



THE BRITISH PAIN SOCIETY

Original Article

Reproducibility of pain manikins: a comparison of paper versus online questionnaires

British Journal of Pain
7(3) 130–137
© The British Pain Society 2013
Reprints and permissions:
sagepub.co.uk/
journalsPermissions.nav
DOI: 10.1177/2049463713487895
bjp.sagepub.com



Gareth T Jones¹, Ramona Kyabaggu¹, Debbi Marais² and Gary J Macfarlane¹

Abstract

Background

In epidemiological studies, pain location is often collected by paper questionnaire using blank body manikins, onto which participants shade the location of their pain(s). However, it is unknown how reliable these will transfer to online questionnaires. The aim of the current study was to determine agreement between online- and paper-based completion of pain manikins.

Methods

A total of 264 children, aged 15–18 years, completed both an online and a paper questionnaire. Participants were asked to identify the location of their pain by highlighting predefined body areas on a manikin (online) or by shading a blank version of the manikin (paper). The difference in the prevalence of 12 regional/widespread pain conditions was determined, and agreement between online and paper questionnaires was assessed using prevalence- and bias-adjusted kappa (PABAK).

Results

For the majority of pain conditions, prevalence was higher when ascertained by paper questionnaire. However, for the most part, the difference in prevalence was modest (range: –1.1 to 5.7%) the exceptions being hip/thigh pain (difference: 10.6%) and upper back pain (difference: 14.8%). For most pain locations, there was good or very good agreement between paper and online manikins (PABAK range: 0.61 to 0.88). However, identification of pain in the hip/thigh and in the upper back had poorer agreement (PABAK: 0.49 and 0.29, respectively).

Conclusions

This is the first study to examine the reproducibility of body manikins on different media, in a population-based survey. We have shown that online manikins can be used to capture data on pain location in a manner satisfactorily comparable to paper questionnaires.

Keywords

Epidemiology, Internet, musculoskeletal pain, pain measurement, questionnaires

Introduction

The bedrock of pain epidemiology – indeed most epidemiology – is the postal questionnaire. In the field of pain specifically, manikins are commonly used for participants to identify the location of pain. They are simple to administer and have intrinsic heuristic value. They can vary from the fairly simple – two-view (front and back) manikins on which the participant shades pain location¹ – to the more complicated, on which

¹Aberdeen Pain Research Collaboration (Epidemiology Group), Institute of Applied Health Sciences, University of Aberdeen, Aberdeen, UK

²Public Health Nutrition Research Group, Institute of Applied Health Sciences, University of Aberdeen, Aberdeen, UK

Corresponding author:

Dr Gareth Jones, Epidemiology Group, Institute of Applied Health Sciences, School of Medicine and Dentistry, University of Aberdeen, Polwarth Building, Foresterhill, Aberdeen, AB25 2ZD, UK.
Email: gareth.jones@abdn.ac.uk

pain with different characteristics (stabbing, aching, shooting, etc.) is marked with different symbols.² Manikins may be either predivided into different body areas overtly, or presented blank. In the case of the latter, scoring is commonly achieved by overlaying the manikin with an acetate marked with predivided body areas, and each area is coded as having pain or no pain, according to the shading underneath. Manikins may also be employed in the clinical setting, although some argue that verbal questions are the gold standard. Face-to-face verbal questions are enhanced by the participant's ability to point to different body areas, whereas in telephone interviews this is not possible, and the researcher must rely on the participant's ability to describe the body area accurately. Thus, the paper questionnaire containing body manikins is still the most common method of identifying pain location in large-scale epidemiological studies.

Response rates to postal questionnaires are falling, generally.³ For this reason, and the potential to increase response rates by providing alternative methods of response (not to mention increasing study efficiency) it is important to consider media other than the paper questionnaire, such as telephone- or internet-based approaches.⁴ Recently, Breivik and colleagues using telephone interviews in 15 European countries, plus Israel, found that 19% of 46,394 respondents reported chronic pain, i.e. pain for at least 6 months, within the last month, and several times during the last week.⁵ In the United States, 9326 of 27,035 respondents to an online pain survey reported pain of at least 6 months' duration (30.7%).⁶ While both of these studies collected data on pain location, they used simple yes-no questions (e.g. 'Do you have pain in your lower back?'). It has been shown that determining pain location using this simple question approach can yield a different prevalence to when collected using a manikin-based approach⁷ and it is important, therefore, to adapt conventional pain manikins to electronic media.

