Mills, K., Maher, C.G. & Hancock, M.J. (2019). Predicting pain recovery in patients with acute low back pain: Updating and validation of a clinical prediction model. *European Journal of Pain*, 23(2), 341–353. https://doi.org/10.1002/ejp.1308.

Staartjes, V.E., de Wispelaere, M.P., Vandertop, W.P. & Schröder, M.L. (2019). Deep learning–based preoperative predictive analytics for patient–reported outcomes following lumbar discectomy: feasibility of center–specific modeling. *The Spine Journal*, 19(5), 853–861. https://doi.org/10.1016/j.spinee.2018.11.009.

Stanton, T.R., Hancock, M.J., Maher, C.G. & Koes, B.W. (2010). Critical Appraisal of Clinical Prediction Rules That Aim to Optimize Treatment Selection for Musculoskeletal Conditions. *Physical Therapy*, 90(6), 843–854. https://doi.org/10.2522/ptj.20090233.

Steyerberg, E.W., Moons, K.G.M., van der Windt, D.A., Hayden, J.A., Perel, P., Schroter, S., *et al.* (2013). *Prognosis Research Strategy (PROGRESS) 3*: Prognostic Model Research. *PLoS Medicine*, 10(2), e1–e1001381. https://doi.org/10.1371/journal.pmed.1001381.

Stynes, S., Konstantinou, K., Ogollah, R., Hay, E.M. & Dunn, K.M. (2018). Clinical diagnostic model for sciatica developed in primary care patients with low back–related leg pain. *PloS One*, 13(4), e0191852. https://doi.org/10.1371/journal.pone.0191852.

Sun, D., Liu, Y., Luo, D., Wu, Y., Yan, Z., Liang, Y., *et al.* (2023). A multidimensional nomogram combining clinical factors and imaging features to predict 1–year recurrence of low back pain with or without radicular pain after spinal manipulation/mobilization. *Chiropractic & Manual Therapies*, 31(1), 27. https://doi.org/10.1186/s12998-023-00500-5.

Tagliaferri, S.D., Angelova, M., Zhao, X., Owen, P.J., Miller, C.T., Wilkin, T., *et al.* (2020). Artificial intelligence to improve back pain outcomes and lessons learned from clinical classification approaches: three systematic reviews. *NPJ Digital Medicine*, 3(1), 93. https://doi.org/10.1038/s41746-020-0303-x.

Tagliaferri, S.D., Mitchell, U.H., Saueressig, T., Owen, P.J., Miller, C.T. & Belavy, D.L. (2022). Classification Approaches for Treating Low Back Pain Have Small Effects That Are Not Clinically Meaningful: A Systematic Review With Meta–analysis. *The Journal of Orthopaedic and Sports Physical Therapy*, 52(2), 67–84. https://doi.org/10.2519/jospt.2022.10761.

Traeger, A.C., Henschke, N., Hübscher, M., Williams, C.M., Kamper, S.J., Maher, C.G., *et al.* (2016). Estimating the Risk of Chronic Pain: Development and Validation of a Prognostic Model (PICKUP) for Patients with Acute Low Back Pain. *PLoS Medicine*, 13(5), e1002019. https://doi.org/10.1371/journal.pmed.1002019.

Tricco, A.C., Lillie, E., Zarin, W., O'Brien, K.K., Colquhoun, H., Levac, D., *et al.* (2018). PRISMA Extension for Scoping Reviews (PRISMA–ScR): Checklist and Explanation. *Annals of Internal Medicine*, 169(7), 467–473. https://doi.org/10.7326/M18-0850.

Truchon, M., Schmouth, M., Côté, D., Fillion, L., Rossignol, M. & Durand, M. (2012). Absenteeism Screening Questionnaire (ASQ): A New Tool for Predicting Long–term Absenteeism Among Workers with Low Back

Pain. *Journal of Occupational Rehabilitation*, 22(1), 27–50. https://doi.org/10.1007/s10926-011-9318-0.

Tsang Chi Chung, R., Lau Sau Ying, J., Kwong So Fong, S., So Ming Loi, E., Law Ka Yee, R., Wong Fu Yan, T., *et al.* (2017). Reliability, Construct and Predictive Validity of the Hong Kong Chinese Orebro Musculoskeletal Pain Screening Questionnaire. *Journal of Occupational Rehabilitation*, 27(4), 584–592. https://doi.org/10.1007/s10926-016-9689-3.

