Do intolerance of uncertainty and resource consumption actually influence the therapeutic strategies chosen by physiotherapists?

Clémence Brun, Oulmann Zerhouni, Alexis Akinyemi, Laurène Houtin, Philippe Meidinger, Jimmy Antunes, Richard Monvoisin, Nicolas Pinsault

Do intolerance of uncertainty and resource consumption actually influence the therapeutic strategies chosen by physiotherapists?

Running title: Therapeutic strategies chosen by physiotherapists

Clémence Brun[1,2], Oulmann Zerhouni[3], Alexis Akinyemi[1,4], Laurène Houtin[1,4], Philippe Meidinger[2], Jimmy Antunes[2], Richard Monvoisin[2], Nicolas Pinsault[2]

AD-HOC Lab,

23 rue Hispano Suiza,

92270 Bois-Colombes, France

Author note

Correspondence concerning this article should be addressed to Clémence Brun, AD-HOC Lab, 23 rue Hispano Suiza, 92270 Bois-Colombes, France; Université Grenoble Alpes, TIMC-IMAG UMR CNRS 5525, ThEMAS team, Domaine de la Merci, 5 Avenue du Grand Sablon, 38700 La Tronche, France. Email: clemence.brun@adhoclab.fr.

[1]: AD-HOC Lab, 23 rue Hispano Suiza, 92270 Bois-Colombes, France

[2]: Univ. Grenoble Alpes, CNRS, UMR 5525, VetAgro Sup, Grenoble INP, TIMC, 38000 Grenoble, France

[3]: University of Rouen Normandy, Centre de Recherche sur les Fonctionnements et Dysfonctionnements Psychologiques (CRFDP, EA 7475) Batiment Satie - Bureau S206

[4]: Laboratoire Parisien de Psychologie Sociale, EA 4386 (équipe PS2C), 200 av. de la République - 92001 Nanterre Cedex, France

ORCID:

Clémence Brun: 0000-0001-6880-6674

Oulmann Zerhouni: 0000-0002-4026-1921

Alexis Akinyemi: 0000-0002-6027-675X

Laurène Houtin: 0000-0002-7619-5639

Nicolas Pinsault: 0000-0001-5034-0802

Philippe Meidinger: 0000-0002-7243-3839

Jimmy Antunes: 0009-0008-7232-0464

Declarations

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

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

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed Consent: N.A.

Acknowledgment: We would like to thank Pauline Girard, Amélie Kechichian, Thomas Nivert and Geoffrey Coisnon for reviewing the clinical cases used in this study and for their advice.

Abstract

The medical environment is prone to diagnostic errors, which can have serious consequences for patients' health status. However, healthcare professionals (HCPs) are often overconfident in their choices. In this context, there is a growing interest in the influence of variables that may negatively influence the reasoning of HCPs, namely intolerance of uncertainty (IU) and resource consumption. We challenged 127 physiotherapists to play a simulation game and choose a therapeutic strategy for three cases of low back pain of increasing difficulty (easy, medium, difficult). We measured their IU level and their resource consumption (*i.e.*, requests for test results for each case). Our results showed that 87.4% of participants chose the most appropriate strategy for the easy case, 46.5% for the medium and 29.1% for the difficult. For the easy case, IU and resource consumption had a negative influence on the chosen therapeutic strategy, which was less adapted. Moreover, resource consumption among highly intolerant individuals had a more adverse influence on the chosen strategy. However, we did not find the same results for the medium and difficult cases. We propose hypotheses to explain this difference (*i.e*., system 1/system 2, consideration of the physiotherapist's speciality, perception of uncertainty). We also outline potential improvements to our tool and suggest future research avenues.

Keywords: critical thinking, clinical reasoning**,** intolerance of uncertainty, physiotherapists, therapeutic strategy, simulation game, virtual patients

Do intolerance of uncertainty and resource consumption actually influence the therapeutic strategies chosen by physiotherapists?

Introduction

The increasing importance of the diagnostic errors problem

The issue of diagnostic errors has become increasingly prominent among medical institutions over the past few decades (Benishek *et al*., 2015; Bonilauri Ferreira *et al*., 2010; Brun *et al*., 2022b; Croskerry, 2009; Crowley *et al*., 2013; Graber, 2022; Harada *et al*., 2021; Iuliucci, 2016; Kourtidis *et al*., 2022; Makridakis *et al*., 2019; Mamede *et al*., 2014; Monteiro *et al*., 2015; Norman & Eva, 2010; Norman *et al*., 2017; Olson *et al*., 2021; Scott *et al*., 2021; Staal *et al.*, 2021, 2022). This is partly due to the fact that these situations are more common than one would think (Makridakis *et al.*, 2019; Peterson *et al*., 2022; Schiff *et al*., 2005; Scott *et al*., 2021), as up to 35% of physicians and 42% of patients reported such errors (Berner & Graber, 2008) and are mostly preventable (Lee *et al*., 2022; Staal *et al*., 2022). Diagnostic errors is an umbrella term that can be commonly defined as "mistakes that result in serious harm, such as death, disability, or additional or prolonged treatment." (Benishek *et al.*, 2015; Blendon *et al*., 2002, as cited in Berner & Graber, 2008, p. 2). Adverse consequences of diagnostic errors are constantly pointed out, as summarised by Durning and colleagues (2020): "reviews of malpractice claims have consistently shown diagnostic error to be the most common contributing factor in settled claims and that diagnostic error claims are the type of claim most likely to result in disability and death". In specialities that provide non-nosological diagnoses (*e.g.*, nursing diagnoses, physiotherapeutic diagnoses), consequences may be similar but in different proportions. That is, deaths are very rare, but errors often lead to additional or prolonged treatment. These

errors could be avoided in 80% of cases, and three-quarters are cognitive and reasoning errors (Nendaz & Perrier, 2012; Scott *et al*., 2021).

Variables of interest likely to influence clinical reasoning

As Noll and colleagues (2001, p. 41) rightly stated, “Because treatment depends on diagnosis, the ability to provide competent patient care depends on the effectiveness and efficiency of clinical reasoning skills”. Clinical reasoning - “deliberate process of critical thinking about a clinical situation to reach a reasonable decision about an outcome, diagnosis, therapeutic action, or resolution of a particular patient problem” - is the thinking process that healthcare professionals (HCPs) follow to reach a diagnosis and treatment proposal (Payan-Carreira *et al*., 2019, as cited in Brun *et al*., 2022b). However, as medical environments are very prone to uncertainty (Benishek *et al*., 2015; Brun *et al*., 2022b; Dahm & Crock, 2022; Patel *et al*., 2022; Simmonds *et al.*, 2012), clinical reasoning is often carried out without a clear knowledge of all the relevant parameters, which can lead to diagnostic uncertainty. Diagnostic uncertainty is the subjective impression that HCPs feel when they cannot offer a correct and clear explanation for the health problem described by their patients (Almond *et al*., 2021). A number of variables and situational characteristics can contribute to diagnostic uncertainty and sometimes diagnostic errors.

HCPs' overconfidence is often mentioned in studies covering factors considered to have a potentially strong negative influence on the occurrence of diagnostic errors (Berner & Graber, 2008; Croskerry & Norman, 2008; Haber *et al*., 2023; Zwaan & Hautz, 2019). Indeed, many studies have shown calibration[1] discrepancies between HCPs' confidence in their diagnosis and the quality of their diagnosis (Berner & Graber, 2008; Høegh-Larsen *et al*., 2023; Norman & Eva, 2010; Zwaan & Hautz, 2019). Confidence in one's diagnosis and

[1] “Calibration is defined as the relationship between the accuracy of a decision and an individual's perception of, and confidence in, that accuracy” (Fernandez-Branson *et al*., 2021, p. 1), “This concept defines the alignment between clinicians’ confidence in their accuracy and their actual diagnostic accuracy” (Staal *et al*., 2023, p. 2).

ability to manage a patient is not necessarily a bad thing, particularly as confidence in one's ability to manage a patient can influence patients' perceptions of their therapeutic relationship with their practitioner, especially if they feel distressed (Brunner *et al*., 2021). Rather, the problem lies in the discrepancies, if any, between practitioners' sense of confidence and the accuracy of their diagnosis, as this may lead practitioners to overlook specific therapies or ignore crucial information (Kourtidis *et al*., 2022; Miller *et al*., 2015; Zwaan & Hautz, 2019). Thus, confidence in one's decisions is increasingly studied in healthcare contexts, such as in the context of vaccination decisions, and needs to be further studied among HCPs within the framework of clinical reasoning (Brun *et al*., 2022a). Authors such as Zwaan and Hautz (2019) consider that the future in this field of research lies in the joint study of uncertainty tolerance in combination with overconfidence and diagnostic accuracy to determine where these calibration errors occur.

Related to the ubiquitous uncertainty in medical environments, poor uncertainty tolerance can have deleterious effects on HCPs (*e.g*., health, reasoning skills), and on patients (especially when HCPs do not communicate their uncertainty to them; Brun *et al.*, 2022b; Dahm & Crock, 2022; Kerr & Thompson, 2022; Patel *et al*., 2022). Direct consequences for patients include overtesting and sometimes inadequate treatment prescriptions (Brun *et al*., 2022b). As such, a high level of intolerance of uncertainty (IU) can have a direct impact on the diagnosis and treatment offered to patients. This is suggested by Simmonds and colleagues (2012) in a study showing that physiotherapists with a high level of IU will tend to suggest to their patients with low back pain a slower and less active return to their physical and professional activities than professionals with a lower level of IU.

Impact of these variables of interest on the therapeutic strategies of physiotherapists

Physiotherapists are among HCPs likely to treat patients following a referral from another HCP (usually a general practitioner). However, they are increasingly required to treat patients as the first HCP they encounter as part of direct access to care, even as they are likely to treat a wide range of conditions, both of which create uncertainty (Infante Guedes *et al*., 2023; Ingram *et al*., 2023). Using the medical prescription, they carry out an assessment of the patient's abilities and deficiencies and set care goals and a suitable rehabilitation program with them (Elvén *et al*., 2018; Ministère de la santé et de la prévention, 2023). Physiotherapists are often consulted for low back pain, a very common problem that most people will experience in their life ("Chronic low back pain prevalence was 4.2% in individuals aged between 24 and 39 years old and 19.6% in those aged between 20 and 59"; Hoy *et al*., 2010; Meucci *et al*., 2015, p. 1; Östhols *et al*., 2019). Highest incidence is during the third decade and overall prevalence increases with age up to 60/65 years, then declines (Hoy *et al*., 2010). While the symptoms are clearly identifiable, the origin is less so: in 85% of cases, low back pain is labelled as non-specific, with only 15% of cases linked to pathologies such as a tumour, osteoporosis and fracture or to nerve-root symptoms (Josephson *et al*., 2011). Furthermore, a study by Fourré and colleagues (2023) showed that a low proportion of physiotherapists are aware of the guidelines for the management of low-back pain, and that this uncertainty leads to guideline-inconsistent behaviour. Authors such as Langridge and colleagues (2015) have previously investigated the clinical reasoning process of physiotherapists and identified seven themes surrounding clinical reasoning in the assessment of patients reporting low back pain (*i.e*., prior thinking, patient interaction, formal testing, safety and accountability, external and internal factors, gut-feeling and time). However, there are few pedagogical tools intended for physiotherapists and fewer tools for

assessing their clinical reasoning[2] (Cowell *et al*., 2023). From the literature on this subject, it seemed to us that research on physiotherapists' clinical reasoning was focused on the identification and classification of the steps of the reasoning process (especially within the framework of the evaluation of students' clinical reasoning skills) with no relation to the relevant resulting variables, namely the diagnosis considered as the most probable for a given situation and above all the therapeutic strategy offered to the patient (Elvén *et al.*, 2018; Hamzeh *et al*., 2021; Langridge *et al*., 2015). In addition, these studies overlooked the impact of contextual factors and factors related to the physiotherapist on clinical decision-making, such as confidence in one's judgment and IU (Holdar *et al*., 2013; Simmonds *et al*., 2012). As such, they ignored the influence of IU on their investigation process and prescriptions, whereas we know that HCPs with a high level of IU tend to prescribe additional tests (*i.e*., overtesting; Brun *et al*., 2022b).

Aim of the study

Our aim was to assess whether physiotherapists propose therapeutic strategies to their patients in line with expert strategies[3] (Meidinger *et al*., 2023). Notably, we wanted to assess the impact of physiotherapists' IU on the therapeutic strategy they chose and determine whether resource consumption interacts with IU to influence the strategy chosen. Secondly, we wanted to check whether physiotherapists exhibited a confidence calibration issue between the strategy offered and its appropriateness for the case (*i.e*., similar to the calibration issues between diagnostics and their quality).

We aimed to test four hypotheses: (H1) A high level of IU and greater resource consumption is associated with the selection of less adapted therapeutic strategies, (H2) A

[2] Some tools can be used to assess diagnostic reasoning in musculoskeletal physiotherapy contexts such as the Diagnostic Thinking Inventory (DTI; Hamzeh *et al*., 2021) and the multidimensional Clinical Reasoning Form (CRF; Cowell *et al*., 2023).

[3] These strategies will be referred to as “most appropriate strategies” in the remainder of the article, as participants were asked to find the “most appropriate strategy” for the cases they were asked to diagnose.

high level of IU is associated with greater consumption of resources (*i.e.*, more requests for test results), (H3) Level of confidence in one's therapeutic strategy will be negatively correlated with the selection of adapted therapeutic strategies (*i.e.*, calibration issue) (H4) IU will be positively associated with confidence in one's diagnoses.

Method

We decided to use the methodology presented by Jette and colleagues (2006) and used by Keller and colleagues (2022). They presented physiotherapists with cases for which they had to propose the most appropriate therapeutic strategy and indicate a suspected diagnosis. We applied the same method, but using a digital tool and adapted it according to our objectives (*e.g.*, adding measurements such as intolerance of uncertainty and resource consumption, changing case format, implementing different difficulty levels in the cases). Among other things, we changed the cases' content by adapting cases from Meidinger and colleagues (2023) on low back pain (LBP) with or without serious spinal pathologies (SSP) as it seemed to be a relevant condition for this study since it is often difficult to characterize the origin of this affection with certainty (Josephson *et al.*, 2011).

Participants

We recruited through email and social networks physiotherapists as a convenience sample. Email diffusion was conducted by submitting the study to an email network of a French university laboratory comprising physiotherapists and contacting the physiotherapists in our network directly. The e-mail presented the study as a clinical reasoning study for physiotherapists and included the estimated time of participation, the game link and the login details for the game. The email also invited physiotherapists to follow a snowballing method and spread the email to their physiotherapy network. The distribution on social networks was

conducted on Facebook, Linkedin and Instagram. On the three networks, we posted a message containing the arguments of the dissemination email, the game link, the login details and invited the physiotherapists to share the posts with their network. In addition, we contacted physiotherapists on Linkedin via private message. We followed the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. Notably, participants' data was anonymised so that they could not be identified, and participation was voluntary.

We recruited 287 physiotherapists. Among the respondents, 160 were excluded because they did not meet our inclusion criteria (*i.e.,* 2 because they were not physiotherapists and 158 because they did not answer all our variables of interest). The final sample was composed of 127 individuals (M_{age} = 33.70, SD_{age} = 8.94, 36 women, 71 men, 1 other, and 1 who did not specify their gender), of which 105 practised in France, 2 in Belgium and one in Ivory Coast ($M_{\text{Years experience}}$ = 10.30, $SD_{\text{Years experience}}$ = 8.95).

Material

Creation of clinical case's content

We decided to present to participants a physiotherapy consultations simulation game featuring three fictitious clinical cases, namely virtual patients (*i.e*. clinical scenarios featured on digital system; Cook & Triola, 2009; Fink *et al*., 2023; Kamath & Ullal, 2023). We chose to assess clinical reasoning through simulation rather than multiple-choice questions because, as Halpern (1998, as cited in Berg *et al.*, 2021) noted, these questions focus on cognitive ability and rote knowledge rather than the analysis and inference inherent in critical thinking and reasoning. Scores on multiple-choice questions would have predicted pure performance but not application in real-life complex situations, such as clinical reasoning in everyday medical practice. Our simulation methodology relies on more ecological measures that

contribute to an assessment closer to real practice in which we focus on a task that requires critical thinking and reasoning skills to be completed. This is in line with the suggestions of Hage and colleagues (2022) that critical thinking should be involved in the examination and management of patients. The three clinical cases used in our study were based on clinical cases from Meidinger and collaborators (2023) which went through a DELPHI assessment process. Three clinical cases from this study were adapted by two physiotherapists that authored our publication, one being an author of the original study[4], so that they would show distinct levels of "difficulty" (see Annexe 11). The introduction of this difficulty variable was meant to address the lack of studies on the reasoning process through different situational levels. As Clarke and colleagues (2023) note, many simulated cases present prototypical situations that fail to provide an opportunity to reason in an uncertain context and to experience the difficulties associated with choosing a strategy from among a wide array of possible ways of managing a case. These methods, based solely on prototypical cases, do not allow the study of the influence of case difficulty on HCPs' choices. As Clarke et colleagues (2023) consider that simple cases are characterised by a clear-cut diagnosis, difficulty level refers in our study to the level of difficulty in making a clear-cut diagnostic choice (Clarke *et al*., 2023). For instance, the "easy" clinical case was supposed to evoke a single diagnosis that practitioners would estimate to have a very high probability of being the right one for the situation (*e.g*., "Diagnostic 1 should be the right one for 99 out of 100 patients"), the clinical case of "medium" level would evoke a few diagnoses of different probabilities (*e.g*., "I think there is a 30% chance that the right diagnosis is Diagnostic 1 and 70% that it is Diagnostic 2") and the "difficult" clinical case would evoke several possible diagnoses with similar probability levels and require more reflection and involvement on the part of practitioners (*e.g*., "Diagnostic 1 has 20% chance of being correct, Diagnostic 2 has 30%, Diagnostic 3 has

[4] The inclusion of one of the authors of the original study helped us to take into account feedbacks from DELPHI experts when designing the new clinical cases.