Electronic body manikins have a number of other advantages over the traditional pencil-and-paper approach, for example: (1) in response to falling participation rates, online manikins provide participants with more than one response method; (2) after the initial set-up, they have considerable cost savings; (3) administrative tasks can be made more efficient and less error-prone with direct capture of data in electronic format; and (4) automation allows for graphical scoring techniques to eliminate inter-observer differences in scoring. To date, benefits in terms of utility have been reported elsewhere,⁸⁻¹³ but, other than in clinical applications with small samples and condition-specific groups, online manikins have not been evaluated for use as a population tool and the

agreement between online manikins and the traditional paper-based assessment is unknown. Thus, the aim of the current study was to determine agreement between online- and paper-based completion of body (pain) manikins.

Methods and materials

Participants and procedures

This study was conducted within a cohort study of pain in schoolchildren, the methods and results of which have been presented previously.¹⁴⁻¹⁶ In brief, 1446 children aged 11-14 years, from 39 schools in north-west England, took part in a survey of musculoskeletal pain. They were then followed up at 1 and 4 years to determine new-onset pain among those initially pain free and, among those with pain, whether symptoms were persistent.

At the 4-year follow-up, when the children were aged 15-18 years, data were collected by online questionnaire. However, because of uncertainties about the reliability of graphical questions (pain manikins) across different media, children who completed the online questionnaire also completed the key pain questions on paper. This study, therefore, focuses on the participants who completed *both* the online and paper questionnaire. Some participants completed the online questionnaire first, followed by the paper questionnaire; and others vice versa. The order was determined by practical considerations in the classroom - i.e. the number of computers available and the size of the class - but both questionnaires were completed within 1 hour.

Insofar as is possible, the online and paper questionnaires were identical in terms of question wording, font, format and order. In both questionnaires participants were asked the same stem question: 'Thinking back over the past month, have you had any pain that has lasted for a day or longer?'. Those answering positively were asked to identify the location of this pain on a body manikin. On online questionnaires, participants were asked to highlight painful body areas by clicking on predefined body areas in a four-view manikin. These areas corresponded to the 29 anatomical sites of the Manchester coding scheme (Figure 1).¹⁷ Participants could click on or off a body area, and the separate manikins were linked such that (for example) if the right knee was clicked on one manikin, the area where the right knee is indicated became highlighted on all three manikins. On paper questionnaires, those answering positively were asked to shade the location of their pain(s) on a four-view blank body manikin (Figure 2). The drawings were then coded by overlaying an acetate with the same 29 body areas demarcated, and pain was

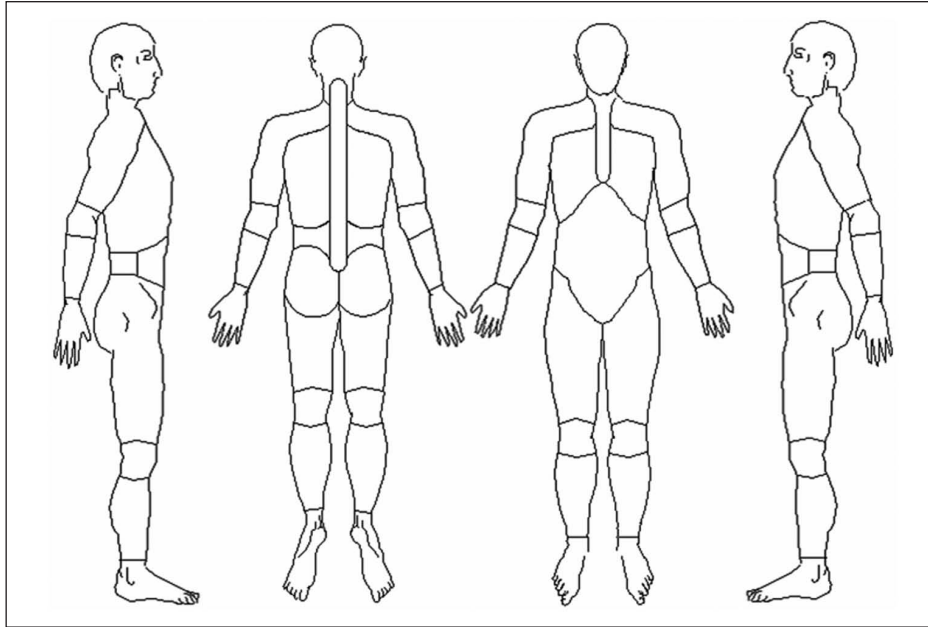


Figure 1. Predefined body manikin, from online questionnaire.

