Validation and utilisation of a digital version of the Survey instrument for Natural history, Aetiology and Prevalence of Patellofemoral pain Studies (eSNAPPS)

Mohamed Yusuf,[1] Paola Dey,[2] Michael Callaghan,[3] Nicola Relph,[2] Phillip Gichuru,[4] Susan Pinner,[1] Jane Ashbrook,[3] Joanne Ashman[3], Peter Goodwin[3], Ruth Macdonald[3], James Selfe[3]

[1] Musculoskeletal Science and Sports Medicine Research Centre, Department of Sport and Exercise Sciences, Manchester Metropolitan University, Manchester.

[2] Faculty of Health, Social Care and Medicine, Edge Hill University, Ormskirk, UK.

[3] Department of Health Professions, Faculty of Health, Psychology and Social Care, Manchester Metropolitan University, Manchester, UK.

[4]Aston Medical School, Aston University, Birmingham, UK

ABSTRACT

Objective: This study validated the newly adapted electronic SNAPPS (eSNAPPS) against the original paper SNAPPS. Subsequently, the study estimated the prevalence of PFP in running participants and spectators attending three mass-participant running events in the United Kingdom by using the eSNAPPS tool.

Methods: This study had two parts. Firstly, a validation of the original paper version of the SNAPPS tool. Secondly, if validation was achieved, eSNAPPS was used in a prevalence study. A convenience sample of running participants and spectators aged 18-40 years attending the mass participation running events was used. 12-month prevalence of PFP was used as the main outcome.

Results: eSNAPPS was valid in identifying those with PFP (ICC 0.99 for Overall agreement, $p<0.0001$). In the prevalence study, a total of 1,080 running participants and spectators completed the eSNAPPS. The overall prevalence of PFP was 17.4% (95%CI: 15.2%, 19.8%); 20.5% of males (16.5, 24.9) and 15.7% of females (13.1, 18.7) had PFP. Prevalence was 17.4% (15.2, 19.8) in spectators and 16.7% in running participants (14.5, 19.0).

Conclusion: The overall PFP prevalence in this study was slightly smaller than those previously reported in the literature. Findings also show that there were similar prevalence estimates in spectators and running participants.

Keywords: Patellofemoral Pain, Knee pain, Epidemiology, Prevalence

Background

Patellofemoral pain (PFP) typically presents as diffuse anterior knee pain, usually with activities such as squatting, running, stair ascent and descent (Crossley, et al., 2016). PFP poses a significant clinical, financial and healthcare burden on the individual and health services (Peat, McCarney, & Croft, 2001; Urwin, et al., 1998; Webb, et al., 2004), and is a potential 'barrier' to physical activity (Doménech, Sanchis-Alfonso, & Espejo, 2014; Smith, Moffatt, et al., 2018). The latter has significant public health implications as lack of physical activity increases the risk of non-communicable diseases such as diabetes and cardiovascular disease (Allender, Foster, Scarborough, & Rayner, 2007; Lee, et al., 2012; Sari, 2009).

Large-scale data on the prevalence and incidence of PFP in general populations are scarce (Callaghan & Selfe, 2007; Smith, Selfe, et al., 2018). Typically, PFP has been measured in homogenous populations, such as adolescents, military personnel and athletes (Neal, et al., 2019; Smith, Selfe, et al., 2018). One reason for the lack of prevalence or incidence data for PFP in larger and community populations is the paucity of measuring instruments. To overcome this challenge, Dey et al. (2016) developed a discriminatory self-report, paper-based questionnaire with high sensitivity, specificity and test-retest reliability scores, that identified those with PFP in the adult general population (SNAPPS - Survey instrument for Natural history, Aetiology and Prevalence of Patellofemoral pain Studies). Using SNAPPS, Dey et al. (2016) reported the 12-month prevalence of PFP to be 22.7% in the general population. Using the same instrument, two additional studies, reported the 12-month prevalence of PFP to be 45.3% in 203 amateur runners in Nigeria (Akodu & Nwakalor, 2018) and 20.7% in 1,153 individuals from the general population in China (Xu, et al., 2018).

Large scale studies can be facilitated by using questionnaires in a digital format. Digital surveys have several advantages over paper surveys, which make them a preferable platform for gathering population-based data in epidemiological studies (Ekman & Litton, 2007; Van Gelder, Bretveld, & Roeleveld, 2010; Wright, 2005), such as rapid completion after survey distribution (Akl, Maroun, Klocke, Montori, & Schünemann, 2005; Kroth, et al., 2009). When collecting from a large population cohort or collecting large volumes of data, digital surveys are cost-effective in terms of printing, distribution, data scoring and data entry (Ebert, Huibers, Christensen, & Christensen, 2018). Additionally, due to their potential for automation, digital surveys have the capability of eliminating or minimising the risk of error and bias in the measurement and transcription of responses (Clark & Maynard, 1998). However, it should not be assumed that instruments in one format are valid when transferred to other formats, despite the content of the questionnaire not changing. In part, this is likely due to a change in the way the participants interact with the instrument (Cronbach & Meehl, 1955). As such, to ensure that the psychometric properties do not change, it is vital to ensure that the adapted instrument is tested for validity against its reference.

In this study, we first aimed to determine the validity of an electronic version of SNAPPS (eSNAPPS) by measuring its agreement with the original paper SNAPPS. Secondly, if validation was demonstrated or confirmed, we aimed to utilise the eSNAPPS in a large population-based study by estimating the prevalence of PFP in participants and spectators attending mass-participation running events in the UK.