40% and Diagnostic 4 has 10%"). We chose to display three cases with these levels to reflect the three levels of difficulty ("easy", "complex" and "very complex") identified by physiotherapists in focus groups on low back pain in Josephson and colleagues' study (2011). The purpose of introducing variability in case difficulty was to relate the difficulty level variable to the participants' level of IU and their propensity to find the most appropriate treatment strategy[5] for the patient situation in each case. Clinical cases were written in a question format, meaning participants had to click on questions to get information (*e.g.*, "What is the patient's height and weight?", "1.85m, 80kg"). This format was chosen as it is more in line with medical practice where patients come in reporting a vague problem and give more diagnostic information to practitioners when questioned further. Besides, unlike traditional clinical cases, the progressive addition of information engages HCPs in multifaceted thinking (Cook, 2023). Once written, the clinical cases were reviewed by another physiotherapist and a general practitioner (GP). Reviewers paid attention to the consistency of the information described in the cases and the examination results. The GP consulted radiologists for specific points related to the scintigraphy results. The clinical cases were evaluated a second time once they had been implemented in a digital tool (see next section) by four physiotherapists who had not participated in their design and who judged them according to the same criteria as during the previous round.

Implementation of the clinical cases in the PowerApps tool

Once the clinical cases were written and passed the first validation round, the clinical cases were implemented in Microsoft PowerApps©[6]. The three clinical cases were presented in the application in sequence so that the participants did not choose the case order. The cases

[5] The most appropriate strategy was the one in line with expert strategies, in accordance with the DELPHI conducted by Meidinger and colleagues (2023).

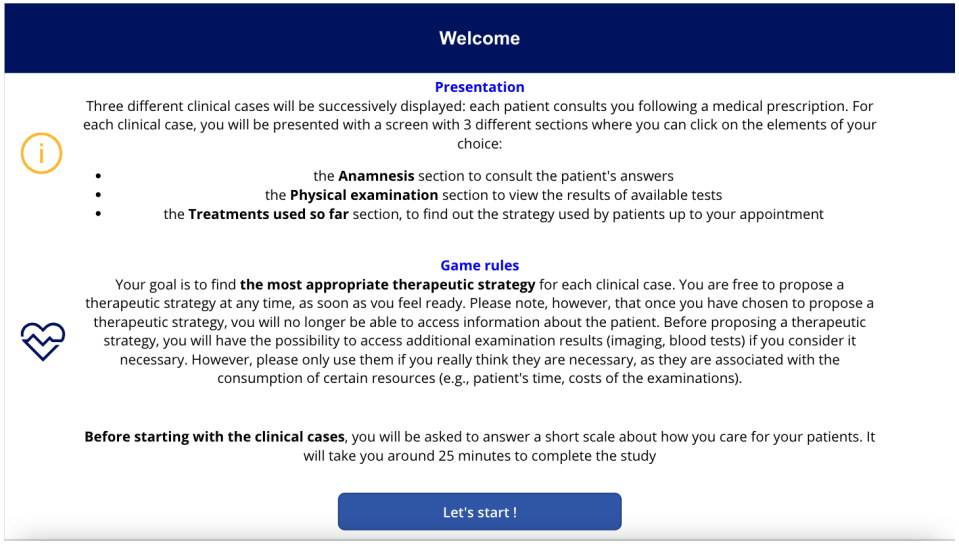
[6] Microsoft PowerApps© is "a suite of applications, services, connectors, and a data platform that provide a rapid development environment for building customized applications" (Microsoft, 2023)

were presented in the following order of difficulty: “easy”, “difficult” and “medium”. This order was chosen based on feedback from reviewers. We chose to present the “easy” case at the beginning of the game to increase participants' engagement in the study. The “difficult” case was placed in the middle of the game because the order initially chosen (*i.e.*, “easy", “medium”, “difficult”) might have influenced participants' choices, as one reviewer told us that she perceived the increasing level of difficulty in the cases and that this influenced her choices for the “difficult” case. Each clinical case was organized in the same following way. An introductory sentence displayed at the top of the screen outlined the very first elements of the case (*i.e.*, the patient's gender, age, and problem) which corresponded to the elements that patients would be likely to communicate first to the physiotherapist in the context of an appointment booking or a brief description of their problem at the beginning of an appointment (*e.g.*, “A 27-year-old male comes in for a consultation for low back pain”). A list of questions that participants could ask patients, organized into three sections (anamnesis[7], physical examination, and treatments), was displayed on the left side of the screen. For each question asked, the answer appeared on the right side of the screen and remained displayed, so that as participants clicked on different questions, they always had the information in front of them as if they were writing a written report on the case presented (see screen “Clinical case screen with “Question/Answer” format” in Table 1). Participants could navigate and see all the questions by scrolling down.

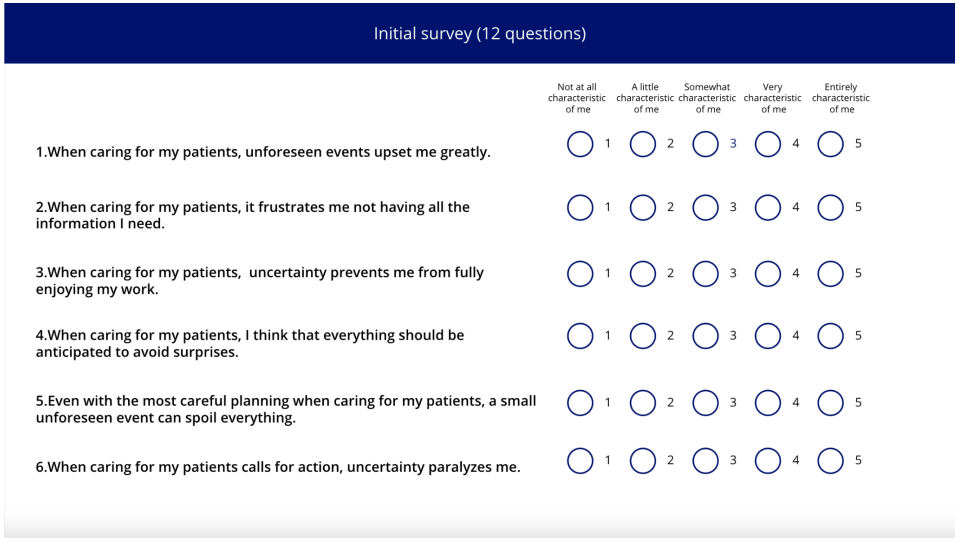
[7] “The anamnesis presents the development of disease as experienced by the patient, *i.e.*, the development of illness. Based on the anamnesis health professionals aim to reach a diagnosis of a disease and a treatment plan (Sadegh-Zadeh, 2012). Hence, anamnesis tracks the disease backwards in time in order to identify it in the present (diagnosis), predict its temporal course (prognosis), and influence the progression of the disease, where possible (treatment).” (Hofmann, 2023, p. 2)

Table 1

Names, visuals and types of game screen

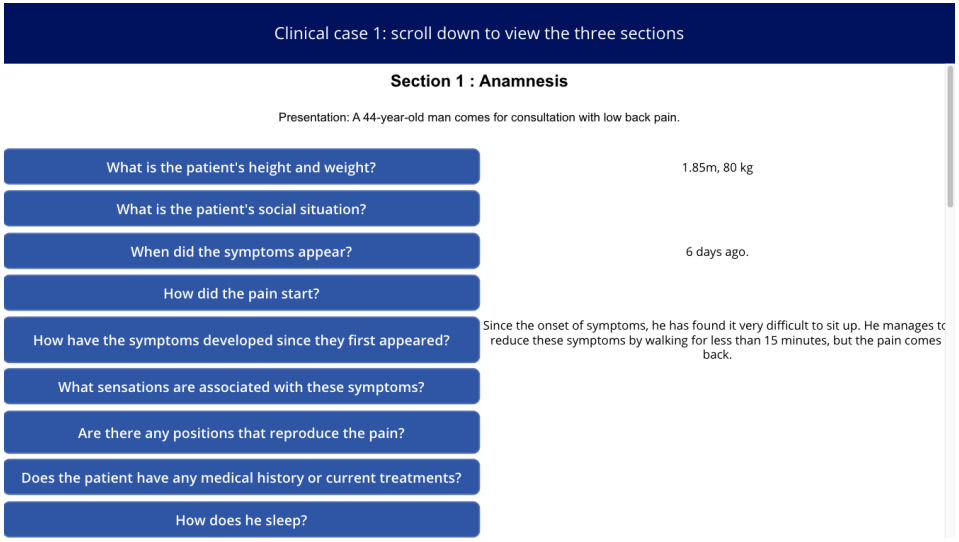
Name	Screen	Screen type[a]
Game presentation screen		Compulsory

Intolerance of uncertainty scale screen



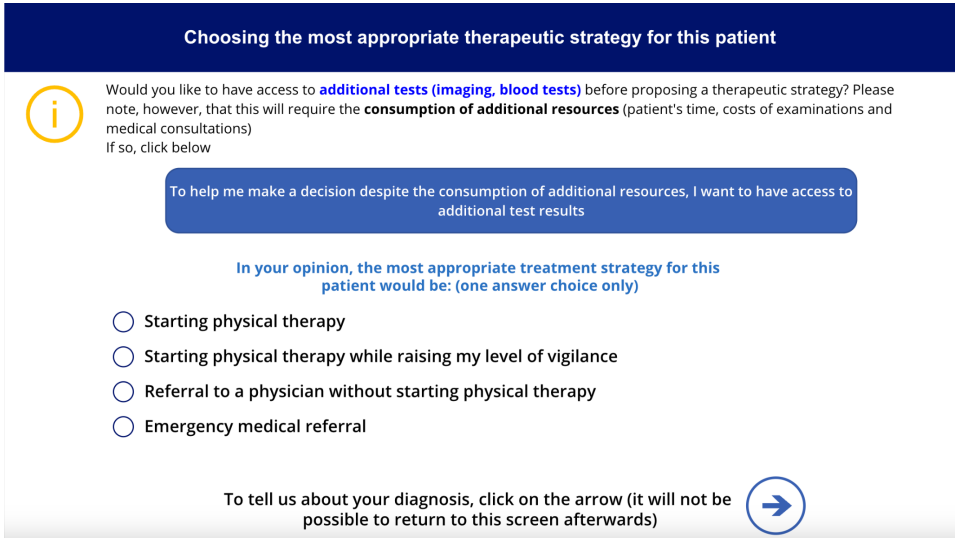
Compulsory

Clinical case screen with"Question/Answer" format



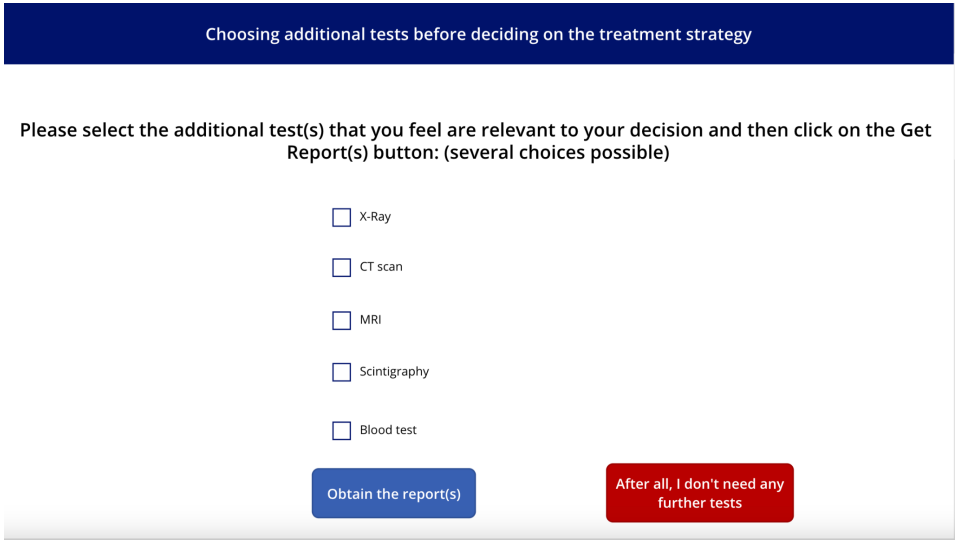
Compulsory

Therapeutic strategy selection screen



Compulsory

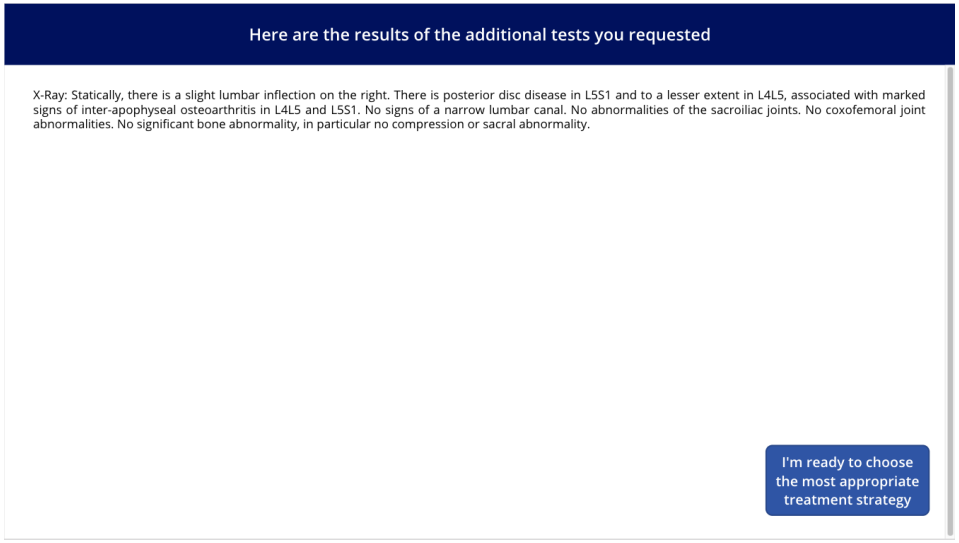
Additional test results selection screen