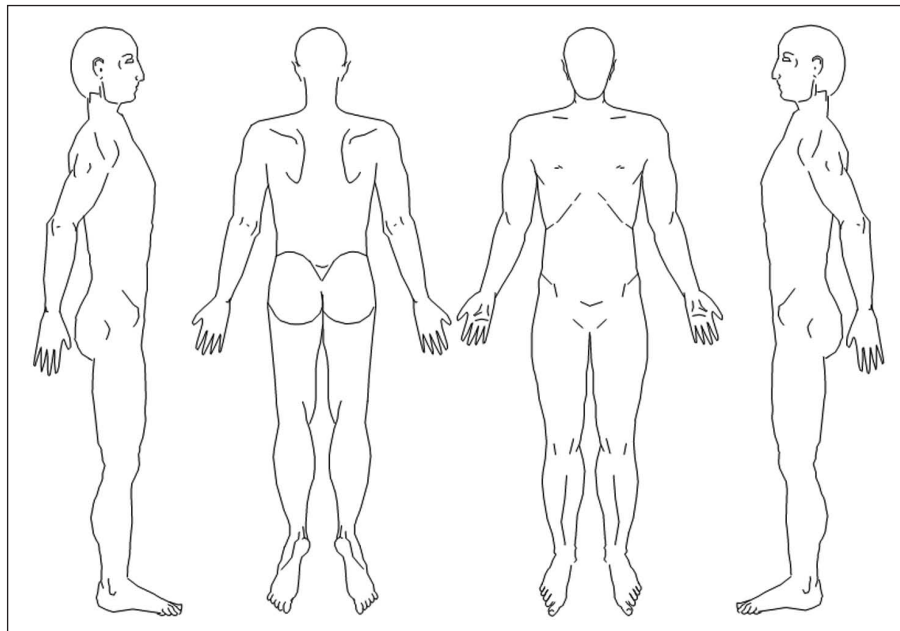


Figure 2. Blank body manikins, from paper questionnaire.

identified for each location if any shading was present within the predefined boundary. This latter method is identical to that used in the baseline and 1-year follow-up surveys.¹⁸

For both data capture approaches, patterns of pain report were then classified into 11 different regional syndromes (Table 1). In addition, widespread pain was coded as per the definition in the American College of Rheumatology (1990) classification criteria

for fibromyalgia, i.e. pain present both above and below the waist, on both the left and right sides of the body and in the axial skeleton.¹⁹

Ethics

The study was approved by the University of Manchester Senate Committee on the Ethics of Research on Human Beings.

Table 1. Prevalence of pain, using different methods of data collection.

	Prevalence (<i>n</i> = 264)		
	Paper questionnaire (%)	Online questionnaire (%)	Difference (%; 95% CI) ^a
Regional pain ^b			
Lower back	28.0	25.0	3.0 [−4.5 to 10.6]
Upper back	26.1	11.3	14.8 [8.2 to 21.3]
Shoulder	18.9	13.6	5.3 [−1.0 to 11.6]
Hip/thigh	18.2	7.6	10.6 [5.3 to 16.3]
Head	14.4	9.5	4.9 [−0.61 to 10.4]
Abdominal	14.0	8.3	5.7 [3.5 to 11.0]
Knee	12.5	11.4	1.1 [−4.4 to 6.7]
Ankle/foot	8.7	7.2	1.5 [−3.1 to 6.1]
Wrist/hand	3.0	4.5	−1.5 [−4.7 to 1.7]
Neck	2.7	3.8	−1.1 [−4.1 to 1.9]
Elbow/forearm	2.6	1.5	1.1 [−1.3 to 3.6]
Widespread pain	4.2	2.3	1.9 [−1.1 to 4.9]

^aPrevalence in paper questionnaire minus prevalence in online questionnaire. Therefore, a positive value represents a higher prevalence on paper.

^bListed in order of descending prevalence, as ascertained by paper questionnaire (gold standard).