Tsang, R.C.C., Lee, E.W.C., Lau, J.S.Y., Kwong, S.S.F., So, E.M.L., Wong, T.F.Y., *et al.* (2019). Development of a short form of the Hong Kong Chinese Orebro musculoskeletal pain screening questionnaire. *Hong Kong Physiotherapy Journal*, 39(1), 57–66. https://doi.org/10.1142/S1013702519500057.

Turner, J.A., Shortreed, S.M., Saunders, K.W., LeResche, L., Berlin, J.A. & Korff, M.V. (2013). Optimizing prediction of back pain outcomes. *Pain*, 154(8), 1391–1401. https://doi.org/10.1016/j.pain.2013.04.029.

Urrútia, G., Burton, A.K., Morral Fernández, A., Bonfill Cosp, X. & Zanoli, G. (2004). Neuroreflexotherapy for non–specific low–back pain. *Cochrane Database of Systematic Reviews*, 2011(2), CD003009. https://doi.org/10.1002/14651858.CD003009.pub2.

Walker–Dilks, C., Wilczynski, N.L. & Haynes, R.B. (2008). Cumulative Index to Nursing and Allied Health Literature search strategies for identifying methodologically sound causation and prognosis studies. *Applied Nursing Research*, 21(2), 98–103. https://doi.org/10.1016/j.apnr.2006.04.004.

Williams, C.M., Hancock, M.J., Maher, C.G., McAuley, J.H., Lin, C.W.C. & Latimer, J. (2014). Predicting rapid recovery from acute low back pain based on the intensity, duration and history of pain: A validation study. *European Journal of Pain*, 18(8), 1182–1189. https://doi.org/10.1002/j.1532-2149.2014.00467.x.

Wolff, R.F., Moons, K.G.M., Riley, R.D., Whiting, P.F., Westwood, M., Collins, G.S., *et al.* (2019). PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Annals of Internal Medicine*, 170(1), 51–58. https://doi.org/10.7326/M18-1376.

Wolff, S.V., Willburger, R., Hallner, D., Rusu, A.C., Rusche, H., Schulte, T., *et al.* (2018). Avoidance–Endurance Fast–Screen (AE–FS). Schmerz, 32(4), 283–292. https://doi.org/10.1007/s00482-018-0310-x.

Appendix A: Search strings

5.1. PubMed

Search	Query	Results
#12	Search: (#6 AND #5) Filters: Meta-Analysis, Systematic Review, Humans, English, from 2000–2024 Sort by: Most Recent	1,720
#11	Search: (#6 AND #5) Filters: Meta-Analysis, Systematic Review, Humans, from 2000–2024 Sort by: Most Recent	1,748
#10	Search: (#6 AND #5) Filters: Meta-Analysis, Systematic Review, from 2000–2024 Sort by: Most Recent	2,016
#9	Search: (#6 AND #5) Filters: Meta-Analysis, from 2000–2024 Sort by: Most Recent	827
#8	Search: (#6 AND #5) Filters: from 2000–2024	33,531
#7	Search: #6 AND #5	37,015
#6	Search: #1 OR #2 OR #3 OR #4	154,049
#5	Search: (((((((Validat* OR Predict*[Title] OR Rule*) OR (Predict* AND (Outcome* OR Risk* OR Model*)) OR ((History OR Variable* OR Criteria OR Scor* OR Characteristic* OR Finding* OR Factor*) AND (Predict* OR Model* OR Decision* OR Identif* OR Prognos*)) OR (Decision* AND (Model* OR Clinical* OR Logistic Models/)) OR (Prognostic AND (History OR Variable* OR Criteria OR Scor* OR Characteristic* OR Finding* OR Factor* OR Model*))) OR ("Stratification" OR "ROC Curve" [Mesh] OR "Discrimination" OR "Discriminate" OR "c-statistic" OR "c statistic" OR "Area under the curve" OR "AUC" OR "Calibration" OR "Indices" OR "Algorithm" OR "Multivariable"))))))	7,578,220
#4	Search: (spin* pain[Title/Abstract]) OR (Low Back Pain[MeSH Terms]) OR (low back pain[Title/Abstract])	47,463
#3	Search: (neuropathic pain [Title/Abstract]) OR (Neuralgia[MeSH Terms])	42,265
#2	Search: (Recess stenosis[Title/Abstract]) OR (Recess stenoses[Title/Abstract]) OR ("spinal stenoses"[Title/Abstract:~2]) OR ("spinal stenosis"[Title/Abstract:~2]) OR ("neurogenic claudication"[Title/Abstract:~2]) OR ("lumbar stenoses"[Title/Abstract:~2]) OR ("lumbar stenosis"[Title/Abstract:~2]) OR (Spinal Stenosis[MeSH Terms])	12,816
#1	Search: (Intervertebral Disc Displacement[MeSH Terms]) OR (Sciatic Neuropathy[MeSH Terms]) OR (Radicul*[Title/Abstract]) OR (Radiculopathy[MeSH Terms]) OR (Sciatic*[Title/Abstract]) OR (Sciatica[MeSH Terms])	72,609