METHODS

Development of eSNAPPS

The paper SNAPPS tool has four sections; it asks questions on clinical features, pain on activity and pain on location, is in multiple-choice format and has embedded skip-logic (Dey, et al., 2016). The first three questions record demographic characteristics including age and sex, subsequent questions are designed to identify those with knee pain (section 1); e.g., 'have you had pain or problems in the last year in or around the knee?' Section 2 asks about clinical features related to knee pain. Section 3 determines whether participants experience pain or difficulty in various activities; this question has 14 'sub' questions, (this section is not used in scoring). Section 4 asks participants to report their location of pain using an image of the knee (Elson, et al., 2011). Section 3 is not used in the scoring as its' inclusion has previously been shown to make no difference in identifying between those with PFP and those with healthy knees (Dey, et al., 2016), and that a combination of section 2 and section 4 was better in identifying between the two groups. To determine the presence of PFP, sections two and four are scored. For section two, the maximum score for each question is 1 and the minimum is 0; the maximum total score is 7. For section four each area (medial patella, lateral patella and patella tendon) is given a maximum score of one. Areas that are outside of these sections are given a score of 0. Each knee has a maximum score of 3 and hence the maximum available score is 6 and the minimum available score is 0. The max available score for the survey is 10. Supplementary Table 1 contains the SNAPPS questions and their scoring.

The eSNAPPS was adapted from the original paper SNAPPS using Online Surveys (www.onlinesurveys.ac.uk). Questions in the eSNAPPS were set-up in the same order as the paper version of SNAPPS. For eSNAPPS, only the knee pain map was modified from the original paper version. Within the paper SNAPPS, this question requires the marking of a cross on a knee image to highlight the location of the pain. However, for digital SNAPPS, this was adapted as the online survey platform does not have an interactive feature allowing marking of the knee image to show pain location. The measurement properties of the knee pain map have been previously determined and validated in the paper SNAPPS (Dey, et al., 2016), therefore, in the digital version, only the interactive element of the question had changed. In line with the scoring of the paper version, a reference scoring grid was devised, this segmented the knee into four sections; these were medial patella, lateral patella patella tendon, and outside of patella. Participants selected a number or multiple numbers that best corresponded to the location of their knee pain. In contrast, in the paper SNAPPS, participants freely marked the location of pain with a cross using a pen.

Study population and data collection

The validation study complied with the Declaration of Helsinki and was approved by Faculty ethics committee Manchester Metropolitan University (EthOS ref: 839) and Edge Hill University (ref: HW 18). Using snowball sampling, participants with and without self-reported current knee pain were recruited from a University, sports therapy clinics and sports clubs in North West England. Participants were randomly assigned to either complete the paper SNAPPS or the eSNAPPS first. They completed the alternative format a week later. Questionnaires were scored as per the scoring algorithm proposed by Dey et al (2016).

The prevalence study complied with the Declaration of Helsinki and was also approved by Faculty ethics committee at Manchester Metropolitan University (EthOS Ref: 2223). This study took place in the North West of England and recruited spectators and running participants aged 18 to 40 years attending three mass-participation running events 'Race for Life' during (17th May 15th and 16th July 2019), these events were organised by Cancer Research UK and are designed to raise awareness and funds for cancer research. Running participants was defined as individuals who were participating in the 'Race for Life' events, while spectators were defined as individuals who were not taking part in any of these events. Using convenience sampling, running participants and spectators were approached by researchers and were asked if they would be willing to take part in the study. Those who agreed were provided with written information about the study, and gave informed consent using the first pages of the electronic questionnaire through tick box methods. The main study outcome was 12 month prevalence of PFP, this was determined by the scoring of eSNAPPS (Dey, et al., 2016). Dey et al. (2016) found an overall score of 6 and above to have highest sensitivity and specificity (sensitivity of 92% and a specificity of 94%) to indicate PFP.

Data analysis

To score the eSNAPPS a series of commands that can be executed at one time was created in R program (Team, 2013), this automated the scoring process without any errors. To ensure consistency in the code, a portion of automated scores for the eSNAPPS responses (10%) were cross-checked against scores that were generated through manual scoring. Responses from the paper SNAPPS were manually entered in an Excel spreadsheet. Manual entry of responses was cross-checked twice to ensure there were no data entry errors. Scores for paper SNAPPS were manually scored by hand.

For the validation of eSNAPPS, to compute the minimum required sample size, a 55% correlation between paper and eSNAPPS was deemed acceptable. Therefore, using a two-sided α of 0.05 and a power of 90%, this recommended a minimum sample size of 30 (Liao, 2010). In addition, to evaluate the agreement of eSNAPPS and paper SNAPPS, an Intraclass Correlation Coefficient (ICC 2, 1) with a single measurement, absolute agreement, two-way random effects model was used to correlate the survey scores between the survey instruments (Weir, 2005). For interpretation, the ICC classification proposed by Shrout and Fleiss was used: 0.00-0.40 = poor, 0.40–0.75 = modest, 0.75–1.00 = excellent (Shrout & Fleiss, 1979).

In the prevalence study, data were segregated into three groups i) PFP, ii) Other knee problems, iii) Healthy (Figure 1). Participants with scores greater than or equal to 6 were categorised as PFP. Those who had knee pain in the past year did not achieve a score of 6 or above were classified as the 'other knee problems' group, and those who did not have knee pain within the past year were grouped as 'healthy'. Prevalence of PFP was estimated as the number of individuals with PFP divided by the total number of individuals who were surveyed. The 12-month prevalence rate of PFP for all respondents, for running participants and spectators were estimated. Whether the prevalence of PFP in running participants and spectators differed based on age and sex was also explored. As the distribution of age and sex differed between running participants and spectators the prevalence estimates for spectators and running participants were directly age-sex standardised using the overall sample as the reference population to facilitate comparison. For the overall prevalence rate, the difference in prevalence rate and the age-standardised prevalence rates, 95% confidence intervals were calculated. Analysis of data was conducted using R program (Team, 2013).