Analysis

The difference in prevalence of regional and widespread pain conditions, as ascertained by paper and online questionnaires, was determined using simple descriptive statistics, and differences in pain extent – i.e. the total number of painful body sites per subject – was determined using the Wilcoxon sign rank test. Cohen's kappa (κ) coefficient was calculated to examine agreement between the two methods of data capture.²⁰ Because of the possibility of low prevalence of some of the pain conditions, prevalence and bias indices were calculated to determine the effects, on kappa, of prevalence and bias. Thereafter, prevalence- and bias-adjusted kappa (PABAK) was computed to adjust kappa accordingly.²¹ Kappa and PABAK coefficients were interpreted according to Landis and Koch's criteria: poor agreement ($\kappa < 0.20$); fair agreement ($0.20 < \kappa < 0.40$); moderate agreement ($0.41 < \kappa < 0.60$); good agreement ($0.61 < \kappa < 0.80$); and very good agreement ($\kappa > 0.80$).²² All statistical analyses were performed using STATA version 10.0 (StataCorp, College Station, TX, USA).

Results

Prevalence

A total of 264 children completed both the online and paper questionnaire (mean age 16.7 years [SD 0.9 years]; 48% female; 79% from state-funded schools). The prevalence of each of the different pain conditions is shown in Table 1, for each method of data collection. With paper questionnaires, prevalence ranged from

2.6% (elbow/forearm) to 28.0% (low back), and with the online questionnaires, prevalence ranged from 1.5% to 25.0%, for the same two pain conditions. For the majority of conditions, prevalence was higher on paper than online, the only two exceptions being neck pain (2.7% and 3.8%) and wrist/hand pain (3.0% and 4.5%) for paper and online questionnaires, respectively. In general, the difference in prevalence between paper and online questionnaires was fairly small (Table 1), the notable exceptions being upper back pain (difference: 14.8%; 95% CI: 8.2 to 21.3%) and hip/thigh pain (difference: 10.6%; 95% CI: 5.3 to 16.3%).

Agreement

One hundred and eighteen participants answered positively to the pain stem question on both the paper and online questionnaire ("Thinking back over the past month, have you had any pain that has lasted for a day or longer?"). Agreement of pain location, as per the different manikins, was therefore assessed in this subgroup. The measure of agreement in the identification of different pain locations between the two data collection methods is shown in Table 2. Cohen's kappa for different pain locations demonstrated wide variation from poor agreement (neck pain and elbow/forearm pain) to good agreement (lower back pain and wrist/hand pain).

The prevalence indices – representing the difference in the probability of pain–yes and pain–no, for each location – ranged from 0.03 (for lower back pain, with the highest prevalence) to 0.91 (for elbow/forearm pain,

Table 2. Agreement of pain location, between different methods of pain location.

	Measure of agreement (<i>n</i> = 118)				Rating ^a
	Kappa	PI	BI	PABAK	
Lower back	0.66	0.03	0.08	0.66	Good
Upper back	0.31	0.24	0.31	0.29	Fair
Shoulder	0.59	0.33	0.14	0.63	Good
Hip/thigh	0.35	0.51	0.20	0.49	Moderate
Head	0.50	0.61	0.13	0.76	Good
Abdominal	0.42	0.58	0.13	0.61	Good
Knee	0.52	0.54	0.03	0.66	Good
Ankle/foot	0.79	0.66	0.01	0.88	Very good
Wrist/hand	0.60	0.83	0.03	0.88	Very good
Neck	0.10	0.89	0.03	0.81	Very good
Elbow/forearm	0.15	0.91	0.03	0.84	Very good
Widespread pain	0.47	0.86	0.05	0.86	Very good

PABAK = prevalence- and bias-adjusted kappa; BI = bias index; PI = prevalence index.²¹

^aMeasure of agreement, as per Landis and Koch's²² criteria.

with the lowest prevalence). This suggested – for areas of low prevalence in particular – that interpreting kappa values on their own may be misleading and that the measure of agreement may be artificially low. For most pain locations, PABAK (the prevalence- and bias-adjusted kappa) demonstrated good or very good agreement (PABAK ranging from 0.61 to 0.88). However, identification of pain in the hip/thigh and in the upper back had poorer agreement (PABAK: 0.49 and 0.29, respectively).