5.2. Embase

No.	Query	Results
#31	#30 AND ‘human’/de	7,762
#30	#29 AND (2000:py OR 2001:py OR 2002:py OR 2003:py OR 2004:py OR 2005:py OR 2006:py OR 2007:py OR 2008:py OR 2009:py OR 2010:py OR 2011:py OR 2012:py OR 2013:py OR 2014:py OR 2015:py OR 2016:py OR 2017:py OR 2018:py OR 2019:py OR 2020:py OR 2021:py OR 2022:py OR 2023:py OR 2024:py OR 2025:py) AND (‘meta analysis’/de OR ‘systematic review’/de)	7,959
#29	#24 AND (#25 OR #26 OR #27)	73,553
#28	predict*:ti,ab OR ‘methodology’/exp OR validat*:ti,ab	10,833,771
#27	validat*:ti,ab	1,182,099
#26	‘methodology’/exp	8,259,141
#25	predict*:ti,ab	2,985,802
#24	#7 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20 OR #21 OR #22 OR #23	251,383
#23	‘low back pain’/mj OR ‘spinal pain’/exp OR ‘spin* pain’:ti,ab OR ‘low* back pain’:ti,ab	65,531
#22	‘low* back pain’:ti,ab	52,918
#21	‘spin* pain’:ti,ab	3,802
#20	‘spinal pain’/exp	4,508
#19	‘low back pain’/mj	33,127
#18	‘neuropathic pain’/exp OR ‘neuralgia’/de OR ‘neuropathic pain’:ti,ab	65,100
#17	‘neuropathic pain’:ti,ab	39,787
#16	‘neuralgia’/de	11,282
#15	‘neuropathic pain’/exp	45,553
#14	#8 OR #9 OR #10 OR #11 OR #12 OR #13	20,980
#13	(neurogenic NEAR/2 claudication):ti,ab	1,337
#12	(lumbar NEAR/2 stenos?s):ti,ab	6,724
#11	(spinal NEAR/2 stenos?s):ti,ab	11,091
#10	‘recess stenos?s’:ti,ab	298
#9	‘neurogenic claudication’/exp	248
#8	‘vertebral canal stenosis’/exp	18,034
#7	#1 OR #2 OR #3 OR #4 OR #5 OR #6	127,556
#6	‘intervertebral disk hernia’/exp	32,258
#5	‘sciatica’/exp	3,463

No.	Query	Results
#4	radicul*:ti,ab	25,328
#3	‘radiculopathy’/exp	48,479
#2	sciatic*:ti,ab	41,316
#1	‘sciatic neuropathy’/exp	9,135