RESULTS

Validation study

In the validation, 33 participants were recruited. The eSNAPPS had no missing responses. However, the paper SNAPPS had six missing responses from six different items (three individuals each missing two questions). However, the missing items did not impact survey scores. In the analysis of overall agreement between paper SNAPPS and eSNAPPS, there was an excellent agreement for all participants (ICC:0.99 95% CI 0.99-1.00, p<0.0001) and for knee pain (ICC: 0.94 95% CI 0.80-0.98, p<0.0001). For the two sections that were scored, the agreement was excellent; section 2 for all participants, (ICC: 0.99 95% CI 0.99-1.00, p<0.0001) and for section 4 for all participants, (ICC: 0.93 95% CI 0.87-0.97, p<0.0001) – see Table 1.

Table 1: Overall score and section agreement of digital versus paper SNAPPS

	All participants		Knee pain	
	ICC (2, 1) [a] (95% CI [b])	Classification	ICC (2, 1) (95% CI)	Classification
Overall Score	0.99 (0.99 – 1.00)	Excellent	0.94 (0.80 – 0.98)	Excellent
Section 2: Clinical features	0.99 (0.99 – 1.00)	Excellent	0.96 (0.90 – 0.98)	Excellent
Section 4: Knee pain map	0.93 (0.87 – 0.97)	Excellent	0.92 (0.81 – 0.97)	Excellent

[a] Interclass correlation, [b] 95% Confidence Interval

Prevalence study

In total 1,190 participants took part. Of these, 9 entries were incomplete and 101 were screened out by the tool because they were outside the eligible age range; 21 individuals were <18 and 80 individuals were >40 years old. Therefore 1,080 complete responses were available for analysis (Figure 1), with more running participants (n=712, 66%) than spectators (n=368, 34%) – see Table 2. The study sample included more females (n=699, 65%) than males (n=381, 35%); the overall mean age was 28 years (± 6 years). In total, 303 (28.1%, 95% CI 25.4–30.4) participants reported having knee pain in the past year: of whom 188 (17.4%) individuals were classified as having PFP and 115 individuals were classified as having other knee problems (10.7%). This left 777 (71.9%) participants who had no knee problems and were classified as 'healthy' (71.9%). The 12-month PFP prevalence rate was 17.4% (95% CI 15.2%-19.8%). Of those with PFP, 46.8% (n=88/188) had previously consulted a doctor about their knee problems compared to 53.9% of those with other knee problems (62/115).

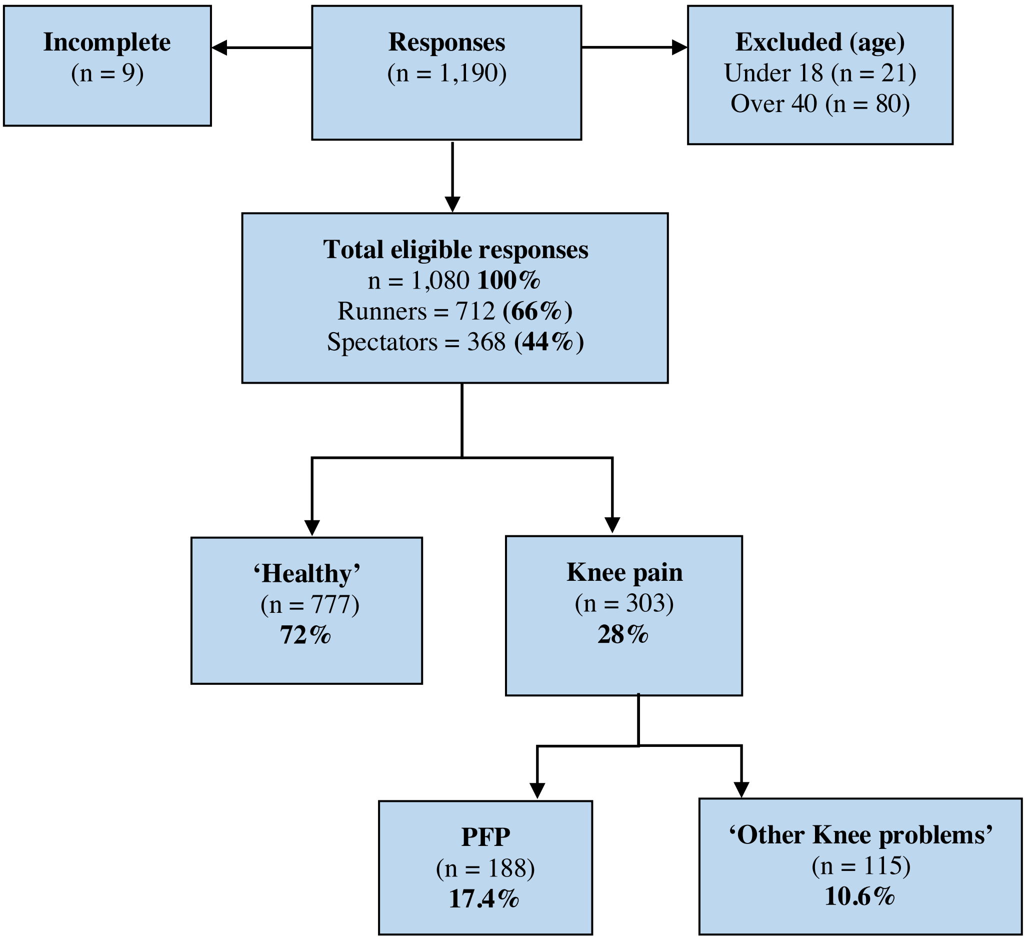

Figure 1: Flow of study participants

Table 2: Demographic characteristics of the study population (N= 1,080)

	n	% (Mean, SD [a])
Sex		
Female	699	64.7
Male	381	35.3
Mean age (in years)	1080	28.4 (± 6.1)
Age group (years)		
18 – 20	122	11.3
21 – 25	272	25.2
26 – 30	288	26.7
31 – 35	224	20.7
36 – 40	174	16.1
Participation		
Spectator	368	34.1
Running participants	712	65.9
Knee pain		
PFP [b]	188	17.4
'Other knee problems'	115	10.7
'Healthy'	777	71.9