Discussion

Improvements in technology, and decreasing questionnaire response rates in epidemiological studies, have led to consideration of the use of online questionnaires. While, conceptually, simple multi-response questions or questions requiring free-text responses may transfer easily to different media, graphical questions such as pain manikins are untested in terms of their reliability. We have shown, among a sample of schoolchildren aged 15–18 years, generally good or very good agreement between the information gained by online and paper manikins.

There are a number of methodological issues to consider in the interpretation of these data. First, the two body manikins – a blank body manikin (paper) versus a manikin with pre-defined body areas (online) (Figures 1 and 2, respectively) – differed with respect to their presentation. Therefore, any disagreement between methods may be due to the presentation of the manikins and not solely to the mode of administration. The aim of the study was to determine the agreement between an online data collection method and the gold-standard blank body manikin of the paper

questionnaire. A better comparison may have been between a blank paper manikin and a blank online manikin where both instruments require shading with a pen, or equivalent. With recent advances in technology there has been a proliferation of stylus-input, or fingertip-input, tablet computers such as Apple's iPad®. However, these are not available to the majority, and the researcher cannot rely on the widespread use of such technology for population-based surveys. Thus, one has to design online questionnaires for internet browser technology that is commonly available to the general population. The question, therefore, is not how can the most advanced computers accurately recreate the pen-and-paper experience, but how well can the data traditionally collected by pen and paper be collected on the majority of computers. However, had it been possible to collect online manikin data using a shading approach, we would argue that the measures of agreement are likely to have been even greater.

An intermediate approach between simple yes–no questions about the presence of pain, and the blank body manikin, is the preshaded manikin (Figure 3). This allows visual representation of the anatomical area of interest, as defined by the investigator, rather than allowing the participant to determine what is meant by (for example) 'the low back'. In a postal survey conducted by Lacey and colleagues,²³ blank and preshaded body manikins elicited a 6.7% difference in prevalence of neck and upper body pain; in the case of the latter, the authors have suggested that higher estimates for the preshaded manikin may be the result of directly asking questions about pain in the preshaded area (i.e. prompting more respondents with less severe pain to report symptoms). In the present study, however, the restrictions of the prespecified areas of the

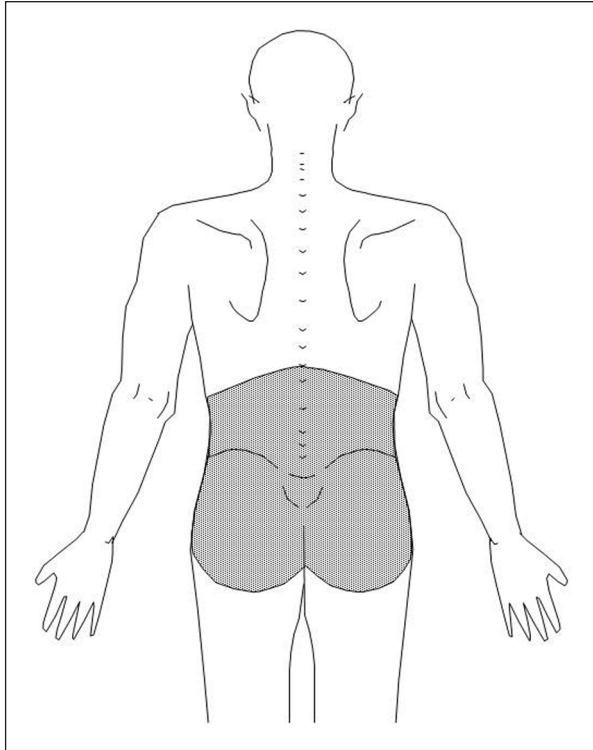


Figure 3. Preshaded body manikin, denoting low back pain.

online manikins may have led to under-reporting of pain, manifested in two ways. First, when clicking on an area the whole area changed colour. This may have seemed too crude a representation of the actual site of pain and it may have subsequently been clicked ‘off’ again. Second, it may be that areas identified on the paper manikins may be overshading from adjacent areas. Bryner first identified the drawing task itself as a factor that can affect the accuracy of pain drawings.⁹ In the current study, when scoring paper manikins, any shading, in any area, was positively coded for pain. This strict criterion has established merit as a widely accepted operational scoring system, but does not control for misclassification due to poor drawing skill. The current study provides evidence that overshading on the standard paper manikins is at least partly responsible for overestimation of some pain syndromes, compared with the online manikins. Upper back pain, hip pain and abdominal pain were the three body areas with both (1) significant differences in prevalence between paper and online manikins and (2) the lowest PABAK measure of agreement (0.29, 0.49 and 0.61, respectively). It may be that pain in these areas is not as high as identified on the paper manikin; rather, these areas are all contiguous to the lower back, and it may be that the high prevalence ascertained by the paper questionnaire is a function of overshading from the lower back area, which has the highest prevalence of

any regional pain site. Alternatively, it may be that the ‘low back’ area on the online manikin does not adequately capture the anatomical location of lower back pain in all patients.