Optional

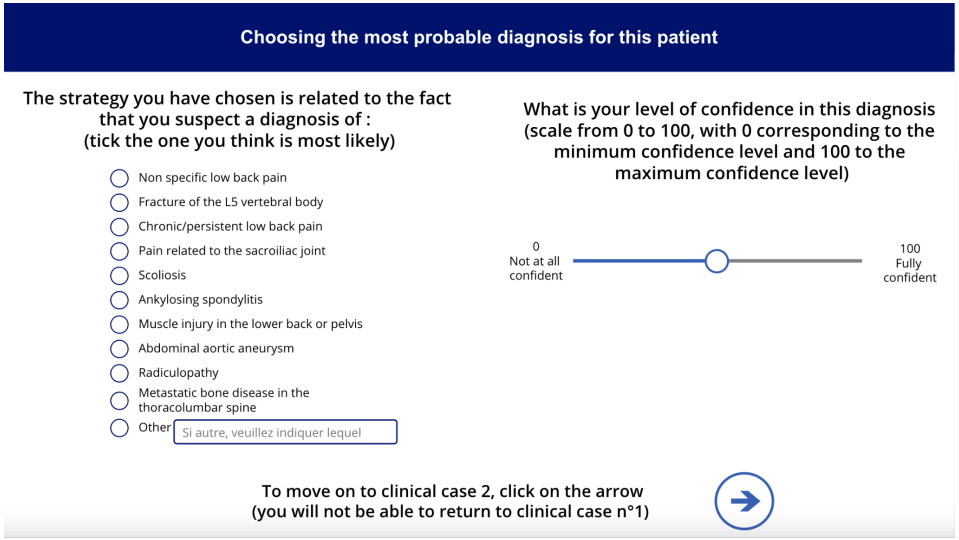
Additional test results screen

Optional

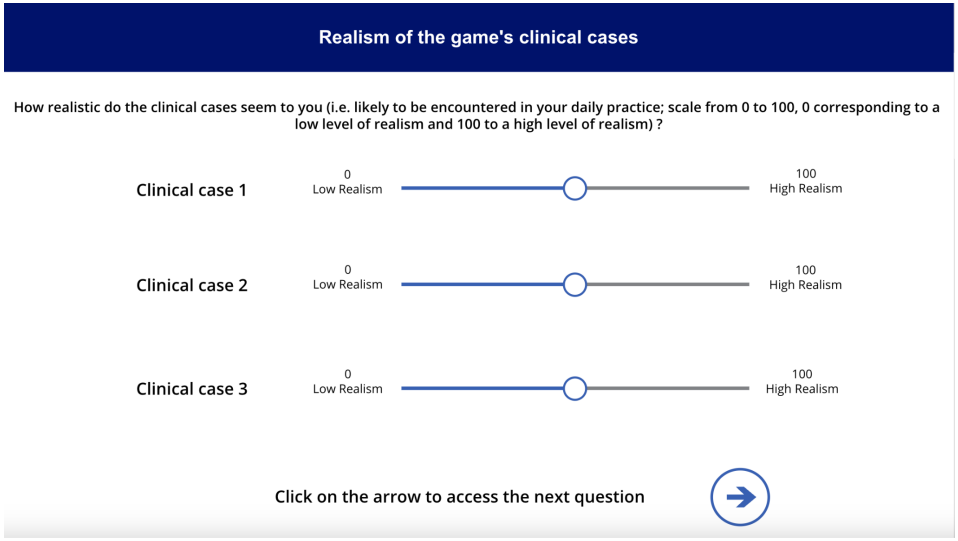


Suspected diagnosis and diagnosis confidence level screen

Compulsory

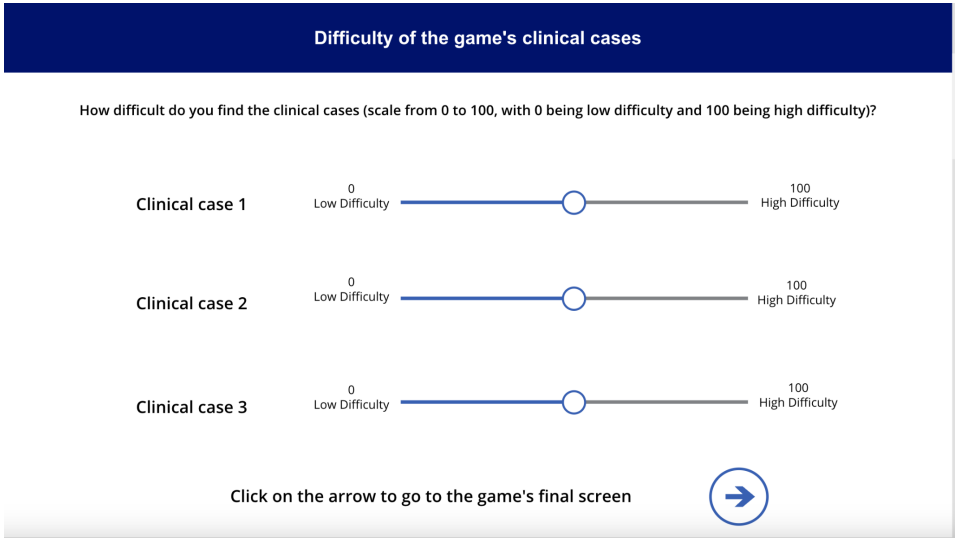


Case realism estimation screen

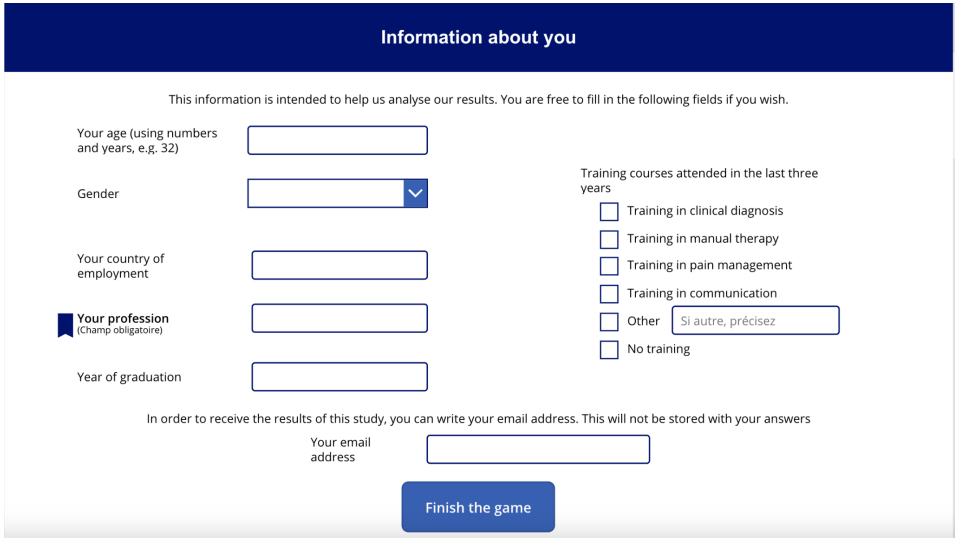


Compulsory

Case difficulty estimation screen



Compulsory

Sociodemographic data screen		Compulsory

[a]Compulsory means the screen was displayed to all, Optional means the screen was displayed only if selected.

Instruments

We measured participants' level of IU when caring for their patients using the IUS-12-H (Brun *et al*., 2023). This scale based on the IUS-12, a generalist IU scale initially validated by Carleton *et al*. (2007), has been validated on a sample of HCPs and has shown good psychometric qualities. The IUS-12 is scored on a Likert scale ranging from "Not at all characteristic of me" (1), "A little characteristic of me" (2), "Somewhat characteristic of me" (3), "Very characteristic of me" (4), to "Entirely characteristic of me" (5).

At the end of each of the three cases, we measured the following variables for each case. We assessed participants' choice of treatment strategy using the following instruction: "In your opinion, the most appropriate treatment strategy for this patient would be: (one answer choice only)". Participants were asked to choose one of the following four treatment strategies as the most adequate one: "Starting physical therapy", "Starting physical therapy while raising my level of vigilance", "Referral to a physician without starting physical therapy", "Emergency medical referral"[8]. We chose this variable as our main dependent variable as most studies on virtual patients focus on measuring diagnostic accuracy of participants (*i.e.*, correctness of the final diagnosis; Fink *et al*., 2023). While this measure has the advantage of being easy to measure numerically and to score objectively, authors such as Daniel and colleagues (2019, as cited in Fink *et al*., 2023) stressed the importance of also measuring other elements that contribute to the assessment of diagnostic success, notably the use of additional diagnostic tests, treatment decisions, prognosis, and justifications for all of these aspects. In particular, they felt that the addition of these variables would reduce the overtreatment and undertreatment of patients and provide a better understanding of

[8] These categories were adapted from those used by Jette and colleagues (2006): "provide intervention without referral", "provide intervention and refer", or "refer before intervention" and those used by Keller and collaborators (2022): "provide a physiotherapy intervention without referral to a physician", "provide a physiotherapy intervention and refer to a physician", or "refer to a physician before a physiotherapy intervention".The adaptations we made make it possible to distinguish between initiating treatment and initiating treatment while remaining vigilant.

misdiagnosis, especially by students. For these reasons, we decided to focus on an important less considered variable, treatment strategy (*i.e.*, thus putting us within the scope of therapeutic reasoning; Duong *et al.*, 2023) and included measures addressing the recommendations of Daniel and colleagues (2019). The scoring grid for each strategy for each case is shown in Table 2.

Table 2

Scoring grid of strategies for each case

	Easy case	Medium case	Difficult case
Beginning of physiotherapy treatment	2	0	0
Beginning physiotherapy treatment while maintaining a high level of vigilance	1	2	1
Referral to a physician without starting physiotherapy treatment	0	1	2
Emergency medical referral	-1	-1	-1

Note: Scoring was based on the notion of distance from a correct answer. A score of 2 was assigned to the strategy considered to be the most appropriate for the case presented. A score of 1 was assigned to the strategy considered to be good but not optimal. A score of 0 was assigned to the strategy considered to be neutral in that it has no harmful effect but no real positive effect. A score of -1 was assigned to the strategy considered to be contraindicated.

On the treatment strategy selection screen, participants could click on a button to access additional test results (“To help me make a decision despite the consumption of additional resources, I want to have access to additional test results”). If clicked on, a new screen with the following instruction prompted them to select the results they wished to see: “Please select the additional test(s) that you feel are relevant to your decision and then click on the Get Report(s) button: (several choices possible)”. A button allowed participants to return to the treatment strategy screen if they no longer wished to view these results, but we recorded the fact that they had accessed this screen. We measured additional test results selected from the following list: “X-Ray”, “CT scan”, “MRI”, “Scintigraphy” and “Blood tests”. On the next screen, participants indicated their suspected diagnosis associated with the treatment strategy they had chosen. The question stated: “The strategy you have chosen is related to the fact that you suspect a diagnosis of : (tick the one you think is most likely)”. Participants chose the diagnosis they thought most likely from the following list: “Non-specific low back pain”, “Fracture of the L5 vertebral body”, “Chronic/persistent low back pain”, “Pain related to the sacroiliac joint”, “Scoliosis”, “Ankylosing spondylitis”, “Muscle injury in the lower back or pelvis”, “Abdominal aortic aneurysm”, “Radiculopathy”, “Metastatic bone disease in the thoracolumbar spine”, “Other”. If the “Other” option was selected, they could write their suggestion in a dedicated writing space. These diagnoses were based on scientific literature (Finucane *et al*., 2017, 2020; Meidinger *et al*., 2023) and the International Classification of Diseases 11th Revision (ICD-11, World Health Organization, n.d.). We also measured the confidence level associated with the participants' given diagnosis with the following question: “What is your level of confidence in this diagnosis (scale from 0 to 100, with 0 corresponding to the minimum confidence level and 100 to the maximum confidence level)”. Participants indicated their level of confidence using an analogous scale

below the instruction, ranging from 0 to 100 with the “Not at all confident” label under 0 and the “Fully confident” label under 100.

After the three clinical cases, participants were asked two questions for each case to assess the level of realism and difficulty associated with each case. To assess the cases' level of realism, participants were asked to rate on an analogous scale from 0 to 100 how realistic the clinical case presented seemed to them (“How realistic do the clinical cases seem to you (*i.e*. likely to be encountered in your daily practice; scale from 0 to 100, 0 corresponding to a low level of realism and 100 to a high level of realism)?”), with the label “Low Realism” below 0 and “High Realism” below 100. To assess the difficulty level of the cases, participants were asked to rate on an analogous scale from 0 to 100 how hard to handle the case was (“How difficult do you find the clinical cases (scale from 0 to 100, with 0 being low difficulty and 100 being high difficulty)?”), with the label “Low Difficulty” under 0 and “High Difficulty” under 100. We decided to assess these measures after the completion of the three cases so that participants could use the other two cases as a reference point to help them answer these questions and to avoid their perception of the realism and difficulty of one case would influence their answers for the next case.

On the last screen, participants were invited to report their socio-demographic data with the following instruction: “This information will help us in the analysis of our results. You are free to fill in the following fields if you wish”. In writing spaces, participants reported their age (“Your age - using numbers and years, *e.g*., 32”), profession (*i.e*., to check that participants were physiotherapists; “Your profession”), country of practice (*i.e.*, to explore whether there were differences in responses across countries of practice and schools of thought; “Your country of employment”) and graduation year (*i.e*., to approximate the participants' experience level; “Year of graduation”). For gender, participants could choose the gender they identified with using a drop-down menu (“Your Gender”, response options:

"Male", "Female", "Do not wish to answer" and "Other" with a writing space if it was selected). Finally, participants were invited to provide their email addresses in a written space ("In order to receive the results of this study, you can write your email address. This will not be stored with your answers").

Procedure

The first screen of the game presented the game structure (*i.e.*, the number of cases and the case sections), the rules of the game (*i.e.*, perform the consultation and choose an appropriate treatment strategy, and if necessary use additional test results) and the content of the following screen, a scale. Once participants had clicked on the "Play" button at the bottom of the page, they were redirected to the first study screen, which featured an uncertainty intolerance scale. The next screens presented the first clinical case introduced with a short presentation and questions/answers[9], the treatment strategy selection screen[82] (and if selected, the additional test results selection screen and then the additional test results screen. After these two additional screens, participants were redirected to the treatment strategy selection screen) and the suspected diagnosis and confidence in diagnosis selection screen[82]. The same screens were presented for the following two clinical cases. On the diagnosis selection and confidence in diagnosis screen, participants could click on a button that redirected them to the next case or to the screens after the clinical cases for the last case. Then, the participants estimated the level of realism of each case on a new screen, then the level of difficulty of each case on the next screen. Finally, participants filled in their socio-demographic data on a new screen and once they had clicked on the "Finish the game" button, they were redirected to the final thank you screen. Screenshots of each screen are available in Table 1.

[9] On these screens, participants were informed that if they clicked on the button redirecting to the next screen, they would not be able to go back and return to the current screen. This was meant to ensure that participants understood that they needed to think carefully before moving on to the next screen (*e.g.*, they needed to be sure that they did not need to consult more information given by the patient before submitting their treatment strategy).

Statistical analysis

We began by conducting our main statistical analyses. We conducted statistical analyses (i) on the therapeutic strategy chosen by HCPs to quantify the strategies chosen in relation to their suitability for the case considered, and to identify factors of influence on the therapeutic strategy chosen (*i.e.*, with frequency table and linear regression model). The aim was to verify that the majority of participants chose the right therapeutic strategy and to check whether participants' level of IU and resource consumption had an influence on the strategy they chose. We then performed (ii) regression analyses to see whether physiotherapists' level of IU influenced their resource consumption. We then conducted correlational analyses to determine (iii) whether physiotherapists' confidence was related to the therapeutic strategy they chose and (iv) whether physiotherapists' level of IU was related to their confidence in their decisions.

For our secondary analyses, (v) we performed frequency analyses on participants' confidence in their choices, level of realism attributed to the cases and level of difficulty attributed to each case. Here we wanted to quantify participants' level of confidence, check that cases were considered realistic (*i.e.*, cases would be considered realistic above a level of 70 on the scale provided for this purpose) and that the perceived level of difficulty between cases was increasing.

Finally, we conducted exploratory analyses. We performed (vi) frequency analyses on the diagnoses selected by the participants for each case. With this analysis, we wanted to confirm that the majority of participants had found the right diagnosis, and to ascertain that the number of diagnoses selected by participants increases as the difficulty of the case increases (as suggested in the “Creation of the clinical case content” section).