5.3. Epistemonikos

#	Query	Date
#4	The below search, dates from 2000 to 2025, filters: systematic review	907
#3	#1 AND #2	07-10-2024 03:32:01 +02:00
#2	(title:(“prediction model”) OR abstract:(“prediction model”)) OR (title:(predict*) OR abstract:(predict*)) OR (title:(“prognostic model”) OR abstract:(“prognostic model”)) OR (title:(prognos*) OR abstract:(prognos*)) OR (title:(model*) OR abstract:(model*)) OR (title:(“ROC curve”) OR abstract:(“ROC curve”)) OR (title:(discriminat*) OR abstract:(discriminat*)) OR (title:(“c-statistic”) OR abstract:(“c-statistic”)) OR (title:(“c statistic”) OR abstract:(“c statistic”)) OR (title:(“area under the curve”) OR abstract:(“area under the curve”)) OR (title:(AUC) OR abstract:(AUC)) OR (title:(calibration) OR abstract:(calibration)) OR (title:(algorithm) OR abstract:(algorithm)) OR (title:(multivariable) OR abstract:(multivariable))	07-10-2024 03:30:15 +02:00
#1	(title:(sciatic*) OR abstract:(sciatic*)) OR (title:(radicul*) OR abstract:(radicul*)) OR (title:(“disc herniation”) OR abstract:(“disc herniation”)) OR (title:(“spinal stenosis”) OR abstract:(“spinal stenosis”)) OR (title:(“neurogenic claudication”) OR abstract:(“neurogenic claudication”)) OR (title:(“recess stenosis”) OR abstract:(“recess stenosis”)) OR (title:(“lumbar stenosis”) OR abstract:(“lumbar stenosis”)) OR (title:(“spinal stenosis”) OR abstract:(“spinal stenosis”)) OR (title:(“spinal pain”) OR abstract:(“spinal pain”)) OR (title:(“low back pain”) OR abstract:(“low back pain”)) OR (title:(“neuropathic pain”) OR abstract:(“neuropathic pain”)) OR (title:(“spine pain”) OR abstract:(“spine pain”))	07-10-2024 03:23:41 +02:00

5.4. CINAHL

#	Query	Limiters/Expanders	Last Run Via	Results
S30	S28 OR S29	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	1,522

#	Query	Limiters/Expanders	Last Run Via	Results
S29	S24 AND S25	Limiters – Publication Date: 20000101–20241231; Human; Publication Type: Meta Analysis Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	504
S28	S24 AND S25	Limiters – Publication Date: 20000101–20241231; Human; Publication Type: Systematic Review Expanders – Apply equivalent subjects Narrow by Language0: – english Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	1,448
S27	S24 AND S25	Limiters – Publication Date: 20000101–20241231 Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	32,964
S26	S24 AND S25	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	35,482
S25	S21 OR S22 OR S23	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	3,029,634
S24	S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14 OR S15 OR S16 OR S17 OR S19 OR S20	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	64,537
S23	TI outcome OR AB outcome OR MW outcome	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	907,663

#	Query	Limiters/Expanders	Last Run Via	Results
S22	TI diagnos* OR AB diagnos* OR MW diagnos*	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	1,264,140
S21	(MH "Study Design+")	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	1,712,679
S20	TI "spin* pain" OR AB "spin* pain"	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	1,215
S19	TI "lumbar pain" OR AB "lumbar pain"	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	471
S17	TI lumbar pain OR AB "low* backpain"	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	1,135
S16	TI "low* backpain" OR AB "low* backpain"	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	6
S15	TI "low* back pain" OR AB "low* back pain"	Expanders – Apply equivalent subjects Search modes – Find all my search terms	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	21,745
S14	(MH "Back Pain")	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	12,238

#	Query	Limiters/Expanders	Last Run Via	Results
S13	(MH "Low Back Pain")	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	23,254
S12	(MH "Neuralgia")	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	5,223
S11	TI "neuropathic pain" OR AB "neuropathic pain"	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	7,822
S10	TI "neurogenic claudication" OR AB "neurogenic claudication"	Expanders – Apply equivalent subjects Search modes – Find all my search terms	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	290
S9	TI lumbar Stenos?s OR AB Lumbar Stenos?s	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	2,430
S8	TI Recess Stenos?s OR AB Recess Stenos?s	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	78
S7	TI Spinal Stenos?s OR AB Spinal Stenos?s	Expanders – Apply equivalent subjects Search modes – Find all my search terms	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	3,700
S6	(MH "Spinal Stenosis")	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	3,211

#	Query	Limiters/Expanders	Last Run Via	Results
S5	(MH "Intervertebral Disk Displacement")	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	5,054
S4	TI Radicul* OR AB Radicul*	Expanders – Apply equivalent subjects Search modes – Find all my search terms	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	4,920
S3	(MH "Radiculopathy")	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	2,625
S2	(MH "Sciatica")	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	1,825
S1	TI sciatic* OR AB sciatic*	Expanders – Apply equivalent subjects Search modes – Find all my search terms	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	4,937