It may be possible to increase the measure of agreement between online and paper questionnaires by further refining the pain areas on the online manikin – by making them smaller, for example – so that participants can indicate the location of their pain, online, with greater resolution. However, it is important to balance the need for accuracy in the identification of detailed differences among many small areas, the ease of use of the instrument, and the purpose for which it is intended. Von Baeyer and colleagues²⁴ illustrate this by example. They argue that studies in rheumatology may require pain data to be collected for each joint, requiring 80 or more separate variables to express pain location, whereas researchers studying pain (irrespective of location) and coping, may require fewer, larger areas.

The choice of appropriate time interval between data collection one and data collection two is a problem for all reliability studies. This interval needs to be sufficiently long, such that memory of the first instrument does not influence the responses to the second instrument, but it also needs to be sufficiently short, such that the underlying trait under investigation does not vary between measurements. In the current study, the online and paper manikins were completed with 1 hour of each other when pain status (pain in the past month) will not have changed. Owing to the practicalities of conducting this research in the classroom (and the limited time available with access to the pupils) it was not possible to randomise participants to complete the online or paper manikin first, as would have been desirable. It is possible, therefore, that completion of one instrument may have influenced the subsequent completion of the other. However, it is not possible to examine this using the data available. Once participants had completed the online manikins and continued to the next part of the online questionnaire, they were unable to go back to review their responses. Participants were therefore unable to check the agreement between their online and paper manikins themselves. Also, in the time between the completion of both, their attention was focused away from pain: there were several sections of the online questionnaire that followed the manikins – about different aspects of health and lifestyle – before participants left the computer terminal and completed the paper questionnaire. Thus, while we are unable to rule out an order effect, we consider that the impact of such an effect in the current study would have been small.

We have shown, first, that for the majority of regional pain syndromes, and widespread pain, there are no

significant differences in prevalence estimates between the two data collection methods and, second, that among those with pain, the location of pain can be captured with online methods with generally good/very good agreement compared with the traditional pen-and-paper blank manikin approach. There are still some concerns over agreement and reproducibility of pain location, with respect to the body areas adjacent to areas of high prevalence (such as the lower back), although we provide some evidence to support the theory that this may reflect shading inaccuracy and over-reporting on paper, rather than under-reporting online. In addition, the current study was conducted in children and there may be different issues of acceptability of online questionnaires in the population generally, particularly with older adults. However, these issues related more to the acceptability of online questionnaires in general, whereas it is harder to argue that the agreement between online and paper manikins would differ greatly between adolescents and the adult population.

Paper-based blank body manikins are characterised as having good reliability and sensitivity to detect the distribution and extent of musculoskeletal pains across different populations. Efforts to develop and validate the online manikin, in addition to a battery of other commonly used pain measures, have been seen as a key factor to advancing the logical integration of information and communication technology in epidemiological studies. The current study is the first of which we are aware that examines the reproducibility of body manikin data collected on different media, from the same participants, in a population-based survey. Our results need to be replicated, using a larger sample, and incorporating a random element into the order of questionnaire completion.

In summary, we would always recommend that the same methodological approach be used in studies with repeated observations (e.g. one should not use paper manikins for baseline and online manikins for follow-up) and for all study participants (e.g. one should not use paper manikins for cases and online manikins for controls). However, we have shown that online manikins can be used to capture data on pain location in a manner satisfactorily comparable to paper questionnaires.

Acknowledgements

The authors thank the 39 schools and the children who participated in the study. In addition, we are grateful to Mark Blything and Stephen Conn for assistance with data collection.

Conflict of interest

The author declares that there is no conflict of interest.

Funding

This study was funded by the Colt Foundation UK (www.coltfoundation.org.uk; registered charity 277189) with additional support from Arthritis Research UK (www.arthritisresearchuk.org; registered charity 207711).