Results

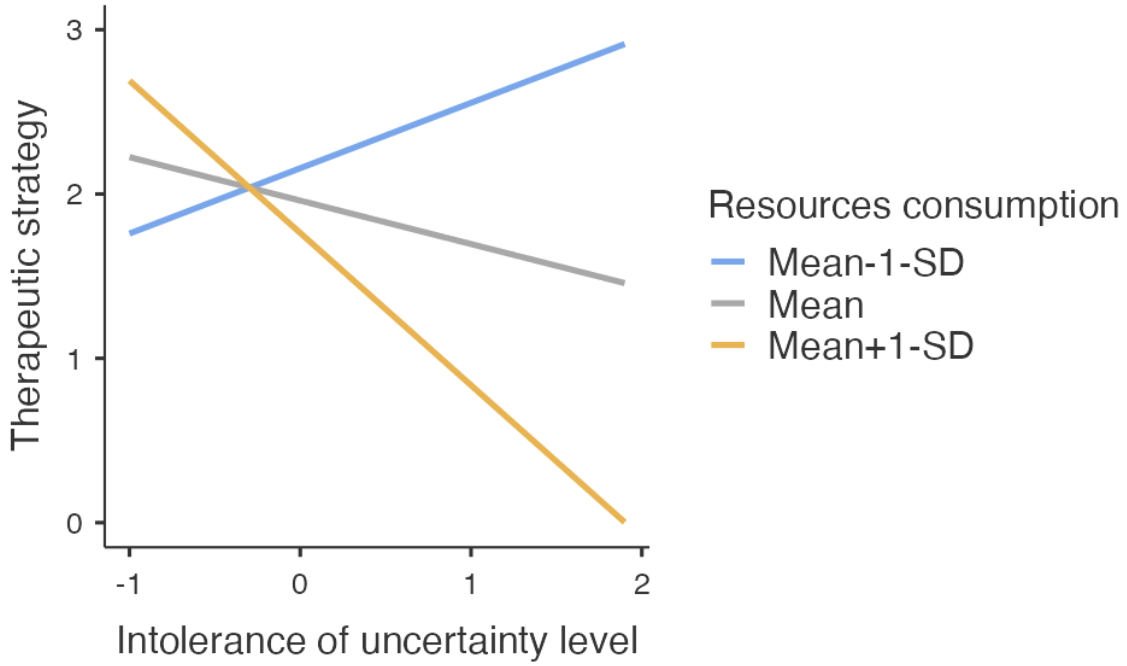
Assessment of the therapeutic strategies chosen and the factors influencing this choice

Analysis for the easy case revealed that 87.4% of participants selected the most appropriate strategy for the case (*i.e.*, beginning of physiotherapy treatment), 11.8% selected the strategy considered to be good but not optimal (*i.e.*, beginning physiotherapy treatment while maintaining a high level of vigilance) and 0.8% selected the neutral strategy (*i.e.*, referral to a physician without starting physiotherapy treatment). None of the participants chose the strategy considered to be contraindicated (*i.e.*, emergency medical referral). Most participants did not use any resources (3.9% requested test results while 96.1% did not). We conducted a regression analysis which showed a significant negative effect of HCPs' level of IU on the therapeutic strategy they chose, β = -.26, $t(121)$ = -4.08, $p < .001$. The model also showed a significant negative effect of resource consumption on the chosen therapeutic strategy, β = -1.20, $t(121)$ = -5.91, $p < .001$. Participants' IU levels significantly interacted with their resource consumption to influence the chosen therapeutic strategy, β = -4.03, $t(121)$ = -3.92, $p < .001$. Specifically, resource consumption had a positive influence on the quality of the therapeutic strategy chosen for uncertainty-tolerant individuals (Mean-1-SD), β = 1.02, $t(121)$ = 1.96, p = .053. But resource consumption had a negative influence on the quality of the chosen therapeutic strategy for individuals with a moderate tolerance of uncertainty (Mean), β = -1.20, $t(121)$ = -5.91, $p < .001$, and individuals intolerant of uncertainty (Mean+1-SD), β = -3.42, $t(121)$ = -5.09, $p < .001$ (cf. Figure 1). IU had a positive influence on therapeutic strategy when participants consumed few resources (Mean-1-SD), β = 0.40, $t(121)$ = 2.59, p = .011. However, it had a negative influence on the therapeutic strategy when participants consumed an average amount of resources (Mean), β = -0.26, $t(121)$ = -4.08, $p < .001$, and when they consumed a lot of resources (Mean+1-SD), β = -0.93,

$t(121) = -4.53$, $p < .001$ (cf. Figure 1). The model also included other variables: age, experience (*i.e.*, relative to graduation date) and gender. None of these variables had a significant effect on the strategy chosen by participants ($p_{age} = .904$, $p_{experience} = .725$, $p_{gender} = .798$).

Figure 1

Quality of therapeutic strategy chosen by level of intolerance to uncertainty and resource consumption



Note: Uncertainty-tolerant physiotherapists (Mean-1-SD, on the left of the graph) suggest therapeutic strategies that are more adapted when they consume a lot of resources (orange line) than when they consume an average amount (grey line), and even more so than when they consume little (blue line). For physiotherapists with an average level of IU (Mean, on the centre of the graph), the therapeutic strategy is all the more adapted when they consume few resources, or at most an average level of resources, than when they consume a lot. This effect is all the more pronounced in uncertainty-intolerant physiotherapists (on the right of the graph), for whom low resource consumption is associated with much better therapeutic strategies than uncertainty-intolerant physiotherapists who consume an average level of resources, themselves choosing more appropriate strategies than uncertainty-intolerant physiotherapists who consume a lot of resources.

Analysis showed that for the medium case, 46.5% of participants selected the most appropriate strategy for the case (*i.e.*, beginning physiotherapy treatment while maintaining a high level of vigilance), 28.3% selected the neutral strategy (*i.e.*, beginning of physiotherapy treatment) and 25.2% selected the good but not optimal strategy (*i.e.*, referral to a physician without starting physiotherapy treatment). The strategy considered to be contraindicated (*i.e.*, emergency medical referral) was not selected by any participant. Just under half of the participants used resources (42.5% asked for test results whereas 57.5% did not). We conducted a regression analysis which did not show a significant effect of HCPs' level of IU on the therapeutic strategy they chose, β = -.03, $t(121)$ = -.19, p = .852. The model also included age, experience, gender and resource consumption but none of these variables had a significant effect on the strategy chosen (p_{age} = .744, $p_{experience}$ = .698, p_{gender} = .716, $p_{ressources\ consumption}$ = .034).

Lastly, regarding the difficult case, 41.7% of participants selected the good but not optimal strategy for the case (*i.e.*, beginning physiotherapy treatment while maintaining a high level of vigilance), 29.1% selected the most appropriate strategy (*i.e.*, referral to a physician without starting physiotherapy treatment), 18.9% selected the contraindicated strategy (*i.e.*, emergency medical referral) and 10.2% selected the neutral strategy (*i.e.*, beginning of physiotherapy treatment). Almost half of the participants used resources (44.1% requested test results compared with 55.9% who did not). The regression analysis we performed did not show a significant effect of HCPs' level of IU on the therapeutic strategy they chose, β = -.01, $t(121)$ = -.71, p = .479. Among the other variables included in the model (*i.e.*, age, experience, gender and resources consumption), none had a significant effect on the chosen strategy (p_{age} = .653, $p_{experience}$ = .595, p_{gender} = .509, $p_{ressources\ consumption}$ = .014).

Assessment of the resources consumption and the factors influencing this choice

For the easy case, 3.9% of participants used resources (*i.e.*, exam results). We tested a linear regression model, which showed no significant effect of participants' level of IU on their tendency to consume resources, β = -0.02, $t(122)$ = -0.56, p = .575. The model also included other variables (*i.e.*, age, gender, experience) that had no significant effect on resource consumption (p_{age} = .478, $p_{experience}$ = .537, p_{gender} = .660).

As for the medium case, 42.5% of participants consumed resources. The linear regression model we tested showed no significant effect of participants' level of IU on their tendency to consume resources, β = 0.14, $t(122)$ = 1.55, p = .125. The other variables included in the model (*i.e.*, age, gender, experience) also had no significant effect on resource consumption (p_{age} = .704, $p_{experience}$ = .637, p_{gender} = .391).

Finally, for the difficult case, 44.1% of participants consumed resources. The linear regression model tested showed no significant effect of participants' level of IU on their tendency to consume resources, β = 0.01, $t(122)$ = 0.06, p = .952, and the other variables included in the model (*i.e.*, age, gender, experience) did not have a significant effect on resource consumption either (p_{age} = .058, $p_{experience}$ = .035, p_{gender} = .565).

Assessments of the relation between confidence and therapeutic strategy

Physiotherapists' confidence in their decisions was positively significantly correlated with the quality of the therapeutic strategy they chose for the easy case (r = .41, $p < .001$). However, confidence was significantly negatively correlated with the quality of the therapeutic strategy proposed for the medium (r = -.31, $p < .001$) and difficult cases (r = -.39, $p < .001$).

Assessment of the relation between IU and confidence

Physiotherapists' level of IU was not significantly correlated with their confidence in their decision for the easy case (r = -.08, p = .404), the medium case (r = -.15, p = .089) and the difficult case (r = -.09, p = .316).

Assessment of the confidence level in one's diagnosis, the level of realism and the level of difficulty for each case

On average, participants' confidence in their decision was rated at 75.7 for the easy case, 61.9 for the medium case and 62.6 for the difficult case. As for the realism attributed to cases, the average level attributed to the easy case was 86.5, 82.9 for the medium case and 80.0 for the difficult case. Finally, the average difficulty level estimated for the easy case was 34.7, 48.8 for the medium case and 55.0 for the difficult case.

Diagnostic selected for each cases

Diagnoses selected by participants for each case are shown in Table 3.

Table 3

Percentage of diagnoses selected for each case

Diagnostic	Easy case	Medium case	Difficult case
Non specific low back pain	94.5%	6.3%	18.1%
Fracture of the L5 vertebral body	0%	0%	3.1%
Chronic/persistent lumbago	3.1%	16.5%	9.4%
Pain related to the sacroiliac joint	0%	21.3%	8.7%
Scoliosis	0%	7.1%	0%
Ankylosing spondylitis	0%	44.1%	3.1%
Muscle injury in the lower back or pelvis	0%	0.8%	3.1%
Abdominal aortic aneurysm	1.6%	0%	0.8%
Radiculopathy	0%	0%	6.3%
Metastatic bone disease in the thoracolumbar spine	0%	0%	41.7%
Other	0.8%	3.9%	5.5%

Discussion

Summary of results

Diagnostic errors are considered to be among the most common and costly and errors in the health field (Tehrani *et al*., 2013). Research seeking to further understand the drivers of diagnostic errors remains lacking due to existing divisions between experts working in the field of clinical reasoning and those working in the field of diagnostic error (Durning *et al*., 2020). Although authors have proposed recent guidelines for improving cognitive processes in clinical reasoning (Scott *et al*., 2021) and for avoiding diagnostic errors (Chakravarty, 2022), these recommendations are far from being widely applied at present, which calls for more studies to be carried out to foster qualitative diagnostic process and therapeutic strategy for patients. Specifically, we need to better understand the influence of factors such as IU in order to limit its potential negative effects on the formulation of the therapeutic strategy chosen.

We analysed the responses of 127 physiotherapists to a serious game in which they had to evaluate three clinical cases of increasing difficulty and suggest a therapeutic strategy. We also assessed their IU level and their resource consumption (*i.e*., test results) to see whether these two variables had an influence on their clinical reasoning process and the therapeutic strategy they recommended to patients. We had four hypotheses : (H1) A high level of IU is associated with the selection of less adapted therapeutic strategies, (H2) A high level of IU is associated with greater consumption of resources (*i.e.*, more requests for test results), (H3) Level of confidence in one's therapeutic strategy will be negatively correlated with the selection of adapted therapeutic strategies (*i.e*., calibration issue) and (H4) IU will be negatively associated with confidence in one's diagnoses.

Our main results showed that for cases of easy or medium difficulty, the majority of physiotherapists offered the most appropriate therapeutic strategy (87.4% for the easy case and 46.5% for the medium case). On the other hand, for difficult cases, there was less consensus, more therapeutic strategies were considered by the sample, and the most chosen strategy was good but not optimal (chosen by 41.7%). While the most appropriate strategy was the second most chosen (29.1%), the contraindicated strategy was still chosen by 18.9% of participants. Participants barely used any resources for the easy case (3.9%), but did so in similar proportions for the medium and difficult cases (42.5% and 44.1% respectively). We should bear in mind that these raw numbers cannot be interpreted directly in terms of treatment offer or patient pathway, since we do not have data for GPs, which is the current standard pathway. However, we will compare these results with those of other studies carried out on physiotherapists later on.

Our regression analyses revealed that IU and resource consumption were significantly associated with choosing a less adapted therapeutic strategy for the easy case but not for the medium and difficult cases, which confirms hypothesis H1 for easy cases but does not confirm it for cases of moderate or high difficulty. Interestingly, our results showed that the IU level and resource consumption interacted for the easy case such that for individuals with a high level of IU, resource consumption seemed to have a negative effect and lead to the choice of less adapted therapeutic strategies. This result is very important, as it questions the use of exams among individuals who are highly intolerant of uncertainty, as they have a negative effect on the treatment they offer while being costly for society. We also note that neither age, gender nor experience revealed a significant effect on the chosen strategy and resource consumption. These results are similar to those of Aron and colleagues (2023) who found no significant effect of age, gender and experience on referral, which in their study was the correct strategy. Regarding H2, our regression analyses showed no significant effect of IU

on resource consumption for the three cases, which means that the hypothesis cannot be confirmed. We believe this result may reflect the current French context, in which physiotherapists receive patients on the basis of a GP's prescription but are not authorised to prescribe additional tests themselves, only to consult those already prescribed by GPs or to suggest a GP prescription. To determine the effect of IU on resource consumption, a study similar to ours should be carried out in countries where physiotherapists can prescribe additional tests (*e.g.*, Canada, Australia, where advanced practice[10] is more widespread), or repeated in France when physiotherapists are required to prescribe additional tests in the future[11]. For H3, we found a negative correlation between therapeutic strategy and physiotherapist confidence for medium and difficult cases (at r = -.31 and r = -.39). Physiotherapists were confident in their choice, even though they did not necessarily choose the most appropriate therapeutic strategy. In our opinion, this result indicates a calibration issue among physiotherapists, which is consistent with the literature on HCPs overconfidence and calibration issues we previously mentioned, especially for difficult cases. The interesting thing about our results is that here the calibration problem is not about the diagnoses, but about the quality of the strategies offered. We did not observe this effect for the easy case since there was a positive correlation between the therapeutic strategy chosen and physiotherapist confidence, but this is logical since the case was easy and the majority of participants chose the most appropriate therapeutic strategy, which led to a good calibration between the chosen strategy and its quality. Finally, hypothesis H4 was not supported as we did not observe any correlation between physiotherapists' IU levels and their confidence in

[10] Advanced practice refers to a form of practice in which HCPs are recognised as competent to perform acts not initially assigned to them, as described by Kechichian and Pinsault (2023): "Physiotherapists are thus able to perform acts that were initially reserved for physicians, such as making a medical diagnosis, prescribing medication or imaging (...) Advanced practice falls within the scope of functions that are recognised as part of the profession's scope of practice, but which may traditionally be performed by other professions. Thus, physiotherapists may be called upon to perform acts initially reserved for medical professions, such as diagnosing, prescribing drugs, referring to other health professionals or prescribing imaging".

[11] Physiotherapists could be increasingly involved in cooperation protocols with other professions, and their tasks could be expanded in this context (Kechichian & Pinsault, 2023).

their diagnosis, even though one might have expected a higher level of IU to be associated with more confidence in their decisions[12]. We believe that this result is due to a likely sampling bias. Our study required a significant cognitive and time investment, which may have led only the most motivated and confident people to participate[13], possibly explaining our high experimental mortality and why we did not observe an effect between IU and confidence. Future studies could address this by conducting generalised studies with larger samples that do not only include motivated and voluntary participants.

Regarding our secondary analyses, participants clearly perceived an increasing difficulty between cases 1 (34.7), 2 (48.8) and 3 (55.0) and all realism levels attributed to each case was above 80 (respectively 86.5, 82.9 and 80.0). Combined, these two elements lead us to believe the cases designed for this study presented reasonable validity. Regarding confidence, the decrease in confidence level as the difficulty of our cases increases is consistent with the results of Staal and colleagues (2023) who reported a similar decrease with increasing difficulty. Interestingly, the mean confidence level was higher for the easy case (75.7) than for the medium case (61.9) and the difficult case (62.6), both of which had similar mean confidence levels. Given that participants considered the medium case to be simpler than the difficult case, we do not attribute this effect to a perception of similar difficulty between the medium and difficult cases. Instead, we hypothesize these results may depend on case content. Perhaps the medium case seemed simpler but featured a clinical picture that participants rarely encountered. We will discuss this point in more detail later. Moreover, our confidence levels seem compatible with confidence levels reported by other researchers such as Haber and colleagues (*i.e.*, 5% of their sample was “Very confident”, 47% “Confident”, 42% “Somewhat confident”, 5% “Not very confident” and 0% “Not at all

[12]This could have reflected a tendency to ignore conflicting information causing uncertainty by quickly choosing a strategy (*i.e.* in the way premature closure occurs; Durning *et al.*, 2020; Norman & Eva, 2010), which would have increased physiotherapists' confidence.

[13] Possibly also physiotherapists who continually undertake training and read scientific literature.

confident"; 2023). Finally, our exploratory analyses on diagnoses showed that a majority of participants found the right diagnosis in each case and support the suggestions we made in our methodology, namely that as the difficulty of cases increases, so does the number of diagnoses considered. Interestingly, we observed the same effect as for confidence, that is, participants overwhelmingly found the right diagnosis for the easy case (94.5% of them), and less so for the medium case (44.1%) and the difficult case (41.7%), which are themselves of similar levels. We suggest the same hypothesis for confidence to explain these results.

Do most physiotherapists recommend good therapeutic strategies?