Appendix B:
Research Question 1 – Further elaboration on the reasons for exclusion

Author	Reason
Fu *et al.*, (2024)	Studies had to include a minimum of 75% of their participants with chronic low back pain. The review excluded studies evaluating persons with LBP caused by disc herniations.
Feller *et al.*, (2024)	The review included prognostic models related to leg pain evaluated outcomes following surgery (Fritzell, Mesterton & Hagg, 2022; Staartjes *et al.*, 2019) or diagnostic models (Stynes *et al.*, 2018).

Author	Reason
	Gabel *et al.*, (2011) evaluated the Original Orebo Musculoskeletal Pain Questionnaire, but it did not include radicular or sciatica or leg pain in the validation study.
	Jellema *et al.*, (2007) evaluated the Orebo, Low Back Pain Perception Scale (LBPPS) and a prediction rule developed by the authors themselves, but it did not mention radicular or sciatica or leg pain.
	Context: The Da Silva model was developed from the Hancock CPR. Williams *et al.*, (2014) externally validated the Hancock CPR (Hancock *et al.*, 2009). Persons with sciatica were included in the validation but not in the development sample. The authors acknowledged that the prediction rule may have limited applicability to persons with sciatica or radicular pain. The authors did not mention the proportion of persons with sciatica or leg pain. In Silva *et al.*, (2019) the authors did not mention the proportion of participants with radicular or sciatica or leg pain.
	Hancock *et al.*, (2009) looked at predicting recovery in persons with acute low back pain with or without leg pain using the Hancock CPR; however, the proportion of patients with leg pain was not mentioned.
	Both Hazard *et al.*, (1996) and Hazard *et al.*, (1997) evaluated the Vermont Disability Prediction Questionnaire, but the authors did not mention the proportion of persons with radicular or sciatica or leg pain.
Silva *et al.*, (2022)	Heneweer *et al.*, (2007) evaluated the Acute Low Back Pain Screening Questionnaire (ALBPSQ), but the authors did not mention the proportion of participants with radicular or sciatica or leg pain.
	Kongsted *et al.*, (2016) compared the SBT with clinicians' expectations. They included participants with non-specific LBP or lumbar nerve root involvement (based on usual clinical practice for diagnostic triage). However, the authors did not mention the proportion of participants with radicular, sciatic or leg pain.
	Law *et al.*, (2013) evaluated the Orebo Musculoskeletal questionnaire, but only 39.0% of the participants had back pain with leg pain.
	Mehling *et al.*, (2015) evaluated the SBT, but the authors did not mention the proportion of participants with radicular, sciatica or leg pain.
	In Mehling *et al.*, (2015), only 27% of the participants had sciatic pain below the knee.
	Traeger *et al.*, (2016) evaluated the PICKUP model, but only 24% of the participants had leg pain in the development sample, and 19% had leg pain in the external validation sample.
	Truchon *et al.*, (2012) evaluated the Absenteeism Screening Questionnaire (ASQ), but the authors did not mention radicular or sciatica or leg pain.
	Both Tsang *et al.*, (2019) and Tsang Chi Chung *et al.*, (2017) evaluated the Hong Kong Chinese version of the Orebo Musculoskeletal pain screening questionnaire.
	Wolff *et al.*, (2018) evaluated the Avoidance-Endurance Fast-Screen (AE-FS) but the manuscript did not mention radicular, sciatic or leg pain.

Author	Reason
Tagliaferri *et al.*, (2020)	Azimi *et al.*, (2015) evaluated a model to predict who will experience recurrent lumbar disc herniations, yet this was not associated with any treatment.
	Barons *et al.*, (2013) evaluated a prognostic model of treatment outcomes (cognitive behavioural therapy) but only in persons with non-specific low back pain.
	Gal *et al.*, (2014) developed a prediction model of treatments for low back pain patients only.
Lheureux & Berquin, (2019) McIntosh *et al.*, (2018) Patel *et al.*, (2013)	None of the models evaluated persons with lumbar-related leg pain.