[a] Standard Deviation, [b] Patellofemoral pain, [c] 95% Confidence Interval

Table 3 presents the prevalence of PFP by demographic characteristics. There were more males (n=78/381; 20.5%, 95% CI 16.5%-24.9%) with PFP than females (n=110/699, 15.7%, 95% CI 13.1%-18.7%). Within the age strata, the proportion of individuals with PFP appears to increase with age. Those aged between 31 and 35 years experienced PFP the most (n=49/224, 21.9%, 95% CI 16.6%-27.9%). Prevalence of PFP was 19.3% in spectators (95% CI 15.4%-23.7%) and 16.4% in running participants (95% CI 13.8%-19.4%).

Running participants were more likely to be female (n=561/712; 78.8%, 95% CI 75.6%-81.7%) than spectators who were more like to be males (n=230/368; 62.5% 95% CI 57.3%-67.5%) and there were no differences in the age distribution between running participants and spectators (spectators: mean 28.5 ± SD 6.1 years, running participants: mean 28.3 ± SD 6.1 years). Prevalence of PFP increased with age and was higher in males compared to females; following age-sex standardisation the rates were 17.4% (95% CI 15.2%-19.8%) for spectators and 16.7% (95% CI 14.5%-19.0%) for running participants – see Supplementary Table 4. When standardised for age, 18.1% (95% CI 14.1%-22.4%) of male running participants had PFP, and for females, it was 15.7% (95% CI 13.1%-18.7%). Similarly, 21.0% (95% CI 17.0%-25.4%) of male spectators had PFP, while 15.5% (95% CI 12.9%-18.4%) of female spectators also had PFP.

Table 3: Prevalence of PFP by demographics

	n	%	95% CI[a]
Sex			
Female	110 / 699	15.7	13.1 – 18.7
Male	78 / 381	20.5	16.5 – 24.9
Age (years)			
18 – 20	16 / 122	13.1	7.7 – 20.4
21 – 25	43 / 272	15.8	11.7 – 20.7
26 – 30	42 / 288	14.6	10.7 – 19.2
31 – 35	49 / 224	21.9	16.6 – 27.9
36 – 40	38 / 174	21.8	15.9 – 28.7
Participation			
Spectator	71 / 368	19.3	15.4 – 23.7
Runner	117 / 712	16.4	13.8 – 19.4

[a] 95% Confidence Interval

DISCUSSION

We have demonstrated that the eSNAPPS tool is valid and can be used in large scale studies to determine the prevalence of PFP. Data can be collected easily and quickly in a technically difficult environment, and electronic storage ensures no loss of data. Consequently, we were able to use the eSNAPPS to collect data from a large sample of runners and spectators attending mass-participant running events and estimated 12-month PFP prevalence. In the current study, 28.1% had self-reported general knee pain, in line with a cross-sectional study using paper SNAPPS to determine a prevalence of 31% knee pain in a UK university cohort (Ibeachu, Selfe, Sutton, & Dey, 2019). This pattern has also been demonstrated in other prevalence studies using the SNAPPS questionnaire, whereby a third of those surveyed had knee pain (Dey, et al., 2016; Xu, et al., 2018).

In the current study, 17.4% of those surveyed presented with PFP, which is lower than that reported elsewhere in the literature (Callaghan & Selfe, 2007; Smith, Selfe, et al., 2018). These results could be due to differences in sampling methodology. Unlike sampling methods where respondents are approached by data collectors, digitally delivered surveys can introduce self-selection bias. This might lead to those with PFP being more inclined to participate in the survey than those without the condition of interest, thus leading to an over representation of participants with PFP. Other differences the explain the lower prevalence might involve the measuring instruments and study populations, as other studies sampled adolescents, military and athletes with varying case definitions and use differing clinical measures and self-report outcomes to determine the presence of PFP. Moreover, when compared to the three other studies using the paper SNAPPS questionnaire, the current study's prevalence estimate is also lower; 21.7% in the general population of China (Xu, et al., 2018), 45.3% in an athletic population in Nigeria (Akodu & Nwakalor, 2018), and in the original SNAPPS study (Dey, et al., 2016), 22.7% in a mixed population of university students and members of the general public in UK (Dey, et al., 2016).

Within the current study, 64.7% of the sample population were female, but more males presented with PFP. This is in contrast with previous studies regarding the prevalence of PFP which suggest that women often have a higher PFP prevalence than men (Boling, et al., 2010; M. S. Rathleff, et al., 2015), however, data from these studies were on different populations - adolescents, athletes and military personnel; therefore, this contrast could be due to the inherent difference between the populations in these studies. Further, the prevalence of PFP increased with age. Looking at the annual consultation prevalence of PFP, Wood et al. (Wood, Muller, & Peat, 2011) showed that adolescents and young adults (aged 15 - 29) have a higher prevalence of patellofemoral disorders than older adults (aged 30 – 44). However, routinely collected general practitioners' data, and the population data within the current study may not be comparable as this is a clinical population that may also differ from community-based populations where PFP prevalence is likely to be higher.