To interpret the therapeutic strategy results selected by physiotherapists, we need to compare them with reference points such as the articles by Jette and colleagues (2006) and Keller and colleagues (2022). Jette and colleagues (2006, p. 1619) found relatively high rates of correct therapeutic strategy: "The average percentages of correct decisions were 87%, 88%, and 79% for musculoskeletal, noncritical medical, and critical medical conditions, respectively". Keller and colleagues (2022, p. 5), although finding lower percentages than Jette and colleagues, reported higher rates than us: "71.2% correct decisions on further management". If we consider only the choice rates of the most appropriate strategy, the physiotherapists in our sample underperformed, with rates of 87.4% for the easy case (making this the only case with a rate comparable to Jette and colleagues'), 46.5% for the medium case and 29.1% for the difficult case. However, if we consider the most appropriate strategies in conjunction with the good but not optimal strategies, the combined rates reach 99.2% for the easy case, 71.7% for the medium case and 70.6%. These rates exceed or are comparable to those of Jette and colleagues (2006) and Keller and colleagues (2022), which testify to the ability of the sample surveyed to propose strategies that are good if not totally adapted to patients' cases. However, while following in the footsteps of Jette and colleagues

(2006) and Keller and colleagues (2022) and confirming their results, our study contributes to the literature by offering cases with a higher level of realism and complexity (*e.g.*, adding confusing information) than the studies by Jette and colleagues (2006) and Keller and colleagues (2022). Indeed, Jette and colleagues' cases were very concise and simple vignettes presenting all the information that could induce the activation of a pattern that physiotherapists could recognize, and Keller and colleagues (2022) used the same vignettes. In comparison, our cases were much more detailed, requiring physiotherapists to seek out the information themselves to confirm or refute their hypotheses, which was more realistic and certainly more conducive to complex reflection. Our easy case, which most closely resembled those of Jette and colleagues (2006) and Keller et colleagues (2022) as it presented a typical clinical situation, had an almost identical rate for the most appropriate strategy, which reinforces our belief that the cases created by Jette and colleagues (2006) are simple and conducive to pattern recognition, a point we will address in the next section.

However, while these results are encouraging, they must be compared with yet more reference points to confirm our conclusions, as Keller and collaborators (2022, p. 13) also noted: “Categorizing PTs' abilities would require comparison with other studies”. New reference points will be available when Lackenbauer and collaborators (2023) publish the results of the study for which they have published a research protocol. GPs could also be an interesting reference point, as they are the other professionals that patients see for conditions that could be treated by physiotherapists, and several studies have shown that GPs and physiotherapists propose diagnoses with the same level of accuracy (Keller *et al*., 2022). A future study could compare the therapeutic strategies proposed by GPs and physiotherapists. However, although physiotherapists have relatively good rates if we combine the rates for the most appropriate strategy and the good but not optimal strategy, as difficulty increases, the most appropriate strategy is chosen less often. In addition, we note that medium and difficult

cases have similar overall results (except for the number of diagnoses considered), which leads us to formulate several hypotheses that could explain why medium and difficult cases are treated differently from easy ones.

Hypotheses on the differences observed between cases

We formulate three explanatory hypotheses that will need to be investigated in the future to clarify the results we uncovered. The first hypothesis is based on Kahneman's system 1/system 2 theory, which states that reasoning is based on two systems. Whereas system 1, a fast and heuristic-based system, is suited to situations where patterns can be recognized, system 2, a slower and analytic system, is more suited to situations where it is necessary to step back and think deliberately (Benishek *et al*., 2015; Cheng & Senathirajah, 2023; Croskerry, 2009; Croskerry & Norman, 2008; Croskerry *et al*., 2013; Peterson *et al*., 2022). We believe the easy case probably triggered system 1, as it presented a simple case that physiotherapists encounter regularly and which favoured recognition of a familiar pattern (Nendaz & Perrier, 2012). The medium and difficult cases, while of different difficulty levels, both featured less familiar cases[14], which likely triggered system 2. We believe the similar results between the medium and difficult cases are due to both triggering the analytical system (Nendaz & Perrier, 2012). Notably, we believe the IU and resource consumption effects we observed only for the easy case depend on which system has been activated. As discussed in the literature, System 1 is conducive to the activation of heuristics that can favour the influence of external factors (Croskerry, 2009; Croskerry & Norman, 2008; Croskerry *et al*., 2013), which could explain why we found an effect of IU and resource

[14] Ankylosing spondylitis (our medium case) may take an average of 6 to 7 years to be diagnosed (Redeker *et al*., 2019). Metastatic bone disease in the thoracolumbar spine (our difficult case) is a condition with low prevalence in primary care (*i.e*., <1%; Finucane *et al*., 2020) and a lower expected prevalence for physiotherapists as they are not usually the first-line primary care practitioners in France. This condition is not inherently difficult to diagnose, provided you can access test results (Henschke *et al*., 2013; Sutcliffe *et al*., 2013). However, physiotherapists do not prescribe tests in France and our results seem to indicate that they do not reflexively look at them, which complicates the diagnosis of this condition.

consumption on the easy case. System 2, on the other hand, mobilizes cognitive resources and conscious reflection that may have left little room for the effects of IU and resource consumption. Consequently, further studies are required to determine how these two factors influence the therapeutic strategies proposed to patients with clinical pictures that are less familiar than the typical cases encountered by physiotherapists.

Another hypothesis we put forward to explain the differences between cases is partly connected to this notion of familiarity, that is, taking into account the scope of practice[15] of physiotherapists. As noted by Keller and colleagues (2022), specialization and professional contexts conducive to confrontation with similar cases can foster their recognition and the suggestion of more appropriate therapeutic strategies. Our easy case perfectly fits this situation, which may explain the very high rates we observed. The effects we attribute to a lack of case familiarity may add to the effects of physiotherapist specialization. While our sample consisted exclusively of physiotherapists, we did not control for their field of expertise (*e.g.* paediatrics, maxillofacial). Yet we only presented cases of low back pain, a field in which the expertise of physiotherapists and consequently their conclusions may vary according to their expertise. The influence of this variable should therefore be controlled in future studies as it may have influenced physiotherapists' answers in the most difficult cases (and potentially reinforced the use of system 2).

Finally, our third explanatory hypothesis is based on physiotherapists' perception of uncertainty. Although it is conceivable that the increasing difficulty of clinical cases may be associated with a growing increase in perceived case uncertainty, we did not measure the effect of this variable on the therapeutic strategy chosen. However, we hypothesize that we observed the effect of IU and resource consumption in the easy case because the

[15] "Scope of practice" is a generic term used to describe the skills and procedures that physiotherapists are authorised to perform. The French Public Health Code employs the term "spécificités d'exercice", which has no equivalent in English, but whose definition may be close to that of specialisation, although the two terms do not fully overlap.

physiotherapists were engaged in the case whereas the medium and difficult cases, as they were unfamiliar, triggered a high perception of uncertainty that resulted in physiotherapists disengagement from the cases. While past studies have shown an influence of IU and not of perceived uncertainty on healthcare decisions (Brun *et al*., 2022a), it is possible that the reasoning of physiotherapists, and HCPs more generally, is a separate issue, where perceived uncertainty causes cognitive disengagement, reflected in our case by lower rates of selection of the most appropriate therapeutic strategy and a lack of influence of IU and resource consumption. Future studies should therefore clarify the influence of perceived uncertainty on the quality of proposed strategies (which is an approximation of their ultimate engagement).

Limitations

Certain limitations in our study must be recognized in order to promote future advances in research. A classic but noteworthy limitation of our study is the generalizability of our findings. Indeed, authors such as Vreugdenhil and collaborators (2023) remind us that different HCPs working in different professions will view and reason about clinical cases differently. Therefore, we cannot ensure that our findings are applicable to other types of medical professions and encourage other researchers to use our method to assess the impact of IU on reasoning, and particularly on the therapeutic strategy chosen by other types of HCPs. Another limitation of our study lies in the way we chose to address one of our study objectives, namely our wish to set into perspective the observed results with the resource consumption of the participants, and particularly to assess whether people with a high IU tended to consume more resources (*e.g.*, requesting more supplementary tests). In the context of this sub-objective, we mentioned the consumption of resources in the instruction screen and insisted on the fact that asking for supplementary tests was associated with a consumption of resources so that participants would not ask for them inappropriately and to

ensure that this consumption would only occur when physiotherapists really wanted to access this information (*e.g*. an important need for certainty and not just a light desire to "confirm their hypotheses"). Although the instruction was written so as to make participants aware of the consumption of resources while not inviting them to avoid additional examinations ("Before proposing a therapeutic strategy, you will have the possibility to access additional examination results (imaging, blood tests) if you consider it necessary. However, please only use them if you really think they are necessary, as they are associated with the consumption of certain resources (*e.g*. patient's time, costs of the examinations")) and was reviewed by physiotherapists, we cannot be sure how these instructions influenced the participants' use of additional examinations. Future studies investigating resource consumption in physiotherapy will have to carefully manipulate such variables in a subtle way and try to assess the impact of such instructions on physiotherapists' consultation behaviour and subsequent treatment strategies.

Regarding the methodology we used, while we put special attention into implementing the clinical cases in a relevant tool, we acknowledge several limitations to our study directly associated with this methodological format. Firstly, one limitation relates to the instructions given to the participants. We told them that they would not be able to go back once they had chosen to suggest a treatment strategy. This means that we have potentially artificially increased the attention paid to the information given by the patients compared to a real situation where physiotherapists probably feel less need to pay attention to the information given by the patients as they can ask them again. This limitation calls for studies comparing similar consultation conditions with varying instructions regarding attention given to information by patients (*e.g*. "You may ask patients as many times as necessary" *vs*. "You may not ask patients repeatedly") in order to assess the impact of this type of instruction and attention to information on the outcome of the consultation. We also had to make choices

regarding the format of the clinical case questions and answers that potentially influenced the results of the study. Specifically, we chose a format with an on-screen presentation where answers to previously asked questions remained on the screen, as opposed to a presentation of information where participants would have to click back on a previously asked question to reexamine an answer (see screen "Clinical case screen with "Question/Answer" format" in Table 1). We chose this format so that the information presented on the side of the screen would be a written record similar to what the physical therapists might have written themselves during or following a consultation. This format also seemed practical to limit completion time and allow participants to focus on their clinical reasoning rather than information gathering (albeit the latter is to some extent necessary and effective in the tool), although it potentially "made it easier" for participants as they did not have to take notes or memorize patient characteristics. The use of this format assumes that participants kept the requested information in front of them and did not have to take notes which may constitute a gap with a normal consultation where physical therapists have access to a limited amount of new information per patient response that they must memorize or take notes on. An alternative for future studies would be to propose a drop-down menu where participants click on a question and have the answer just underneath which keeps the spirit of a "conversational" style close to a real consultation without burdening the tool ergonomics. In addition, other transformations could make the consultation more realistic such as adding audio content instead of text or adding images to illustrate the consultation.

Lastly, regarding our measures, we decided to include an item about the perceived "realism" of the cases in order to assess the quality of the cases (*i.e.*, were they similar to cases that participants might actually encounter in their practice), which were based on clinical cases that had already been evaluated with a DELPHI method. We felt that this item was necessary but sufficient in that we did not want to overburden the participants with too

many questions about each case and that the validity of the cases was partly ensured by the fact that they were based on validated cases. However, we have not formally assessed the validity of the cases we created and we believe that future studies should include in their methodology the creation and validation of a grid created specifically to assess the validity of the clinical cases. Scientific articles explaining the required procedure to ensure validation provide clear steps to follow. Examples include Downing's (2003) seminal article which offers a review of the contemporary meaning of the term “validity” and describes the five typical sources of evidence for validity (*i.e.*, content, response process, internal structure, relationship to other variables and consequences) in relation to two example assessments in medical education. Another example is Schmutz and colleagues' (2014) article which describes five clear steps for designing checklists to evaluate clinical performance (cf. Table 4). One should bear in mind that this kind of validation methodology is time-consuming and involves many specialists as noted by Schmutz and collaborators (2014), which is why we consider this procedure as a separate side study to be carried out separately and a procedure that we were unfortunately not able to conduct before the study presented in this article.

Table 4

Schmutz and colleagues' (2014) steps for designing checklists to evaluate clinical performance

Step	Step content
Step 1: Development of a draft checklist	Creation of a preliminary checklist draft based on relevant literature and clinical experience.
Step 2: DELPHI review rounds	Sending of the draft checklist to five experts for review using an adapted DELPHI technique.
Step 3: Design of the final checklist and pilot testing	Design of three scoring categories for items after pilot testing.
Step 4: Final DELPHI review round	Submission of the checklist to an additional DELPHI review round to ensure the changes made after pilot testing were valid.
Step 5: Final DELPHI review round	Rating of all checklist items in terms of their importance on a scale from 1 (not important) to 5 (essential) by 10 specialists in the field presented in the assessed content to weight items needed for accurate performance assessment.

Possible improvements to the tool

We consider our study to be a first step towards a better understanding of clinical reasoning among physiotherapists using digital consultation simulation tools. As Daniel and collaborators (2019, p. 902) note: “The sheer number and diversity of clinical reasoning assessment methods create challenges for selecting assessments fit for the purpose”. We are aware that there is room for improvement in how we assessed clinical reasoning among our participants, and we have several avenues that we would like to explore and develop to refine our results. We must admit that we adopted certain positions when designing our tool which

did not leave enough room for consideration of the construction and evaluation grids available in the literature. A major adaptation of our tool would be to adapt it based on this literature. One way to refine our measures of interest, which are currently very focused on therapeutic outcome and whether it is appropriate to the situation, would be to assess more specific clinical reasoning components. Daniel and colleagues (2019) in their review of assessment methods for clinical reasoning suggest that it can be assessed through the following seven components: information gathering, hypothesis generation, problem representation, differential diagnosis, leading or working diagnosis, diagnostic justification, management and treatment. This classification is one of several that we could use to create a clinical reasoning assessment method that focuses more on the components of reasoning and that we could connect to the therapeutic outcomes considered by HCPs. This revision of the tool could be accompanied by a revision of the clinical cases. Given that cases 2 and 3 showed similar results for several variables, it would be worth testing the hypotheses we formulated on this subject and modifying the cases if the results observed were attributable to the writing and content of the cases.

Regarding our other measures, we are considering complementing some of them. Specifically, the IUS-12 is the gold standard for IU assessment but has specific limitations: a) it does not capture the way individuals represent uncertainty, b) it does not estimate or quantify how individuals react in specific uncertain situations, c) it does not assess the different types of uncertainty that individuals may face, and d) it does not capture the cognitive domains in which individuals may be particularly intolerant to uncertainty (Sandhu *et al*., 2023). Overall, while the IUS-12 is a good tool for measuring IU[16], it would benefit from being combined with other IU measurements that counteract its limitations. Future studies on physiotherapists' clinical reasoning and IU could incorporate a full range of

[16] Other scales for assessing intolerance to uncertainty specifically intended for physiotherapists are emerging, such as the “Tolerance to Uncertainty in Physiotherapy” (TUP) recently developed by Infante Guedes and colleagues (2023).

IU measurements to quantify the elements that we could not capture in this study. This is consistent with the alternative approach described by Sandhu and colleagues (2023) which is emerging in computational psychiatry. The idea is to use cognitive tasks in which experimenters ortogonalise the different forms of uncertainty being assessed and use computational modelling to characterise how individuals estimate and behave in the face of these uncertainties (Sandhu *et al*., 2023). This approach provides a more holistic and granular quantification of how people process uncertainty and could help describe the different routes that lead to its intolerance at the individual level (Sandhu *et al*., 2023). In line with these modifications suggested to gain a deeper understanding of the variables studied and to propose a more refined tool we are considering additions to the content of the tool that could enhance immersion. We would like to add more questions in the anamnesis section, as some reviewers told us that they asked more questions than those included in the tool during their daily consultations. Such additions would be especially beneficial as the history-taking process is an essential stage in the reasoning process that influences the chosen diagnosis and marks the point at which many HCPs make errors (Yanagita *et al*., 2023). We also wish to present the imagery results not using sentences (as we did in this study) but with real imagery that the players would have to visually interpret. As the interpretation of imagery largely depends on the specialist and the reason why the imagery was prescribed by a fellow medical specialist (O'Keefe, 2023), the introduction of real imagery could upset the reasoning of HCPs and increase the game's complexity, or even have an influence on the uncertainty perceived by HCPs.

Finally, we are considering the addition of a new module for the evaluation of diagnostic hypotheses and the resulting impact on the therapeutic strategy chosen. We chose to work within the paradigm of therapeutic reasoning in this study, which is more circumscribed than clinical reasoning, to fill in the gaps in the literature as described by

Duong and colleagues (2023, p. 16): “Outcomes of therapeutic reasoning were also rarely assessed. Future researchers should study the connection between process and outcomes of therapeutic reasoning”. We believe that, although we have partially addressed these limitations mentioned by Duong and collaborators (2023), we could incorporate new measures to further study the reasoning processes. We aim to include a real-time hypothesis measurement module. Indeed, Duong and colleagues (2023, p. 16) call for the evaluation of the maturation of therapeutic reasoning: “Finally, we recommend future studies to not only focus on describing processes, but also on assessing the effectiveness, predicting outcomes, and evaluating maturation of therapeutic reasoning systems.” We believe this could be accomplished by incorporating a module similar to the methodology described in the study by Pelaccia and colleagues (2014). They developed a methodology to measure at several points in time the diagnostic hypotheses considered by emergency physicians, which they also had to rank. This methodology enables the monitoring of the evolution of the considered diagnoses. We want to integrate a similar module for the considered therapeutic strategy to see, among other things, the impact of information given by the patients or of test results on the considered therapeutic strategy. This is consistent with the findings of Fink and colleagues (2023) who suggested on the basis of several studies that the quality of hypothesis generation is positively associated with diagnostic success, especially in the case of virtual patients, calling for more attention to the hypothesis generation process.