References

1. Margolis RB, Tait RC and Krause SJ. A rating system for use with patient pain drawings. *Pain* 1986; 24(1):57–65.
2. Ransford OA, Cairns D and Mooney V. The pain drawing as an aid to the psychologic evaluation of patients with low back pain. *Spine* 1976; 1(2): 127–134.
3. Morton LM, Cahill J and Hartge P. Reporting participation in epidemiologic studies: a survey of practice. *Am J Epidemiol* 2006; 163(3): 197–203.
4. van Gelder MM, Bretveld RW and Roeleveld N. Web-based questionnaires: the future in epidemiology? *Am J Epidemiol* 2010; 172(11): 1292–1298.
5. Breivik H, Collett B, Ventafridda V, et al. Survey of chronic pain in Europe: Prevalence, impact on daily life, and treatment. *Eur J Pain* 2006; 10(4): 287–333.
6. Johannes CB, Le TK, Zhou X, et al. The prevalence of chronic pain in United States adults: results of an internet-based survey. *J Pain* 2010; 11(11): 1230–1239.
7. Pope DP, Croft PR, Pritchard CM, et al. Prevalence of shoulder pain in the community: the influence of case definition. *Ann Rheum Dis* 1997; 56(5): 308–312.
8. North RB, Nigrin DJ, Fowler KR, et al. Automated 'pain drawing' analysis by computer-controlled, patient-interactive neurological stimulation system. *Pain* 1992; 50(1): 51–57.
9. Bryner P. Extent measurement in localised low-back pain: a comparison of four methods. *Pain* 1994; 59(2): 281–285.
10. Jamison RN, Fanciullo GJ and Baird JC. Computerized dynamic assessment of pain: comparison of chronic pain patients and healthy controls. *Pain Med* 2004; 5(2): 168–177.
11. Sanders NW and Mann NH 3rd. Automated scoring of patient pain drawings using artificial neural networks: efforts toward a low back pain triage application. *Comput Biol Med* 2000; 30(5): 287–298.
12. Wenngren A and Stalnacke BM. Computerized assessment of pain drawing area: a pilot study. *Neuropsychiatr Dis Treat* 2009; 5: 451–456.
13. Felix ER, Galoian KA, Aarons C, et al. Utility of quantitative computerized pain drawings in a sample of spinal stenosis patients. *Pain Med* 2010; 11(3): 382–389.
14. Watson KD, Papageorgiou AC, Jones GT, et al. Low back pain in schoolchildren: occurrence and characteristics. *Pain* 2002; 97(1–2): 87–92.
15. Jones GT, Watson KD, Silman AJ, et al. Predictors of low back pain in British schoolchildren: a population-based prospective cohort study. *Pediatrics* 2003; 111(4 Pt 1): 822–828.
16. Jones GT and Macfarlane GJ. Predicting persistent low back pain in schoolchildren: a prospective cohort study. *Arthritis Rheum* 2009; 61(10): 1359–1366.

17. Hunt IM, Silman AJ, Benjamin S, et al. The prevalence and associated features of chronic widespread pain in the community using the 'Manchester' definition of chronic widespread pain. *Rheumatology (Oxford)* 1999; 38(3): 275–279.
18. Jones GT, Silman AJ and Macfarlane GJ. Predicting the onset of widespread body pain among children. *Arthritis Rheum* 2003; 48(9): 2615–2621.
19. Wolfe F, Smythe HA, Yunus MB, et al. The American College of Rheumatology 1990 Criteria for the Classification of fibromyalgia. *Arthritis Rheum* 1990; 33(2): 160–172.
20. Cohen JA. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960; 20: 37–46.
21. Byrt T, Bishop J and Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol* 1993; 46(5): 423–429.
22. Landis JR and Koch GG. A review of statistical methods in the analysis of data arising from observer reliability studies (Part I). *Statistica Neerlandica* 1975; 29: 101–123.
23. Lacey RJ, Lewis M and Sim J. Presentation of pain drawings in questionnaire surveys: influence on prevalence of neck and upper limb pain in the community. *Pain* 2003; 105(1–2): 293–301.
24. von Baeyer CL, Lin V, Seidman LC, Tsao JC and Zeltzer LK. Pain charts (body maps or manikins) in assessment of the location of pediatric pain. *Pain Manag* 2011; 1(1): 61–68.