Author	Reason
Karran *et al.*, (2017a)	Kongsted *et al.* (2016) compared the SBT with clinicians' expectations. They included participants with non-specific LBP or lumbar nerve root involvement (based on usual clinical practice for diagnostic triage). However, the authors did not mention the proportion of participants with radicular, sciatic, or leg pain.
	In Beneciuk *et al.*, (2013), 66.4% of the participants had LBP with leg pain radiation.
	In Field & Newell, (2012), 36.6% (n = 404) of the participants had radiating leg pain with LBP.
	In the SBT derivation study (n=851), only 62.3% of the participants had radiation pain in the leg, and only 32.1% had radiating pain below the knee (Hill *et al.*, 2011).
	Newell, Field & Pollard, (2015) evaluated the SBT but only 45.4% of the participants (n=749) had radiation pain in the leg.
	Gabel *et al.*, (2011) evaluated the Original Orebro Musculoskeletal pain questionnaire but did not mention radicular, sciatica, or leg pain in the validation study.
	Law *et al.*, (2013) evaluated the Orebo Musculoskeletal pain questionnaire, but only 39.0% of the participants had back pain with leg pain.
	Nonclercq & Berquin, (2012) evaluated the French version of the Orebo musculoskeletal pain questionnaire, but there is no mention of radicular or sciatica or leg pain.
	Schmidt *et al.*, (2016) evaluated the German Orebo Musculoskeletal pain questionnaire, but there was no mention of radicular, sciatic or leg pain.
	Heneweer *et al.*, (2007) evaluated the Acute Low Back Pain Screening Questionnaire (ALBPSQ) but did not mention radicular, sciatic or leg pain.
	Within Grotle, Vøllestad and Brox, (2006), in the acute LBP group, only 20% of the participants had LBP with leg pain, and 56% in the chronic LBP group had leg pain.
	Williams *et al.*, (2014) sought to externally validate the Hancock CPR (Hancock *et al.*, 2009). Persons with sciatica were included in the validation but not in the development sample. The authors acknowledge that the prediction rule may have limited applicability to persons with sciatica or radicular pain. The proportion of persons with sciatica or leg pain was not provided.
	Truchon *et al.*, (2012) evaluated the Absenteeism Screening Questionnaire (ASQ), but the manuscript did not mention radicular pain, sciatic or leg pain.
	Jellema *et al.*, (2007) evaluated the Orebo, Low Back Pain Perception Scale (LBPPS) and a prediction rule developed by the authors themselves. The manuscript did not mention radicular pain, sciatic or leg pain.
	Both Hazard *et al.*, (1996) and Hazard *et al.*, (1997) evaluated the Vermont Disability Prediction Questionnaire, but the manuscript did not mention radicular pain or sciatica or leg pain.
	Turner *et al.*, (2013) evaluated the Chronic pain risk score. Only 21.4% of the participants had pain below the knee.
	Shaw, Pransky & Winters, (2009) evaluated the Back Disability Risk Questionnaire. Only 7.0% of the participants had leg pain radiating below the knee.