Practical implications for resource consumption and management of IU

As Infante Guedes and colleagues (2023) note, studies on the tolerance of clinical uncertainty among physiotherapists remain scarce. To our knowledge, studies examining its influence on physiotherapists' reasoning, and its interaction with variables of interest in terms of costs sustained by society, such as resource consumption, are all the more scarce, if not

non-existent. Physiotherapists' resource consumption will become an increasingly important issue, especially if, as we have observed, it may be associated with poorer appropriate strategy choices among individuals highly intolerant of uncertainty. Future studies should investigate the hypotheses we advanced, to determine whether these effects can be observed for medium and difficult cases, which require deliberate reflection, or whether this effect is limited to easy cases that rely on pattern recognition. New studies on uncertainty tolerance and resource consumption will also have to differentiate effects by test, since Begin and collaborators (2022) showed that HCPs with a high level of uncertainty tolerance order fewer diagnostic tests (*e.g.*, thyroid tests) and that Morgan and colleagues (2021) found that the adjustment of probabilities associated with various diagnoses differs according to the tests considered[17]. On a practical level, awareness of medical costs could be managed by displaying examination costs for example, as it could encourage reduced use of examinations (*e.g.*, as shown by Ishizuka and colleagues with a gamified tool, 2023).

To conclude and contextualize the importance of the topic studied, the study of physiotherapists' clinical reasoning will become all the more important as the profession progressively enters the paradigm of direct access to care in France (Aron *et al*., 2023; Bastide & Nouvel, 2022; Bruant-Bisson *et al*., 2022; Östhols *et al*., 2019). Physiotherapists' reasoning and therapeutic strategies will therefore be the first clinical steps undertaken on a patient's case, which calls for even greater vigilance and rigour.

[17]For the French context specifically, given that physiotherapists do not yet prescribe tests, resource consumption could be reflected in the proposition of additional sessions or referrals of patients at an advanced stage of their condition. Future studies could investigate the indirect costs of resource consumption using these metrics.

References

Almond, A., Zou, Y., & Forbes, R. (2021). Navigating diagnostic uncertainty in musculoskeletal practice: The perspectives and experiences of new graduate physiotherapists. *Musculoskeletal Science and Practice*, *52*, 102354. https://doi.org/10.1016/j.msksp.2021.102354

Aron, A., Cunningham, S., Yoder, I., Gravley, E., Brown, O., & Dickson, C. (2023). Diagnostic momentum in physical therapy clinical reasoning. *Journal of Evaluation in Clinical Practice*. https://doi.org/10.1111/jep.13884

Bastide, N., & Nouvel, L. (2022). L'accès direct aux soins de masso-kinésithérapie. Partie 2: vers un nécessaire changement de paradigme de la prescription médicale. *Kinésithérapie, la Revue*, *22*(244), 33-39. https://doi.org/10.1016/j.kine.2021.09.010

Begin, A. S., Hidrue, M. K., Lehrhoff, S., Lennes, I. T., Armstrong, K., Weilburg, J. B., del Carmen, M. G., & Wasfy, J. H. (2022). Association of Self-reported Primary Care Physician Tolerance for Uncertainty With Variations in Resource Use and Patient Experience. *JAMA Network Open, 5*(9), e2229521. https://doi.org/10.1001/jamanetworkopen.2022.29521

Benishek, L. E., Weaver, S. J., & Newman-Toker, D. E. (2015). The cognitive psychology of diagnostic errors. *Scientific American Neurology*. https://doi.org/10.2310/NEURO.6288

Berg, C., Philipp, R., & Taff, S. D. (2021). Scoping review of critical thinking literature in healthcare education. *Occupational Therapy In Health Care*, 1-18. https://doi.org/10.1080/07380577.2021.1879411

Berner, E. S., & Graber, M. L. (2008). Overconfidence as a cause of diagnostic error in medicine. *The American Journal of Medicine, 121*(5), S2-S23. https://doi.org/10.1016/j.amjmed.2008.01.001

Bonilauri Ferreira, A. P. R., Ferreira, R. F., Rajgor, D., Shah, J., Menezes, A., & Pietrobon, R. (2010). Clinical reasoning in the real world is mediated by bounded rationality: implications for diagnostic clinical practice guidelines. *Plos One, 5*(4), e10265. https://doi.org/10.1371/journal.pone.0010265

Bruant-Bisson, A., Laffon, P., & Marty, M. (2022). *Expérimentation de l'accès direct aux actes de masso-kinésithérapie.* https://www.igas.gouv.fr/IMG/pdf/2021-092r.pdf

Brun, C., Akinyemi, A., Houtin, L., Zerhouni, O., Monvoisin, R., & Pinsault, N. (2023). DecidHealth : Psychometric validation of the IUS-12 and IUS- 12-H scales on French healthcare professionals. https://psyarxiv.com/q2uer

Brun, C., Akinyemi, A., Houtin, L., Zerhouni, O., Monvoisin, R., & Pinsault, N. (2022a). Intolerance of Uncertainty and Attitudes towards Vaccination Impact Vaccinal Decision While Perceived Uncertainty Does Not. *Vaccines*, *10*(10), 1742. https://doi.org/10.3390/vaccines10101742

Brun, C., Zerhouni, O., Akinyemi, A., Houtin, L., Monvoisin, R., & Pinsault, N. (2022b). Impact of uncertainty intolerance on clinical reasoning: A scoping review of the 21st-century literature. *Journal of Evaluation in Clinical Practice*, *29*(3), 539-553. https://doi.org/10.1111/jep.13756

Brunner, E., Dankaerts, W., O'Sullivan, K., Meichtry, A., Bauer, C., & Probst, M. (2021). Associations between alliance, physiotherapists' confidence in managing the patient and patient-reported distress in chronic low back pain practice. *European Journal of Physiotherapy*, *23*(3), 196-200. https://doi.org/10.1080/21679169.2019.1672786

Carleton, R. N., Norton, M. A. P. J., & Asmundson, G. J. G. (2007). Fearing the unknown: A short version of the Intolerance of Uncertainty Scale. *Journal of Anxiety Disorders*, *21*(1), 105-117. https://doi.org/10.1016/j.janxdis.2006.03.014

Chakravarty, A. (2022). *Pitfalls in the diagnosis of neurological disorders*. Jaypee Brothers

Medical. https://www.google.fr/books/edition/Pitfalls_in_the_Diagnosis_of_Neurologica/E6S7EAAAQBAJ?hl=fr&gbpv=1&dq=Pitfalls+in+The+Diagnosis+of+Neurological+Disorders&printsec=frontcover

Cheng, L., & Senathirajah, Y. (2023). Using Clinical Data Visualizations in Electronic Health Record User Interfaces to Enhance Medical Student Diagnostic Reasoning: Randomized Experiment. *JMIR Human Factors*, *10*, e38941. https://doi.org/10.2196/38941

Clarke, S. O., Ilgen, J. S., & Regehr, G. (2023). Fostering Adaptive Expertise Through Simulation. *Academic Medicine*, 10-1097. https://doi.org/10.1097/ACM.0000000000005257

Cook, C. (2023). *Using Unfolding Case Studies to Develop Critical Thinking in First-Semester Nursing Students* [Doctoral dissertation, Université Gardner-Webb]. https://digitalcommons.gardner-webb.edu/cgi/viewcontent.cgi?article=1081&context=nursing-dnp

Cook, D. A., & Triola, M. M. (2009). Virtual patients: a critical literature review and proposed next steps. *Medical Education*, *43*(4), 303-311. https://doi.org/10.1111/j.1365-2923.2008.03286.

Cowell, I., McGregor, A., O'Sullivan, P., O'Sullivan, K., Poyton, R., & Murtagh, G. (2023). Physiotherapists' perceptions on using a multidimensional clinical reasoning form during psychologically informed training for low back pain. *Musculoskeletal Science and Practice*, 102797. https://doi.org/10.1016/j.msksp.2023.102797

Croskerry, P. (2009). A universal model of diagnostic reasoning. *Academic Medicine, 84*(8), 1022-1028. https://doi.org/10.1097/ACM.0b013e3181ace703

Croskerry, P., & Norman, G. (2008). Overconfidence in clinical decision making. The

American Journal of Medicine, 121(5), S24-S29. https://doi.org/10.1016/j.amjmed.2008.02.001

Croskerry, P., Singhal, G., & Mamede, S. (2013). Cognitive debiasing 1: origins of bias and theory of debiasing. *BMJ Quality & Safety*, *22*(Suppl 2), ii58-ii64. https://doi.org/10.1136/bmjqs-2012-001712

Crowley, R. S., Legowski, E., Medvedeva, O., Reitmeyer, K., Tseytlin, E., Castine, M., Jukic, D., & Mello-Thoms, C. (2013). Automated detection of heuristics and biases among pathologists in a computer-based system. *Advances in Health Sciences Education, 18*(3), 343-363. https://doi.org/10.1007/s10459-012-9374-z

Dahm, M. R., & Crock, C. (2022). Understanding and communicating uncertainty in achieving diagnostic excellence. *JAMA*, *327*(12), 1127-1128. https://doi.org/10.1001/jama.2022.2141

Daniel, M., Rencic, J., Durning, S. J., Holmboe, E., Santen, S. A., Lang, V., Ratcliffe, T., Gordon, D., Heist, B., Lubarsky, S., Estrada, C. A., Ballard, T., Artino, A. R., Sergio Da Silva, A., Cleary, T., Stojan, J., & Gruppen, L. D. (2019). Clinical Reasoning Assessment Methods: A Scoping Review and Practical Guidance. *Academic Medicine, 94*(6), 902-912. https://doi.org/10.1097/acm.0000000000002618

Downing, S. M. (2003). Validity: on the meaningful interpretation of assessment data. *Medical Education*, *37*(9), 830-837. https://doi.org/10.1046/j.1365-2923.2003.01594.x

Duong, Q. H., Pham, T. N., Reynolds, L., Yeap, Y., Walker, S., & Lyons, K. (2023). A scoping review of therapeutic reasoning process research. *Advances in Health Sciences Education*, 1-22. https://doi.org/10.1007/s10459-022-10187-7

Durning, S. J., Trowbridge, R. L., & Schuwirth, L. (2020). Clinical reasoning and diagnostic

error- a call to merge two worlds to improve patient care. *Academic Medicine, 95*(8), 1159-1161. https://doi.org/10.1097/ACM.0000000000003041

Elvén, M., Hochwälder, J., Dean, E., & Söderlund, A. (2018). Development and initial evaluation of an instrument to assess physiotherapists' clinical reasoning focused on clients' behavior change. *Physiotherapy Theory and Practice*, *34*(5), 367-383. https://doi.org/10.1080/09593985.2017.1419521

Fernandez-Branson, C., Williams, M., Chan, T. M., Graber, M. L., Lane, K. P., Grieser, S., Landis-Lewis, Z., Cooke, J., Upadhyay, D. K., Mondoux, S., Singh, H., Zwaan, L., Friedman, C., & Olson, A. P. J. (2021). Improving diagnostic performance through feedback: the Diagnosis Learning Cycle. *BMJ Quality & Safety, 30*(12), 1002-1009. https://doi.org/10.1136/bmjqs-2020-012456

Fink, M. C., Heitzmann, N., Reitmeier, V., Siebeck, M., Fischer, F., & Fischer, M. R. (2023). Diagnosing virtual patients: the interplay between knowledge and diagnostic activities. *Advances in Health Sciences Education*, 1-20. https://doi.org/10.1007/s10459-023-10211-4

Finucane, L. M., Downie, A., Mercer, C., Greenhalgh, S. M., Boissonnault, W. G., Pool-Goudzwaard, A. L., Beneciuk, J. M., Leech, R. L., & Selfe, J. (2020). International Framework for Red Flags for Potential Serious Spinal Pathologies. *Journal of Orthopaedic & Sports Physical Therapy, 50*(7), 350-372. https://doi.org/10.2519/jospt.2020.9971

Finucane, L., Greenhalgh, S., & Selfe, J. (2017). What are the Red flags to aid the early detection of metastatic bone disease as a cause of back pain?. *Physiotherapy Practice and Research*, *38*(2), 73-77.

Fourré, A., Vanderstraeten, R., Ris, L., Bastiaens, H., Michielsen, J., Demoulin, C., Darlow,

B., & Roussel, N. (2023). Management of Low Back Pain: Do Physiotherapists Know the Evidence-Based Guidelines? *International Journal of Environmental Research and Public Health, 20*(9), 5611. https://doi.org/10.3390/ijerph20095611

Graber, M. L. (2022). Reaching 95%: decision support tools are the surest way to improve diagnosis now. *BMJ Quality & Safety, 31*(6), 415-418. https://doi.org/10.1136/bmjqs-2021-014033

Haber, T., Hinman, R. S., Dobson, F., Vicenzino, B., Darlow, B., Kayll, S., & Hall, M. (2023). Clinical reasoning in managing chronic hip pain: One in two Australian and New Zealand physiotherapists diagnosed a case vignette with clinical criteria for hip OA as hip OA. A cross‐sectional survey. *Musculoskeletal Care*. https://doi.org/10.1002/msc.1751

Hage, R., Fourré, A., Ramonfosse, L., Leteneur, S., Jones, M., & Dierick, F. (2022). Description and rules of a new card game to learn clinical reasoning in musculoskeletal physiotherapy. *Journal of Manual & Manipulative Therapy*, 1-10.https://doi.org/10.1080/10669817.2022.2132346

Hamzeh, H., Madi, M., & Hensman, M. (2021). Using the diagnostic thinking inventory in musculoskeletal physiotherapy: a validity and reliability study. *Physiotherapy Research International*, *26*(2), e1895. https://doi.org/10.1002/pri.1895

Harada, T., Miyagami, T., Kunitomo, K., & Shimizu, T. (2021). Clinical decision support systems for diagnosis in primary care: A scoping review. *International Journal of Environmental Research and Public Health, 18*(16), 8435. https://doi.org/10.3390/ijerph18168435

Henschke, N., Maher, C. G., Ostelo, R. W., de Vet, H. C., Macaskill, P., & Irwig, L. (2013).

Red flags to screen for malignancy in patients with low‐back pain. *Cochrane database of systematic reviews*, (2). https://doi.org/10.1002/14651858.CD008686.pub2

Høegh-Larsen, A. M., Gonzalez, M. T., Reierson, I. Å., Husebø, S. I. E., Hofoss, D., & Ravik, M. (2023). Nursing students' clinical judgment skills in simulation and clinical placement: a comparison of student self-assessment and evaluator assessment. *BMC Nursing*, *22*(1), 64. https://doi.org/10.1186/s12912-023-01220-0

Hofmann, B. (2023). Temporal uncertainty in disease diagnosis. *Medicine, Health Care and Philosophy*, 1-11. https://doi.org/10.1007/s11019-023-10154-y

Holdar, U., Wallin, L., & Heiwe, S. (2013). Why do we do as we do? Factors influencing clinical reasoning and decision‐making among physiotherapists in an acute setting. *Physiotherapy Research International*, *18*(4), 220-229. https://doi.org/10.1002/pri.1551

Hoy, D., Brooks, P., Blyth, F., & Buchbinder, R. (2010). The epidemiology of low back pain. *Best practice & research Clinical rheumatology*, *24*(6), 769-781. https://doi.org/10.1016/j.berh.2010.10.002

Infante Guedes, A., Quintana Montesdeoca, M. D. P., & Etopa Bitata, M. D. P. (2023). Educational Needs for Coping with Clinical Uncertainty in Physiotherapy. *Education Sciences*, *13*(7), 657. https://doi.org/10.3390/educsci13070657

Ingram, S., Stenner, R., & May, S. (2023). The experiences of uncertainty amongst musculoskeletal physiotherapists in first contact practitioner roles within primary care. *Musculoskeletal Care*. https://doi.org/10.1002/msc.1735

Iuliucci, N. (2016). *Patient History Elicitation and Diagnostic Decision Making*. Rochester Institute of Technology.