A recent systematic review on running injuries has shown the knee to be the most commonly injured location (28% of all injuries) and PFP to be the most commonly occurring running-induced injury (17% of all injuries) (Francis, Whatman, Sheerin, Hume, & Johnson, 2019). Besides, when comparing PFP prevalence between running participants and spectators, our results suggest that there was no difference between these different respondents. These findings are inconsistent with the existing literature, as PFP has previously been thought to affect mainly athletes (Akodu & Nwakalor, 2018; Smith, Selfe, et al., 2018; Taunton, et al., 2002). However, these previous studies reported PFP prevalence estimates in athletes rather than non-athletes. Our results suggest that, PFP is a condition of the general population as well as of athletes. Furthermore, in the current study, some of the spectators were themselves runners but were unable to participate in the events at the time due to an injury this could also potentially explain the similarity in the prevalence estimates of running participants and spectators. In the future, more information could be gathered from study participants regarding their running history and why they were not participating in the running events.

We also found that 47% of those with PFP and 54% of those with other knee problems had previously visited a doctor about their knee problem. A health care seeking pattern differed little between running participants and spectators (21% vs 24%, respectively). These findings highlight that PFP may have a burden on primary and secondary musculoskeletal care services similar to other general knee pain disorders, such that they lead to significant health care seeking behaviour in individuals presenting with these conditions. The similarity in health care seeking behaviour of those with PFP and those with other knee problems further highlights the need for a full assessment of the economic and healthcare system burden associated with PFP. The estimates for health seeking behaviour are likely to be higher than observed, as the eSNAPPS tool specifically asks about visiting the doctors. It is possible that some participants may have also seen physiotherapists, sports therapists or other health care clinicians rather than a doctor, thus leading to potential under-reporting of the burden of PFP.

Similar results can be observed in a population-based study evaluating the health care seeking behaviour of adolescents with knee pain (Rathleff et al., 2013). Rathleff et al (2013). reported those with knee pain of insidious, atraumatic onset had similar outcomes as traumatic knee pain for pain severity, pain duration and reductions in health-related quality of life. In the current study, similarities between PFP and 'other knee problems' (which included traumatic knee pain) further highlighted a potentially unmet burden of PFP in the population. Currently, there appears to be a large mismatch with research funding and priorities for PFP when compared to other knee problems (Smith, Selfe, et al., 2018). For example, current literature is dominated by research on anterior cruciate ligament and knee osteoarthritis while there is limited research around PFP, although, PFP has been proposed to be part of the patellofemoral osteoarthritis continuum (Crossley, 2014; Eijkenboom, Waarsing, Oei, Bierma-Zeinstra, & van Middelkoop, 2018; Mehl, et al., 2019; Thomas, Wood, Selfe, & Peat, 2010; Utting, Davies, & Newman, 2005; Wyndow, Collins, Vicenzino, Tucker, & Crossley, 2016).

STRENGTHS AND LIMITATIONS

A key strength of this study was that we validated eSNAPPS against the paper SNAPPS, ensuring that the psychometric properties of the tool were not lost. We utilised eSNAPPS, which was data managed and scored automatically using programming code, this minimised the risk of error and bias in the measurement, as well as transcription of survey response. This also reduced the costs associated with data handling (Clark & Maynard, 1998). The present study should be interpreted with the following limitations in mind. Self-selection bias is a limitation because those with a history of knee problems might be more likely to want to take part. To minimise this likelihood, participants were told that everyone at the event was being surveyed about their knee health, regardless of whether they had knee pain or not. Additionally, the survey was restricted to those between the ages of 18 and 40. The survey is restricted to this age group because it is difficult to distinguish between PFP and patellofemoral osteoarthritis in those over this age and juvenile knee conditions in those below this age (Dey, et al., 2016). The study design utilised a self-report measure to determine the prevalence of PFP. The SNAPPS tool is not an objective measure or diagnostic test of PFP and may induce recall bias. However, this tool has been previously validated in a clinical and general population setting (Dey, et al., 2016). For pragmatic reasons, this study utilised a non-probability sampling which meant that individuals were not randomly sampled. No information on the demographic breakdown of the whole events was available to determine a sampling framework whereby individuals could be randomly selected. In addition to this, the study was an opportunistic one whereby those attending the mass-participant events were conveniently enrolled. To add to this, the number of individuals who were approached but who declined to participate was not recorded. This would have been useful for evaluating the refusal rate and potential bias due to response rate.

CONCLUSION

In this study, we present eSNAPPS, a validated, electronic questionnaire that can be deployed in large populations to collect data expeditiously, with a low measurement risk of error. Using this tool, it was determined that the 12-month prevalence of PFP of attendees at mass-participation events was 17.4%, lower than the estimates previously reported within the literature. With 15.7% of females and 20.5% of males presenting with PFP. Additional findings show PFP prevalence of spectators and running participants were similar.

ABREVIATIONS: PFP - Patellofemoral pain; SNAPPS - Survey instrument for Natural history, Aetiology and Prevalence of Patellofemoral pain Studies; eSNAPPS - electronic Survey instrument for Natural history, Aetiology and Prevalence of Patellofemoral pain Studies; ICC - Intraclass Correlation Coefficient

FUNDING: This project was funded by the Faculty of Health, Psychology and Social Care at Manchester Metropolitan University. The scheme was Research Excellence Award (code: 294300)

CONFLICTS OF INTEREST: All authors in this study declare that they have no conflict of interest.

DATA AVAILABILITY: Additional tables are available in online supplementary file 1. Paper and eSNAPPS questionnaire along with the scoring guideline and R script can be found at www.snappspfp.com. The SNAPPS questionnaire has already been translated into ten languages, and this is anticipated to increase with time.