Author	Reason
Haskins, Rivett & Osmotherly, (2012)	This review included 25 CPRs on the physiotherapy management of LBP, two of which were prognostic models. George, Bialosky & Donald, (2005) evaluated the data from 28 participants after screening 202 participants. 21 out of 28 participants had both leg and back pain. The study evaluated whether the centralisation phenomenon and fear avoidance belief predict pain and disability at 6 months in persons with acute LBP/leg pain after attending 4 weeks of physiotherapy sessions
	Hancock *et al.*, (2009) looked at predicting recovery in persons with acute LBP with or without leg pain; however, the proportion of patients with leg pain was not provided.
Stanton *et al.*, (2010)	Congcong, Yong & Kian, (2009) included 129 patients with LBP who were referred for physiotherapy. All participants had a diagnosis related to the lumbosacral spine and had a chief complaint of pain and/or numbness in the lumbar spine, buttock, and/or lower extremity. However, only 24.8% of the participants had neurological involvement and 35.7% had pain radiation below the knee.
	Flynn *et al.*, (2002) excluded persons with signs consistent with nerve root compression (positive straight leg raise at 45°, or diminished lower extremity strength, sensation, or reflexes). The authors developed the original 5-item rule.
	Fritz, Childs & Flynn, (2005) evaluated the 2-item version of the Flynn rule (mentioned above). Participants had a primary complaint of LBP with or without referral into the lower extremity, but the proportion of patients with leg pain was not reported.
	Hicks *et al.*, (2005) included fifty-four patients with complaints of LBP with or without leg pain. In the abstract, it mentioned that persons with non-radicular LBP were studied. They excluded persons with 2 or more signs of nerve root compression: diminished lower-extremity strength, sensation, or reflexes.
May & Rosedale, (2009)	Fritz *et al.*, (2007) included participants aged between 18 and 60 years, with pain and/or numbness extending distal to the buttock in the past 24 hours, Oswestry score 30%, signs of nerve root compression [positive straight leg raise (reproduction of symptoms at <45°), or reflex, sensory, or muscle strength deficit]. However, only 76.5% (49/64) of the patients had symptoms distal to the knee.
Kent & Keating, (2008)	It included studies with less than 15% of participants presenting with neuro-compressive symptoms.
Chiodo & Haley, (2024)	The included primary studies either evaluated only cases of LBP or excluded cases with radiculopathy.

Appendix C
PROBAST Assessment

Kovacs *et al.*, (2012) model		
Step 1 – All prediction models of treatment outcomes for lumbar-related leg pain		
Step 2 – Development only		
Step 3		
Domain 1. Participants.		
1.1	PN	Routine care registries
1.2	PN	Pg. 1011, second paragraph. This registry was a neuroreflexology therapy (NRT) marketing surveillance study. 94.8% of the participants had NRT (table 1), while only 15% had physiotherapy. This does not reflect usual clinical care for such conditions.
ROB – high risk of bias		
Applicability – Low (they match the review question)		
Domain 2. Predictors.		
2.1	PN	Since the data came from routine care data registry.
2.2	NI	However, predictors were measured a long time (3 months) before the outcome occurred, so they were blinded to the outcome. Therefore, the domain can still be rated as low risk of bias.
2.3	N	It included "have been treated with neu-reflexology." The treatment is not available at the time the model would be applied, making the model unusable. Also, this variable (treated with NRT) has a very high association with the outcome (coefficient = 1.47) compared to the strength of the other coefficients.
ROB – high risk of bias		
Applicability – low concern. However, the MRI scanner might not be available in the 7 primary care centres.		
Domain 3. Outcome.		
3.1	Y	They used VAS for LBP and leg pain and RMQ
3.2	Y	The standard outcome definition was VAS leg pain.
3.3	Y	They included baseline leg pain, which is partially independent of the outcomes.
3.4	PY	Outcome definition and determination were the same for all participants.
3.5	Y	The predictors in the final model were determined by independent statisticians who had no contact with the clinicians.
3.6	Y	Justified reasons for the 3-month follow-up period.
ROB – low risk of bias.		
Applicability – low concern.		

Domain 4. Analysis		
4.1	PY	Candidate predictors included in the model: Sex Age Baseline severity LBP Baseline severity leg pain Roland Morris Duration of current episodes – acute, subacute, chronic. Employment status – passive, working, receiving financial aid Recruitment setting – primary, specialised History of lumbar surgery Diagnosis of FBSS Diagnostic tests undertaken at the moment during the study period – x-ray, CT, MRI, emg, other Findings in imaging procedures – disc degeneration, facet jt, scoliosis, spondylolisthesis, spinal stenosis, annular tears, disc protrusion, disc herniation, lumbarisation of S1, sacralization of L5, other findings, no findings Treatments used – drugs, other, PT, NRT, surgery. 37 variables for 833 events. Therefore 22.5 events per variable (EPV).
4.2	Y	All continuous predictors were kept as continuous numerical variables.
4.3	PY	Despite being a routine care registry, the model included data from all the participants with leg pain.
4.4	PY	The authors used multiple imputations and included variables that were selected as predictors in at least 2 of the 5 imputed datasets. Rubin rules were used to average regression coefficients.
4.5	Y	Predictors were not selected on the basis of univariable analysis prior to multivariable modelling.
4.6	PY	For patients with subsequent episodes of LBP, it was decided that only the data from the first episode would be analysed. Patients with a baseline score below the MCIC for a given variable, except those who worsened at discharge, were excluded from the analysis.
4.7	N	Discrimination was provided for the model on leg pain as the AUC = 0.655. However, the measure of calibration was not provided. Instead, they provided the p-value (0.156) of the Hosmer-Lemeshow test as the only measure to assess the calibration of the model. However, this p-value indicates neither the presence nor the magnitude of nay miscalibration.
4.8	PY	The sample size was large (n=3,359), the model had an EPV of 22.5, and the authors conducted bootstrapping.
4.9	Y	Both predictors and coefficients correspond to the reported results of the multivariable regression model.
ROB – High. No actual value of calibration is provided.		