Jette, D. U., Ardleigh, K., Chandler, K., & McShea, L. (2006). Decision-making ability of

physical therapists: physical therapy intervention or medical referral. *Physical Therapy*, *86*(12), 1619-1629. https://doi.org/10.2522/ptj.20050393

Josephson, I., Bülow, P., & Hedberg, B. (2011). Physiotherapists' clinical reasoning about patients with non-specific low back pain, as described by the International Classification of Functioning, Disability and Health. *Disability and Rehabilitation*, *33*(22-23), 2217-2228. https://doi.org/10.3109/09638288.2011.563819

Kamath, A., & Ullal, S. D. (2023). Learning and clinical reasoning experience of second-year medical pharmacology students and teachers with virtual patients developed using OpenLabyrinth. *Electronic Journal of General Medicine*, *20*(5). https://doi.org/10.29333/ejgm/13289

Kechichian, A., & Pinsault, N. (2023). Modèles de partage des tâches impliquant les kinésithérapeutes pour la prise en charge des troubles musculosquelettiques en soins primaires : une revue narrative. *Santé Publique*, *35*(2).

Keller, F., Allet, L., Meichtry, A., Scascighini, L., Scheermesser, M., Wirz, M., & Nast, I. (2022). Diagnostic and decision-making abilities of Swiss physiotherapists in a simulated direct access setting. *Physiotherapy Theory and Practice*, 1-16. https://doi.org/10.1080/09593985.2022.2077269

Kerr, A., & Thompson, C. (2022). Medical Students' Reactions to Uncertainty During Clinical Rotations. *Family Medicine*, *54*(4), 285-289. https://doi.org/10.22454/FamMed.2022.947719

Kourtidis, P., Nurek, M., Delaney, B., & Kostopoulou, O. (2022). Influences of early diagnostic suggestions on clinical reasoning. *Cognitive Research: Principles and Implications*, *7*(1), 1-17. https://doi.org/10.1186/s41235-022-00453-y

Lackenbauer, W., Gasselich, S., Lickel, M. E., Beikircher, R., Keip, C., Rausch, F., Wieser,

M., Selfe, J., & Janssen, J. (2023). The Ability of Austrian Qualified Physiotherapists to Make Accurate Keep-Refer Decisions and to Detect Serious Pathologies Based on Clinical Vignettes: Protocol for a Cross-sectional Web-Based Survey. *JMIR Research Protocols, 12*, e43028. https://doi.org/10.2196/43028

Langridge, N., Roberts, L., & Pope, C. (2015). The clinical reasoning processes of extended scope physiotherapists assessing patients with low back pain. *Manual Therapy*, *20*(6), 745-750. https://doi.org/10.1016/j.math.2015.01.005

Lee, C. Y., Lai, H. Y., Lee, C. H., & Chen, M. M. (2022). Medical Dispute Cases Caused by Errors in Clinical Reasoning: An Investigation and Analysis. *Healthcare*, *10*(11), 2224. https://doi.org/10.3390/healthcare10112224

Makridakis, S., Kirkham, R., Wakefield, A., Papadaki, M., Kirkham, J., & Long, L. (2019). Forecasting, uncertainty and risk; perspectives on clinical decision-making in preventive and curative medicine. *International Journal of Forecasting, 35*(2), 659-666. https://doi.org/10.1016/j.ijforecast.2017.11.003

Mamede, S., Van Gog, T., Van Den Berge, K., Van Saase, J. L., & Schmidt, H. G. (2014). Why do doctors make mistakes? A study of the role of salient distracting clinical features. *Academic Medicine, 8*9(1), 114-120. https://doi.org/10.1097/ACM.0000000000000077

Meidinger, P., Kéchichian, A., & Pinsault, N. (2023). A model of triage of serious spinal pathologies and therapeutic options based on a Delphi study. *Medicina*, *59*(7), 1283. https://doi.org/10.3390/medicina59071283

Meucci, R. D., Fassa, A. G., & Faria, N. M. X. (2015). Prevalence of chronic low back pain: systematic review. *Revista de Saúde Pública*, *49*.https://doi.org/10.1590/S0034-8910.2015049005874

Microsoft. (2023). *Qu'est-ce que Power Apps ?*. Microsoft.com. Retrieved on the 24 march,

2023, from https://learn.microsoft.com/fr-fr/power-apps/powerapps-overview

Miller, D. J., Spengler, E. S., & Spengler, P. M. (2015). A meta-analysis of confidence and judgment accuracy in clinical decision making. *Journal of Counseling Psychology*, *62*(4), 553.https://doi.org/10.1037/cou0000105

Ministère de la santé et de la prévention. (2023). *Masseur-kinésithérapeute*. Sante.gouv. Retrieved on the 30 march, 2023, from https://sante.gouv.fr/metiers-et-concours/les-metiers-de-la-sante/les-fiches-metiers/article/masseur-kinesitherapeute

Monteiro, S. D., Sherbino, J. D., Ilgen, J. S., Dore, K. L., Wood, T. J., Young, M. E., Bandiera, G., Blouin, D., Gaissmaier, W., Norman, G. R., & Howey, E. (2015). Disrupting Diagnostic Reasoning: Do Interruptions, Instructions, and Experience Affect the Diagnostic Accuracy and Response Time of Residents and Emergency Physicians? *Academic Medicine, 90*(4), 511-517. https://doi.org/10.1097/acm.0000000000000614

Morgan, D. J., Pineles, L., Owczarzak, J., Magder, L., Scherer, L., Brown, J. P., Pfeiffer, C., Terndrup, C., Leykum, L., Feldstein, D., Foy, A., Stevens, D., Koch, C., Masnick, M., Weisenberg, S., & Korenstein, D. (2021). Accuracy of Practitioner Estimates of Probability of Diagnosis Before and After Testing. *JAMA Internal Medicine, 181*(6), 747. https://doi.org/10.1001/jamainternmed.2021.0269

Nendaz, M., & Perrier, A. (2012). Diagnostic errors and flaws in clinical reasoning: mechanisms and prevention in practice. *Swiss Medical Weekly*, *142*(4344), w13706-w13706. https://doi.org/10.4414/smw.2012.13706

Noll, E., Key, A., & Jensen, G. (2001). Clinical reasoning of an experienced physiotherapist: insight into clinician decision‐making regarding low back pain. *Physiotherapy Research International*, *6*(1), 40-51. https://doi.org/10.1002/pri.212

Norman, G. R., & Eva, K. W. (2010). Diagnostic error and clinical reasoning. *Medical Education, 44*(1), 94-100. https://doi.org/10.1111/j.1365-2923.2009.03507.x

Norman, G. R., Monteiro, S. D., Sherbino, J., Ilgen, J. S., Schmidt, H. G., & Mamede, S. (2017). The causes of errors in clinical reasoning: cognitive biases, knowledge deficits, and dual process thinking. *Academic Medicine, 92*(1), 23-30. https://doi.org/10.1097/ACM.0000000000001421

O'Keefe, K. (2023). *An exploration of radiographer decision-making regarding rejected or sub-optimal plain x-ray images* [Doctoral dissertation, Queensland University of Technology].https://eprints.qut.edu.au/238332/7/Katrina_OKeefe_Thesis.pdf

Olson, A. P., Linzer, M., & Schiff, G. D. (2021). Measuring and improving diagnostic safety in primary care: Addressing the "Twin" pandemics of diagnostic error and clinician burnout. *Journal of General Internal Medicine, 36*(5), 1404-1406. https://doi.org/10.1007/s11606-021-06611-0

Östhols, S., Boström, C., & Rasmussen-Barr, E. (2019). Clinical assessment and patient-reported outcome measures in low-back pain–a survey among primary health care physiotherapists. *Disability and Rehabilitation*, *41*(20), 2459-2467. https://doi.org/10.1080/09638288.2018.1467503

Patel, P., Hancock, J., Rogers, M., & Pollard, S. R. (2022). Improving uncertainty tolerance in medical students: A scoping review. *Medical Education*. https://doi.org/10.1111/medu.14873

Pelaccia, T., Tardif, J., Triby, E., Ammirati, C., Bertrand, C., Dory, V., & Charlin, B. (2014). How and when do expert emergency physicians generate and evaluate diagnostic hypotheses? A qualitative study using head-mounted video cued-recall interviews. *Annals of Emergency Medicine*, *64*(6), 575-585. https://doi.org/10.1016/j.annemergmed.2014.05.003

Peterson, B. D., Magee, C. D., Martindale, J. R., Dreicer, J. J., Mutter, M. K., Young, G., Sacco, M. J., Parsons, L. C., Collins, S. R., Warburton, K. M., & Parsons, A. S. (2022). REACT: Rapid Evaluation Assessment of Clinical Reasoning Tool. *Journal of General Internal Medicine, 37*(9), 2224-2229. https://doi.org/10.1007/s11606-022-07513-5

Redeker, I., Callhoff, J., Hoffmann, F., Haibel, H., Sieper, J., Zink, A., & Poddubnyy, D. (2019). Determinants of diagnostic delay in axial spondyloarthritis: an analysis based on linked claims and patient-reported survey data. *Rheumatology*, *58*(9), 1634-1638. https://doi.org/10.1093/rheumatology/kez090

Sandhu, T. R., Xiao, B., & Lawson, R. P. (2023). Transdiagnostic computations of uncertainty: towards a new lens on intolerance of uncertainty. *Neuroscience & Biobehavioral Reviews*, 105123. https://doi.org/10.1016/j.neubiorev.2023.105123

Schiff, G. D., Kim, S., Abrams, R., Cosby, K., Lambert, B., Elstein, A. S., & Mcnutt. (2005). Diagnosing diagnosis errors: lessons from a multi-institutional collaborative project. *Advances in Patient Safety: From Research to Implementation*, *2*.

Schmutz, J., Eppich, W. J., Hoffmann, F., Heimberg, E., & Manser, T. (2014). Five steps to develop checklists for evaluating clinical performance: an integrative approach. *Academic Medicine*, *89*(7), 996-1005. https://doi.org/10.1097/ACM.0000000000000289

Scott, I. A., Hubbard, R. E., Crock, C., Campbell, T., & Perera, M. (2021). Developing critical thinking skills for delivering optimal care. *Internal Medicine Journal, 51*(4), 488-493. https://doi.org/10.1111/imj.15272

Simmonds, M. J., Derghazarian, T., & Vlaeyen, J. W. (2012). Physiotherapists' knowledge,

attitudes, and intolerance of uncertainty influence decision making in low back pain. *The Clinical Journal of Pain*, *28*(6), 467-474. https://doi.org/10.1097/AJP.0b013e31825bfe65

Staal, J., Alsma, J., Mamede, S., Olson, A. P. J., Prins-van Gilst, G., Geerlings, S. E., Plesac, M., Sundberg, M. A., Frens, M. A., Schmidt, H. G., Van den Broek, W. W., & Zwaan, L. (2021). The relationship between time to diagnose and diagnostic accuracy among internal medicine residents: a randomized experiment. *BMC Medical Education, 21*(1). https://doi.org/10.1186/s12909-021-02671-2

Staal, J., Hooftman, J., Gunput, S. T. G., Mamede, S., Frens, M. A., Van den Broek, W. W., Alsma, J., & Zwaan, L. (2022). Effect on diagnostic accuracy of cognitive reasoning tools for the workplace setting: systematic review and meta-analysis. *BMJ Quality & Safety, 31*(12), 899-910. https://doi.org/10.1136/bmjqs-2022-014865

Staal, J., Katarya, K., Speelman, M., Brand, R., Alsma, J., Sloane, J., Van den Broek, W. W., & Zwaan, L. (2023). Impact of performance and information feedback on medical interns' confidence-accuracy calibration. *Advances in Health Sciences Education*. https://doi.org/10.1007/s10459-023-10252-9

Sutcliffe, P., Connock, M., Shyangdan, D., Court, R., Kandala, N. B., & Clarke, A. (2013). A systematic review of evidence on malignant spinal metastases: natural history and technologies for identifying patients at high risk of vertebral fracture and spinal cord compression. *Health Technology Assessment*, *17*(42), 1. https://doi.org/10.3310/hta17420

Tehrani, A. S. S., Lee, H., Mathews, S. C., Shore, A., Makary, M. A., Pronovost, P. J., & Newman-Toker, D. E. (2013). 25-Year summary of US malpractice claims for diagnostic errors 1986–2010: an analysis from the National Practitioner Data Bank. *BMJ Quality & Safety*, *22*(8), 672-680. https://doi.org/10.1136/bmjqs-2012-001550

Vreugdenhil, J., Somra, S., Ket, H., Custers, E. J., Reinders, M. E., Dobber, J., & Kusurkar, R. A. (2023). Reasoning like a doctor or like a nurse? A systematic integrative review. *Frontiers in Medicine*, *10*. https://doi.org/10.3389/fmed.2023.1017783

World Health Organization. (n.d.). *International Statistical Classification of Diseases and Related Health Problems.* WHO.int. Retrieved on the 24 july, 2023, from https://www.who.int/standards/classifications/classification-of-diseases

Yanagita, Y., Shikino, K., Ishizuka, K., Uchida, S., Li, Y., Yokokawa, D., Tsukamoto, T., Noda, K., Uehara, T., & Ikusaka, M. (2023). Improving diagnostic accuracy using a clinical diagnostic support system for medical students during history-taking: a randomized clinical trial. *BMC Medical Education, 23*(1). https://doi.org/10.1186/s12909-023-04370-6

Zwaan, L., & Hautz, W. E. (2019). Bridging the gap between uncertainty, confidence and diagnostic accuracy: calibration is key. *BMJ Quality & Safety, 28*(5), 352-355. https://doi.org/10.1136/bmjqs-2018-009078

Supplementary Material

Clinical Case 1 (Difficulty Level: Easy - Suggestive of common non-specific lumbago)

Block 1: History
Presentation: A 44-year-old man comes for consultation with low back pain.
Question: What is the patient's height and weight?
Answer: 1.85 m, 80 kg
Question: What is the patient's social situation?
Answer: The patient is a trader, currently going through divorce proceedings with his ex-wife.
Question: When did the symptoms appear?
Answer: 6 days ago
Question: How did the pain start?
Answer: He woke up one morning with a pain in his back for no particular reason.
Question: How have the symptoms developed since they first appeared?
Answer: Since the onset of symptoms, he has found it very difficult to sit down. He manages to reduce these symptoms by walking for more than 15 minutes, but the pain soon returns, accompanied by a pain in his back.
Question: What sensations are associated with these symptoms?
Answer: The pain in his back bothers him in his daily activities.
Question: Are there any positions that reproduce the pain?
Answer: He is particularly uncomfortable sitting down. He says that he also has difficulty with everyday tasks such as putting on trousers and socks. If he does it too suddenly, it triggers severe back pain.
Question: Does the patient have any medical history or current treatments?
Answer: The patient has type 1 diabetes and ruptured his right anterior cruciate ligament 5 years ago.
Question: How does he sleep?
Answer: The patient tells you that he is bothered by pain at night, but that he still gets plenty of restful sleep.
Question: What are the patient's leisure activities?
Answer: He's not doing any physical activity at the moment, as he doesn't have the time because of his work schedule.
Question: What impact do the symptoms have on the patient?
Answer: He finds it difficult to concentrate at work. This handicaps him, especially as the atmosphere at work is not good. Recently, the company's results have not been good, which adds to the pressure from his superiors and the goals to be achieved.
Question: What are the patient's main objectives and expectations in terms of treatment?
Answer: To improve his situation, to be able to work without discomfort, to resume physical activity

Block 2: Physical examination
Question: What visual observations do you make?
Answer: You look at the patient from the front and notice that he is standing relatively straight, with his left shoulder higher than his right. He also has varus knees. From the back, the left scapula is more protruding than the right, and the head is tilted more to the left. In

profile, the lumbar lordosis is reduced. There are no skin or trophic signs (redness, scarring, swelling in the area of the patient's pain).
Question: Are there any identified nerve disorders?
Answer: No central or peripheral neurological disorders (large and small fibres) were identified.
Question: What are the main findings of the physical examination (non-exhaustive list)?
Answer :
Palpatory: sensitive lumbar spinal muscles, reproduce the patient's pain. Mechanosensitivity not reproduced by palpation.

Clinical tests:

- Gillet test: negative
- PKB (Prone Knee Bend test)/SKB (Slump Knee Bend test)/Léri's sign: negative
- Schober's test: +17 cm
- Fingertip-to-floor test: no discomfort, 22 cm distance from the ground
- Patrick's test (=FABER test): negative on both sides
- Slump test: negative
- Downing test (lower limb lengthening/shortening test): negative
- Cluster of Laslett: 0/5
- Lasègue test/SLR (Straight Leg Raise): negative
- DN4 (Douleur Neuropathique 4): negative
- No directional preference identified

Question: What are the results of the specific questionnaires carried out?
Answer

- Orëbro < 49 (low risk of future work disability)
- Start back tools: 3/9 (low risk of chronicity)
- FABQ (Fear-Avoidance Beliefs Questionnaire): 23/96 (the higher the score, the higher the beliefs and fears in relation to work and physical activity)
- Roland-Morris Disability Questionnaire: 6/24 (the higher the score, the greater the functional impact of low back pain)

Block 3: Treatments used so far
Question: What non-drug strategies have been used to ease his pain?
Answer: He applied a hot water bottle to the painful area, but it didn't really improve.
Question: What medication was used to relieve his pain?
Answer: He didn't take any medication.
Question: In view of all this, you decide to summarize the information you have gathered.
Answer: The patient goes on to thank you for understanding his problem, adding that he spends about 10 to 12 hours sitting down because of his work.