CONTRIBUTORS: MY led the writing and drafting of the manuscript. MY, JS, PD, MC, NR and PG have conceived and designed the cross-sectional and validation study, while NR and PG have contributed to the design of the validation study. MY collected the validation study data. MY, JS, MC, SP, JA, JA, PG and RM collected the cross-sectional study data. All authors contributed to the writing and editing of the manuscript and gave approval for the final version.

ACKNOWLEDGMENT: We would like to thank Kirsti Thompson and Cancer Research UK for allowing us to attend the 'Race for life' events to collect data, and Michelle Howard for facilitating this. Also, we would like to thank the following individuals for helping with the data collection; Jason Frain, Sophia Tzioni, Daniel Richardson, Zhao Xiaoxiao and Alex Wharton.

TRANSPARENCY: The lead author affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

ETHICAL APPROVAL: Ethical approval; Faculty ethic committee at Manchester Metropolitan University (ethics number for validation study: 839 & ethics number for cross-sectional study: 2223).

References

Akl, E. A., Maroun, N., Klocke, R. A., Montori, V., & Schünemann, H. J. (2005). Electronic mail was not better than postal mail for surveying residents and faculty. *Journal of Clinical Epidemiology, 58*, 425-429.

Akodu, & Nwakalor. (2018). Patellofemoral pain syndrome: prevalence and coping strategies of amateur runners in Lagos state. *Sports Medicine Journal/Medicina Sportivâ, 14*.

Allender, S., Foster, C., Scarborough, P., & Rayner, M. (2007). The burden of physical activity-related ill health in the UK. *Journal of Epidemiology & Community Health, 61*, 344-348.

Boling, M., Padua, D., Marshall, S., Guskiewicz, K., Pyne, S., & Beutler, A. (2010). Gender differences in the incidence and prevalence of patellofemoral pain syndrome. *Scandinavian journal of medicine & science in sports, 20*, 725-730.

Callaghan, M. J., & Selfe, J. (2007). Has the incidence or prevalence of patellofemoral pain in the general population in the United Kingdom been properly evaluated? *Physical therapy in Sport, 8*, 37-43.

Clark, R., & Maynard, M. (1998). Research Methodology: Using Online Technology for Secondary Analysis of Survey Research Data—" Act Globally, Think Locally". *Social Science Computer Review, 16*, 58-71.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological bulletin, 52*, 281.

Crossley, K. M. (2014). Is patellofemoral osteoarthritis a common sequela of patellofemoral pain? In: BMJ Publishing Group Ltd and British Association of Sport and Exercise Medicine.

Crossley, K. M., van Middelkoop, M., Callaghan, M. J., Collins, N. J., Rathleff, M. S., & Barton, C. J. (2016). 2016 Patellofemoral pain consensus statement from the 4th International Patellofemoral Pain Research Retreat, Manchester. Part 2: recommended physical interventions (exercise, taping, bracing, foot orthoses and combined interventions). *Br J Sports Med, 50*, 844-852.

Dey, P., Callaghan, M., Cook, N., Sephton, R., Sutton, C., Hough, E., James, J., Saqib, R., & Selfe, J. (2016). A questionnaire to identify patellofemoral pain in the community: an exploration of measurement properties. *BMC musculoskeletal disorders, 17*, 237.

Doménech, J., Sanchis-Alfonso, V., & Espejo, B. (2014). Changes in catastrophizing and kinesiophobia are predictive of changes in disability and pain after treatment in patients with anterior knee pain. *Knee surgery, sports traumatology, arthroscopy, 22*, 2295-2300.

Ebert, J. F., Huibers, L., Christensen, B., & Christensen, M. B. (2018). or web-based questionnaire invitations as a method for data collection: cross-sectional comparative study of differences in response rate, completeness of data, and financial cost. *Journal of medical Internet research, 20*.

Eijkenboom, J., Waarsing, J., Oei, E., Bierma-Zeinstra, S., & van Middelkoop, M. (2018). Is patellofemoral pain a precursor to osteoarthritis? Patellofemoral osteoarthritis and patellofemoral pain patients share aberrant patellar shape compared with healthy controls. *Bone & Joint Research, 7*, 541-547.

Ekman, A., & Litton, J.-E. J. E. j. o. e. (2007). New times, new needs; e-epidemiology. *22*, 285-292.

Elson, D. W., Jones, S., Caplan, N., Stewart, S., Gibson, A. S., & Kader, D. F. (2011). The photographic knee pain map: Locating knee pain with an instrument developed for diagnostic, communication and research purposes. *Knee, 18*, 417-423.

Francis, P., Whatman, C., Sheerin, K., Hume, P., & Johnson, M. I. (2019). The Proportion of Lower Limb Running Injuries by Gender, Anatomical Location and Specific Pathology: A Systematic Review. *Journal of sports science & medicine, 18*, 21.

Ibeachu, C., Selfe, J., Sutton, C. J., & Dey, P. (2019). Knee problems are common in young adults and associated with physical activity and not obesity: the findings of a cross-sectional survey in a university cohort. *BMC musculoskeletal disorders, 20*, 116.

Kroth, P. J., McPherson, L., Leverence, R., Pace, W., Daniels, E., Rhyne, R. L., Williams, R. L., & Medicine, P. N. C. J. T. A. o. F. (2009). Combining web-based and mail surveys improves response rates: a PBRN study from PRIME Net. *7*, 245-248.

Lee, I.-M., Shiroma, E. J., Lobelo, F., Puska, P., Blair, S. N., Katzmarzyk, P. T., & Group, L. P. A. S. W. (2012). Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *The lancet, 380*, 219-229.