Methodological summary: The authors first conducted univariate analysis followed by multivariate analyses (providing discrimination and calibration and performing bootstrapping). The predictive factors in the multivariate analysis were discussed with experts. The algorithm included the StarT back and the variables included in the model (providing sensitivity, specificity, and likelihood ratios). Hence, they did a binary logistic regression model (whether or not people were referred to specialist services) and then integrated these variables into an algorithm incorporating the STarT back tool. Therefore, instead of individualised predictions, the algorithm provides group-based predictions, which are less precise for the individual and assume that the entire group has similar prognosis.

Algorithm used within SCOPIC trial (Konstantinou *et al.*, 2020; Konstantinou *et al.*, 2019)		
Step 1 – All prediction models of treatment outcomes for lumbar-related leg pain		
Step 2 – Development only		
Step 3		
Domain 1. Participants.		
1.1	PY	The data was taken from the ATLAS, which was a prospective cohort study and retrospectively analysed.
1.2	Y	It included patients with back and leg pain who were consulting by their GP.
ROB – low		
Applicability – Low (they match the review question)		
Domain 2. Predictors.		
2.1	Y	It is mentioned in the Additional file 1.
2.2	NI	Since this is a retrospective analysis, the domain can still be rated as low risk of bias.
2.3	Y	All variables were available at the time the model is intended to be used.
ROB – low. Although 2.2 is rated as NI, the algorithm developed used retrospective cohort data, and hence, it can still be rated as low risk of bias.		
Applicability – low concern.		
Domain 3. Outcome.		
3.1	Y	The outcome in the model was binary, i.e. referral or no referral to specialist services.
3.2	N	Step 2c: Identifying patients for fast-track referral to spinal specialist services: algorithm design. In this step of the iterative process, we investigated a number of possibilities in terms of combinations of factors from the clinical assessment and information on risk of poor prognosis, using the STarT Back Tool score, for identifying which patients with sciatica to refer or 'fast-track' to spinal specialist services.
3.3	Y	Predictors and outcomes were unrelated.
3.4	Y	
3.5	Y	
3.6	Y	The time interval was 1 year.
ROB – high.		
Applicability – low concern. The outcome definition is still within the review's eligibility criteria since epidural steroid injections and multidisciplinary rehabilitation can be provided within specialist services.		

Domain 4. Analysis		
4.1	PN	57 out of 429 were referred to the specialist services. They included a total of 22 candidate predictors. EPV = 2.6 + StarT back variables.
4.2	PY	Continuous variables were handled appropriately. The risk categories of the StarT back tool were originally validated in primary care LBP cases, but 32% of patients had pain radiating below the knee.
4.3	PY	Patients at low risk were excluded from the analysis, including the STarT back and for onward referral. This is justified since only 1 out of 57 patients was referred to specialist services.
4.4	NI	Table 1 of the article mentions that a small proportion has missing data for some baseline characteristics, but we are not given information about how this is tackled in the model.
4.5	PY	Univariate analysis was done, but the choice of predictors was not based on it.
4.6	NI	There is no mention of complexities, apart from Table 2, item b (in the footer).
4.7	PY	Discrimination (AUC=0.695), which after bootstrapping became 0.678. The calibration slope was 1.0 (0.57 to 1.43). The authors then provide information on the overall algorithm's sensitivity, specificity, and positive and negative likelihood ratio.
4.8	Y	Bootstrapping was done for the model.
4.9	NI	The coefficients and variables in the final model do not account for the entire list of variables included in the final algorithm.
ROB – Unclear.		