Block 4: Imaging
Question: Would you like an X-Ray?
Answer:
X-Ray: Mild right lumbar curve can be seen.. There is degenerative disc disease in L5S1 and to a lesser extent in L4L5, associated with severe degenerative facet joint arthrosis from L4L5 through L5S1. No evidence of lumbar spinal stenosis. No abnormalities of the sacroiliac joints. No hip joint abnormalities. No significant bone abnormality, in particular no compression or sacral abnormality.
Question: Would you like a CT scan?
Answer:
CT scan: Spinal canal dimensions are normal. No evidence of disc bulges from L3L4 to L4L5. Left facet joint osteoarthritis at the L4L5 level with inferior and superior facet hypertrophy with patent left intervertebral foramen. At the L5S1 level, there is a left posterolateral disc bulge without causing nerve root impingement. There were no abnormalities in the sacroiliac or hip joints.
Question: Would you like an MRI scan?
Answer:
MRI: The thecal sac and conus are normal. Vertebral body marrow signal and disc signal are normal. At L3L4 and L4L5 levels, moderate degenerative disc disease without disc protrusion. Mild Ligamentum flavum hypertrophy is seen at the L4L5 level. At the L5S1 level, diffuse intervertebral disc bulge with a mild central protrusion. Vertebral body degenerative changes seen on multiple levels, especially from L4L5 through L5D1. No central canal stenosis. The sacroiliac and hip joints are unremarkable.
Question: Would you like a scintigraphy?
Answer:
Scintigraphy: Scintigraphy with no particular abnormalities.
Question: Would you like a biological examination?
Answer:

- Blood test (Reference value = RV)
- Red blood cells 5.3 M/mm3 (RV 4.50-6.50)
- Haemoglobin 14.1 g/100 mL (RV 13.0-17.0)
- Haematocrit 44.7% (RV 40.0-54.0)
- White blood cells 8900 /mm3 (RV 4000-10000)
- Platelets 381000 /mm3 (RV 150000-500000)
- Erythrocyte Sedimentation rate 1st hour 1mm (RV<10)/ 2nd hour 2mm (RV<20)
- C-Reactive Protein (CRP) 2 mg/L (RV<6mg/L)
- Fasting blood glucose 0.82 g/L (RV 0.70-1.05)
- Creatinine 9.4 mg/L (RV 6.0-11.0)
- Calculation of estimated clearance (Cockroft and Gault formula) 113mL/min
- No HLA B27 marker

Clinical Case 2 (Difficulty Level: Medium - Suggestive of ankylosing spondylitis)

Block 1: History
Presentation: A 27-year-old man comes for consultation with low back pain.
Question: What is the patient's height and weight?
Answer: 1.75m, 68.5 kg
Question: What is the patient's social situation?
Answer: He lives with his partner. He is in his final year of a doctorate in biochemistry.
Question: When did the symptoms appear?
Answer: This episode occurred 5 days ago, but he has had similar pain several times before.
Question: How did the pain start?
Answer: The patient does not know what could have caused the pain.
Question: How have the symptoms changed since they first appeared?
Answer: The patient has not noticed any significant change since the onset of the pain.
Question: What sensations are associated with these symptoms?
Answer: When these symptoms appear, he is usually able to maintain his physical activities except during episodes of excessive pain. The patient has noticed that the onset of his painful periods is associated with stress due to professional constraints and/or excessive sporting activity. Finally, he tells you about another pain in his left heel that can sometimes bother him when running.
Question: Are there any positions that reproduce the pain?
Answer: Prolonged sitting.
Question: Does the patient have a history of the condition and any ongoing treatment?
Answer: The patient's history includes a right ankle sprain in 2017, a fractured right shoulder following a scooter fall in 2012 and psoriasis diagnosed in 2020. He is not taking any medication.
Question: How is his sleep?
Answer: He has noticed that his sleep is often altered during painful episodes and that he finds it difficult to go back to sleep once he has woken up.
Question: What are the patient's leisure activities?
Answer: He does a lot of physical activity: running, crossfit (3 to 5 times a week).
Question: What impact do the symptoms have on the patient?
Answer: He says that he has often had periods of lower back pain since he was 20. During periods of severe pain, he feels rusty, especially in the morning. He manages to live with it, but it worries him when it deprives him of activities that bring him well-being, as is currently the case. He insists on the importance of maintaining an acceptable level of physical activity and pain, now that his thesis is due.
Question: What are the patient's main objectives and expectations in terms of treatment?
Answer: To resume all physical activities as before, to put an end to the discomfort he has suffered for many years, to be helped with his pain and to find strategies for managing their back pain

Block 2: Physical examination
Question: What visual observations do you make?
Answer: You look at the patient from the front and notice that he is standing with his shoulders rolled up and that his right shoulder is higher than his left. He also has valgus knees. From the back, you can see that the spine is not aligned with the vertical but deviates to the left, with the head leaning more to the left. In profile, the thoracic kyphosis is more marked and the lumbar spine is straight. There are no skin or trophic signs (redness, scarring, swelling in the area of the patient's pain).

Question: Are there any identified nerve disorders?
Answer: No central or peripheral neurological disorders (large and small fibres) were identified.
Question: What are the main findings of the physical examination (non-exhaustive list)?
Answer:
Palpatory: sensory lumbar spines, psoas, piriformis and quadratus lumborum do not reproduce the patient's pain. Mechanosensitivity not reproduced by palpation.
Clinical tests:

- Gillet test: Positive at G (the left Posterior Superior Iliac Spine ascends at the end of flexion).
- PKB (Prone Knee Bend test)/SKB (Slump Knee Bend test)/Léri's sign: negative
- Schober test: 12cm
- Fingertip-to-floor test: reproduces pain, 18cm distance from the ground
- Patrick's test (=FABER): positive at G
- Slump test: negative
- Downing test (lower limb lengthening/shortening test): positive at D(anterior ilium)
- Cluster of Laslett: 3/5
- Lasègue test/SLR (Straight Leg Raise): negative
- DN4 (Douleur Neuropathique 4): negative
- No directional preference identified

Question: What are the results of the specific questionnaires carried out?
Answer:

- Orëbro > 49 (higher risk of future work disability).
- Start back tools: 4/9 and sub questionnaire < 4 (average risk of chronicity)
- FABQ (Fear-Avoidance Beliefs Questionnaire): 16/96 (the higher the score, the higher the beliefs and fears in relation to work and physical activity)
- Roland-Morris Disability Questionnaire: 5/24 (the higher the score, the greater the functional impact of low back pain)

Block 3: Treatments used so far
Question: What non-drug strategies have been used to ease his pain?
Answer: He limits physical activity when the pain is severe. When it's bearable, he tries to go for a walk, which seems to do him good.
Question: What drugs have been used to relieve his pain?
Answer: He says that during painful episodes he is prescribed non-steroidal anti-inflammatory drugs (NSAIDs), which give him relief but cause him stomach pains. He would like to stop taking NSAIDs.
Question: In view of all this, you decide to summarise the information you have gathered.
Answer: The patient goes on to confirm that you have understood his problem and sees nothing more to add.

Block 4: Imaging
Question: Would you like an X-Ray?
Answer:
X-Ray: A pelvic obliquity on the left is seen in static position.. Examination reveals mainly multilevel degenerative disc disease with pinching lateral to the left at L3-L4, pinching lateral to the right at the L4-L5 level and the L5-S1 level. Presence of sacroiliitis. No hip joint abnormalities. No significant bone abnormality, in particular no compression or sacral abnormality.

Question: Would you like a CT scan?
Answer:
CT scan: No lumbar static abnormalities. No transitional anomalies. No lytic or condensing bone lesions. No evidence for degenerative changes of the vertebral bodies or facet joints osteoarthritis.. Spinal canal dimensions are normal. No spontaneously detectable paravertebral soft tissue abnormalities. Medial and left paramedian disc protrusion, filling the anterior epidural fat and exerting a mass effect on the thecal sac and the homolateral L3 nerve root at the level of its emergence. Right disc protrusion at the L4L5 level without impingement. Discrete paramedian disc protrusion at the L5S1 level. Sacroilitis is identified on the CT scan.
Question: Would you like an MRI scan?
Answer:
MRI: Degenerative disc disease is seen in the last three lumbar levels with disk space narrowing and posterior disc protrusion. At L3-L4 level: there is a left posterolateral disc protrusion in contact with the dural sac. At L4-L5 level: mild disc protrusion lateralized on the right side. At L5-S1: overall disc bulge with a medial disc protrusion. No clear discovertebral anomaly of the dorsal spine. No spinal cord abnormality. No suspicious bony abnormality. No facet joints anomaly. No significant central canal stenosis. No significant foraminal stenosis. Pseudo widening of the sacroiliac joint space with bone oedema in T2 sequence suggestive of sacroiliitis. The hip joints are intact.
Question: Would you like a scintigraphy?
Answer: Scintigraphy with no particular abnormalities.
Question: Would you like a biological examination?
Answer: Blood test (Reference value = RV)

- Red blood cells 5.1 M/mm3 (RV 4.50-6.50)
- Haemoglobin 13.9 g/100 mL (RV 13.0-17.0)
- Haematocrit 47.4% (RV 40.0-54.0)
- Leukocytes 8800 /mm3 (RV 4000-10000)
- Platelets 302000 /mm3 (RV 150000-500000)
- Sedimentation rate 1st hour 8mm (RV<10)/ 2nd hour 19mm (RV<20)
- C-Reactive Protein (CRP) 15 mg/L (RV<6mg/L)
- Fasting blood glucose 1.01g/L (RV 0.70-1.05)
- Creatinine 9.7 mg/L (RV 7.0-13.0)
- Calculation of estimated clearance (Cockroft and Gault formula) 111 mL/min
- No HLA B27 marker

Clinical Case 3 (Difficulty Level: Difficult - Suggestive of metastatic bone disease in the thoracolumbar spine)

Block 1: History
Presentation: A 54-year-old woman comes for consultation with low back pain.
Question: What is the patient's height and weight?
Answer: 1.62 m, 58.5 kg
Question: What is the patient's social situation?
Answer: She is married and lives in a first floor flat. She is a secondary school teacher.
Question: When did the symptoms first appear?
Answer: Her pain started 3 or 4 weeks ago.
Question: How did the pain start?
Answer: She can't think of any particular reason, perhaps when she helped her husband unload the boot on returning from a trip.
Question: How have the symptoms developed since they first appeared?
Answer: From the beginning, the pain has gradually increased.
Question: What sensations are associated with these symptoms?
Answer: She shows you the middle and lower back. She says it goes down behind her right leg. She has also had difficulty walking since the pain increased.
Question: Are there any positions that reproduce the pain?
Answer: Nothing of note.
Question: Does the patient have any history of pain or ongoing treatment?
Answer: She remembers breaking her wrist several years ago. She has been taking Levothyrox daily since her thyroid cancer 3 years ago (the thyroid has since been surgically removed).
Question: How does she sleep?
Answer: The patient tells you that her nights are very complicated at the moment, she's been sleeping less since the pain appeared. She finds her sleep less restorative.
Question: What leisure activities does the patient do?
Answer: She goes Nordic walking every Sunday and does some painting.
Question: What impact do the symptoms have on the patient?
Answer: She does fewer things than before. She tried to resume her daily activities (Nordic walking, painting) 15 days ago, but stopped after a few days because she couldn't see any improvement. She's really worried and dreads having a pulled or torn muscle.
Question: What are the patient's main objectives and expectations in terms of treatment?
Answer: She doesn't quite understand your question, the doctor told her that with physiotherapy it would go away, simply no more pain, do what is written on the prescription, it was her doctor who told her to go and see a physiotherapist.

Block 2: Physical examination
Question: What visual observations do you make?
Answer: You look at the patient from the front and notice that she is standing with her shoulders back and that her shoulders are at the same height. In addition, her knees are valgum/varum free. From the back, there is no deviation of the spine. In profile, there is no increase in kyphosis or lordosis. There are no dermatological signs (redness, scarring, swelling in the area of the patient's pain).
Question: Have any nervous disorders been identified?
Answer: No central or peripheral neurological disorders (large and small fibres) were identified.
Question: What are the main findings of the physical examination (non-exhaustive list)?

Answer:
Palpatory: lumbar spinal tendons, psoas, piriformis and quadratus lumborum reproduce the patient's pain. Mechanosensitivity not reproduced by palpation.

Clinical tests:

- Gillet test: Positive at G (the left Posterior Superior Iliac Spine ascends at the end of flexion).
- PKB (Prone Knee Bend test)/SKB (Slump Knee Bend test)/Léri's sign: negative
- Schober's test: 13cm
- Fingertip-to-floor test: no discomfort, 25cm distance from the ground
- Patrick's test (=FABER test): positive at R and L
- Slump test: negative
- Downing test (lower limb lengthening/shortening test): short.
- Cluster of Laslett: 1/5
- Lasègue test/SLR (Straight Leg Raise): negative
- DN4 (Douleur Neuropathique 4): negative
- No directional preference identified

Question: What are the results of the specific questionnaires carried out?
Answer:

- Orëbro > 49 (higher risk of future work disability).
- Start back tools: 5/9 and sub questionnaire > 4 (high risk of chronicity)
- FABQ (Fear-Avoidance Beliefs Questionnaire): 17/96 (the higher the score, the higher the beliefs and fears in relation to work and physical activity)
- Roland-Morris Disability Questionnaire: 10/24 (the higher the score, the greater the functional impact of low back pain)

Block 3: Treatments used so far
Question: What non-drug strategies have been used to ease her pain?
Answer: Her husband offered to massage her, which felt very good at the time but didn't last over time.
Question: What medication was used to ease her pain?
Answer: She tells you that she's not a fan of medication, that she took a medicine for a week, can't remember the exact name, but it contained codeine. In any case, it hadn't helped her much.
Question: In view of all this, you decide to summarise the information you have gathered.
Answer: After the summary, the patient thinks she has forgotten to give you some information. She notes that the pain is worse at night. She says that taking the codeine-based medication has increased her stomach pains. She called her doctor to check whether it was the dosage of Levothyrox that was causing her stomach pains, as she had had the same pains at the start of treatment. However, she has no problems with bowel movements, as she had at the time.

Block 4: Imaging
Question: Would you like an X-Ray?
Answer:
X-Ray: Normal static front and side views. No significant pelvic tilt. Posterior degenerative disc disease at L5S1 level. No abnormal bone structure. Signs of incipient facet joints osteoarthritis. No evidence of lumbar spinal stenosis. Normal appearance of sacroiliac and hip joints.

Question: Would you like a CT scan?
Answer:
CT scan: No abnormalities of lumbar static views. No transitional abnormalities. Lumbar canal of normal calibre. Global paramedian disc bulge at L5S1 level. Signs of bone metastases in the vertebral bodies on several dorsolumbar levels, in relation to the patient's history of neoplasia. Slight facet joints osteoarthritic changes. No abnormalities in the sacroiliac or hip joints.
Question: Would you like an MRI scan?
Answer:
MRI: Degenerative disc disease in the last two lumbar levels with disk space narrowing and posterior disc bulges. At the L4-L5 level: discrete overall disc protrusion. At the L5-S1 level: medial disc protrusion. No significant central canal stenosis. No significant foraminal stenosis. MRI showed several focal lesions of different sizes in the vertebral bodies on several dorsolumbar levels: they appeared in high signal intensity on T1-weighted images and on T2-weighted images in favour of bone metastases. No signs of extension to adjacent soft tissues. Integrity of vasculo-nervous structures. No visible abnormalities in the sacroiliac joints. Integrity of the hip joints.
Question: Would you like a scintigraphy?
Answer:
Scintigraphy: Multiple areas of hyperfixation in the axial skeleton. The entire pelvic complex seems to be spared.
Question: Would you like a biological examination?
Answer:

- Blood test (reference value = RV) :
- Red blood cells 4.8 M/mm3 (RV 4.50-6.50)
- Haemoglobin 12.1 g/100 mL (RV 12.0-16.0)
- Haematocrit 42.4% (RV 40.0-54.0)
- Leukocytes 9100 /mm3 (RV 4000-10000)
- Platelets 358000 /mm3 (RV150000-500000)
- Sedimentation rate 1st hour 7mm (RV<10)/ 2nd hour 19mm (RV<20)
- C-Reactive Protein (CRP) 15 -20 mg/L (RV<6mg/L)
- Fasting blood glucose 0.78g/L (RV 0.70-1.05)
- Creatinine 10.1 mg/L (RV 6.0-11.0)
- Calculation of estimated clearance (Cockroft and Gault formula) 59mL/min
- No HLA B27 marker