Liao, J. J. (2010). Sample size calculation for an agreement study. *Pharmaceutical statistics, 9*, 125-132.

Mehl, J., Otto, A., Baldino, J. B., Achtnich, A., Akoto, R., Imhoff, A. B., Scheffler, S., & Petersen, W. (2019). The ACL-deficient knee and the prevalence of meniscus and cartilage lesions: a systematic review and meta-analysis (CRD42017076897). *Archives of orthopaedic and trauma surgery, 139*, 819-841.

Neal, B. S., Lack, S. D., Lankhorst, N. E., Raye, A., Morrissey, D., & van Middelkoop, M. (2019). Risk factors for patellofemoral pain: a systematic review and meta-analysis. *Br J Sports Med, 53*, 270-281.

Peat, G., McCarney, R., & Croft, P. (2001). Knee pain and osteoarthritis in older adults: a review of community burden and current use of primary health care. *Annals of the rheumatic diseases, 60*, 91-97.

Rathleff, M., Skuldbøl, S., Rasch, M. N., Roos, E. M., Rasmussen, S., & Olesen, J. L. (2013). Care-seeking behaviour of adolescents with knee pain: a population-based study among 504 adolescents. *BMC musculoskeletal disorders, 14*, 225.

Rathleff, M. S., Vicenzino, B., Middelkoop, M., Graven-Nielsen, T., van Linschoten, R., Hölmich, P., & Thorborg, K. (2015). Patellofemoral pain in adolescence and adulthood: same same, but different? *Sports Medicine, 45*, 1489-1495.

Sari, N. (2009). Physical inactivity and its impact on healthcare utilization. *Health Economics, 18*, 885-901.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin, 86*, 420.

Smith, B. E., Moffatt, F., Hendrick, P., Bateman, M., Rathleff, M. S., Selfe, J., Smith, T. O., & Logan, P. (2018). The experience of living with patellofemoral pain—loss, confusion and fear-avoidance: a UK qualitative study. *BMJ open, 8*, e018624.

Smith, B. E., Selfe, J., Thacker, D., Hendrick, P., Bateman, M., Moffatt, F., Rathleff, M. S., Smith, T. O., & Logan, P. (2018). Incidence and prevalence of patellofemoral pain: A systematic review and meta-analysis. *PloS one, 13*, e0190892.

Taunton, J. E., Ryan, M. B., Clement, D., McKenzie, D. C., Lloyd-Smith, D., & Zumbo, B. (2002). A retrospective case-control analysis of 2002 running injuries. *British journal of sports medicine, 36*, 95-101.

Team, R. C. (2013). R: A language and environment for statistical computing.

Thomas, M. J., Wood, L., Selfe, J., & Peat, G. (2010). Anterior knee pain in younger adults as a precursor to subsequent patellofemoral osteoarthritis: a systematic review. *BMC musculoskeletal disorders, 11*, 201.

Urwin, M., Symmons, D., Allison, T., Brammah, T., Busby, H., Roxby, M., Simmons, A., & Williams, G. (1998). Estimating the burden of musculoskeletal disorders in the community: the comparative prevalence of symptoms at different anatomical sites, and the relation to social deprivation. *Annals of the Rheumatic Diseases, 57*, 649-655.

Utting, M., Davies, G., & Newman, J. (2005). Is anterior knee pain a predisposing factor to patellofemoral osteoarthritis? *The knee, 12*, 362-365.

Van Gelder, M. M., Bretveld, R. W., & Roeleveld, N. (2010). Web-based questionnaires: the future in epidemiology? *American journal of epidemiology, 172*, 1292-1298.

Webb, R., Brammah, T., Lunt, M., Urwin, M., Allison, T., & Symmons, D. (2004). Opportunities for prevention of ‘clinically significant’knee pain: results from a population-based cross sectional survey. *Journal of Public Health, 26*, 277-284.

Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *The Journal of Strength & Conditioning Research, 19*, 231-240.

Wood, L., Muller, S., & Peat, G. (2011). The epidemiology of patellofemoral disorders in adulthood: a review of routine general practice morbidity recording. *Primary health care research & development, 12*, 157-164.

Wright, K. B. (2005). Researching Internet-Based Populations: Advantages and Disadvantages of Online Survey Research, Online Questionnaire Authoring Software Packages, and Web Survey Services. *Journal of Computer-Mediated Communication, 10*, JCMC1034-JCMC1034.

Wyndow, N., Collins, N., Vicenzino, B., Tucker, K., & Crossley, K. (2016). Is there a biomechanical link between patellofemoral pain and osteoarthritis? A narrative review. *Sports Medicine, 46*, 1797-1808.

Xu, X., Yao, C., Wu, R., Yan, W., Yao, Y., Song, K., Jiang, Q., & Shi, D. (2018). Prevalence of patellofemoral pain and knee pain in the general population of Chinese young adults: a community-based questionnaire survey. *BMC musculoskeletal disorders, 19*, 165.

SUPPLEMENTARY TABLES

Supplementary Table 1: SNAPPS Questions and their scoring

Questions	Scored	Scoring
Age		
Q1.1: Are you aged over 18?	No	
Q1.2: Are you aged under 40?	No	
Q1.3: How old are you?		
Section 1		
Q2: Have you **ever** been to a doctor because of knee problems?	No	
Q3: Have you had pain or problems in **the last year** in or around the knee?	No	No = end of SNAPPS
Section 2 (Clinical features)		
Q4: In which knee have you had pain or problems?	Yes	Both knees=1
Q5: Have you had surgery to your knee? (Including arthroscopy, keyhole surgery, camera in your knee)?	Yes	No=1
Q6: Have you ever had a kneecap that has gone out of joint (dislocated)?	Yes	No=1
Q7: Since starting with your knee problem, does your knee **ever swell up**?	Yes	No=1
Q8: Have you had pain and discomfort for **more than one month**?	Yes	Yes=1
Q10a & Q11a: Thinking about your **right** (left) knee, what do you consider is your main problem with your knee?	Yes	Pain (discomfort)=1
Q10b & Q11b: Thinking about your **right** (left) knee, did your current knee problem come on…	Yes	Gradually=1
Section 3 (Activity related pain)		
Q9a: Because of your knee problems would you suffer from pain or difficulty with **sitting for a long time**?	No	
Q9b: Because of your knee problems would you suffer from pain or difficulty with **going up stairs**?	No	
Q9c: Because of your knee problems would you suffer from pain or difficulty with **going downstairs**?	No	
Q9d: Because of your knee problems would you suffer from pain or difficulty with **squatting**?	No	
Q9e: Because of your knee problems would you suffer from pain or difficulty with **standing for long periods**?	No	
Q9f: Because of your knee problems would you suffer from pain or difficulty with **walking on a level surface**?	No	
Q9g: Because of your knee problems would you suffer from pain or difficulty with **getting up out of a chair**?	No	
Q9h: Because of your knee problems would you suffer from pain or difficulty with **kneeling**?	No	
Q9i: Because of your knee problems would you suffer from pain or difficulty with **walking on uneven surfaces**?	No	

Q9j: Because of your knee problems would you suffer from pain or difficulty with **walking down slopes**?	No	
Q9k: Because of your knee problems would you suffer from pain or difficulty with **walking up slopes**?	No	
Q9l: Because of your knee problems would you suffer from pain or difficulty with **hopping**?	No	
Q9m: Because of your knee problems would you suffer from pain or difficulty with j**umping**?	No	
Q9n: Because of your knee problems would you suffer from pain or difficulty with **running**?	No	
Section 4 (Knee map)		
Q12R_L: Pain on right knee **lateral** patella	Yes	Yes=1
Q12R_M: Pain on right knee **medial** patella	Yes	Yes=1
Q12R_T: Pain on right knee patella **tendon**	Yes	Yes=1
Q12R_O: Pain on right knee **outside** of patella	Yes	Yes=0
Q12L_L: Pain on left knee **lateral** patella	Yes	Yes=1
Q12L_M: Pain on left knee **medial** patella	Yes	Yes=1
Q12L_T: Pain on left knee patella **tendon**	Yes	Yes=1
Q12L_O: Pain on left knee **outside** of patella	Yes	Yes=0
Section 5:		
Q13: Considering both your knees which would you say is the knee that gives you most problems?	No	

Supplementary Table 2: Demographics of runners and spectators

	Runner (n=712)		Spectator (n=368)	
	n	%	n	%
Sex				
Female	561	78.8	138	37.5
Male	151	21.2	230	62.5
Age (years)				
18 – 20	86	12.1	36	9.8
21 – 25	172	24.2	100	27.2
26 – 30	190	26.7	98	26.6
31 – 35	156	21.9	68	18.5
36 – 40	108	15.2	66	17.9

Supplementary Table 3: Prevalence of PFP by demographics of runners and spectators

	Runners (n=712)		Spectators (n=368)	
	n	%	N	%
Sex				
Female	89 / 561	15.9	21 / 138	15.2
Male	28 / 151	18.5	50 / 230	21.7
Age (years)				
18 – 20	13 / 86	15.1	3 / 36	8.3
21 – 25	25 / 172	14.5	18 / 100	18.0
26 – 30	23 / 190	12.1	19 / 98	19.4
31 – 35	36 / 156	23.1	13 /68	19.1
36 – 40	20 / 108	18.5	18 /66	27.3

Supplementary Table 4: Age-sex standardised Prevalence of PFP for runners and spectators

			Runners (n=712)				Spectators (n=368)			
Sex	Age	Standard population	Observed events (PFP)	Observed population	Observed prevalence (%)	Expected prevalence (%)	Observed events (PFP)	Observed population	Observed prevalence (%)	Expected prevalence (%)
Male	18 – 20	58	6	36	16.7	9.686	2	22	9.1	5.780
	21 – 25	97	6	37	16.2	15.714	12	60	20.0	19.400
	26 – 30	92	5	33	15.2	13.984	13	59	22.0	20.240
	31 – 35	73	9	31	29.0	21.170	10	42	23.8	17.374
	36 – 40	61	2	14	14.3	8.723	13	47	27.7	16.897
Females	18 – 20	64	7	50	14.0	8.960	1	14	7.1	4.544
	21 – 25	175	19	135	14.1	24.675	6	40	15.0	26.250
	26 – 30	196	18	157	11.5	22.540	6	39	15.4	30.184
	31 – 35	151	27	125	21.6	32.616	3	26	11.5	17.365
	36 – 40	113	18	94	19.1	21.583	5	19	26.3	29.719
Total		**1080**				**180**				**188**

Age-sex standardised PFP prevalence for Runner: 179.651/1080 = **16.6%** (95% CI 15.2-19.8)
Age-sex standardised PFP prevalence for Spectators: 187.753/1080 = **17.4%** (95% CI 14.5-19.0)
Age-sex standardised PFP prevalence Male runner: 69.277/381 = **18.1%** (95% CI 14.1-22.4)
Age-sex standardised PFP prevalence Female runners = 110.374/699 = **15.7%** (95% CI 13.1-18.7)
Age-sex standardised PFP prevalence Male spectators = 79.691/381 = **21.0%** (95% CI 17.0-25.4)
Age-sex standardised PFP prevalence Female spectators = 108.062/699= **15.5%** (95% CI 12.9-18.4)