APPENDIXES

Appendix 1. Supplementary explanations of methods

Recruitment via online methods:

All participants were directed to the studies through either word of mouth via clinicians, or through online methods (supplementary methods). via: the Body in Mind blog (www.bodyinmind.org), social media (Facebook and Twitter), Pelvic Pain SA, the Pelvic Pain Foundation of Australia, and EndoActive. All participants accessed the study online via Survey Monkey (www.surveymonkey.com).

eDelphi technique:

The Delphi technique collects information from a panel of experts.[9] In a conventional Delphi, experts are sent a questionnaire on a particular topic and a second questionnaire is created based on the results of the first. A critical aspect of the current study was that the experts were women with pelvic pain. These rounds of questionnaires are continually refined until a consensus is reached. Because a Delphi is traditionally employed using paper questionnaires, one criticism of the process is the potential for poor response rates.[5] Often Delphi studies that are run online (eDelphi) are more convenient for participants, leading to higher response rates.[3] For this reason, we ran this study entirely online.

Point-allocation approach to item weighting:

In round two of the eDelphi, participants were asked to distribute, across the 25 aspects, a total of 100 points, according to the level of impact that each aspect had on their life. Aspects that did not impact their life at all would receive zero points. Round three was a repeat of the 100-point distribution task used in round two, and was used to ensure responses were consistent at different time points. Aspects were ranked from highest average score to lowest average score.

Rasch Analysis

We used Rasch analysis to determine the psychometric properties of the PPIQ. Rasch analysis is based on a probabilistic model that uses the level of endorsement of questions to investigate the integrity of the questionnaire. Participants are said to have 'endorsed' a question if they have selected a response of 1 or above on a question. We used Rasch analysis to assess whether two main assumptions were met: first, that a participant who is greatly impacted by pelvic pain will have a greater probability of endorsing any PPIQ question than a participant who is less impacted; and second, that the probability of any participants endorsing a question indicative of high impact is less than that of any participant endorsing a question indicative of lower impact. Rasch analysis is an appropriate model to assess our sample as it provides a method in which to assess whether the questions targeted the sample appropriately, whether the questions formed a unidimensional scale, and whether PPIQ scores can be used as an interval level measurement. Unidimensionality and interval level measurement are important requirements if scores are to be summated to provide a measure of impact, which is critical for easy use clinically. For a comprehensive overview of the Rasch model see Bond & Fox (2013).[2]

Targeting refers how well the PPIQ questions assess the sample of participants. Ideally, a well-functioning questionnaire contains a spread of questions that target participants with varying levels of impact. We assessed targeting by visually inspecting the distribution of participant scores across the PPIQ questions and through a comparison of the summary statistics. The average question endorsability was anchored at zero logits by software default. A logit is an equal-interval unit that is used to report relative differences between participants' ability estimates (i.e. their level of impact) and question endorsability.[1] Therefore, a positive average participant score would suggest the sample were more impacted by pelvic pain than the average of the questionnaire, and a negative average value would suggest the opposite.[14]

To determine whether the sample used the Likert rating scale in the intended manner, we analysed the category ordering. We analysed the PPIQ using the Andrich Rating Scale model because all of the questions shared the same Likert scale categories.[8] Analysis of ordering determined whether the participants were able to discriminate between the incremented categories of impact. The PPIQ has five response categories (0 to 4) and thus four thresholds, termed 'step-calibrations', at which the likelihood of endorsing one category is equal to that of endorsing the next. We assessed whether each category was used by the participants and whether the category thresholds advanced in the expected manner (e.g. can the participants differentiate between 'A little bit' and 'somewhat'?).

Stage 1 identified the aspects of life most impacted by pelvic pain and these aspects were used to form the PPIQ questions. An intended purpose of the PPIQ is to provide an overall measure of pelvic pain impact, so each question should assess a component of this unidimensional construct. That is, each question should share in common an aspect of impact yet be sufficiently different so as not to be redundant. Assessment of unidimensionality looks to identify clusters of questions that together may be assessing a secondary dimension, thus threatening measurement of the primary dimension. We assessed unidimensionality through analysis of question fit statistics and through principal components analysis (PCA) of residuals.[13]

The chi-square based fit statistics, reported as mean-squares (in logits), have an expected value of 1 logit. Values greater than 1 indicate randomness in the data while values less than 1 indicate the data are too predictable. We assessed both the infit (information-weighted fit statistic) and outfit (outlier-sensitive) statistics and considered them excessive if they were in greater than 1.4, or less than 0.6 logits.[15] Question characteristic curves display the Rasch-model prediction of how participants with differing levels of impact would score on a question. The curves of misfitting questions were visually inspected to assess how the questions were used. PCA of residuals investigates the part of the data that do not fit within the Rasch model (i.e. the discrepancies). The residuals are determined by the observed responses minus their expectations; again, these are reported as logits. The analysis assesses whether clusters of questions share the same pattern of discrepancies, indicating that the questionnaire possesses a second dimension. PCA was conducted, and the residual correlation matrix was inspected visually to identify unexpected clusters of questions with excessive positive or negative residuals.[2] Questions with substantial positive or negative loadings equivalent to an eigenvalue greater than 2 (the strength of two questions) were reviewed.[12]

PCA also allows for a test of local independence of questions.[7] High correlations between questions suggest the response to one question is reliant on the response to the other and may lead to false interpretation of resultant data. We decided *a priori* to subjectively review the content of any question pairs that had a correlation greater than 0.5. Person fit was assessed to determine whether the participants used the scale in a predictable manner. Here, we considered person fit to be excessive if their fit residuals were greater than 2 logits.[14] Misfitting persons were compared to those who fit the model across variables using a chi-square test of significance (for diagnosis and duration) and an independent samples t-test (for age).[6] Response strings of those misfitting persons were visually analysed to identify patterns in their responses.
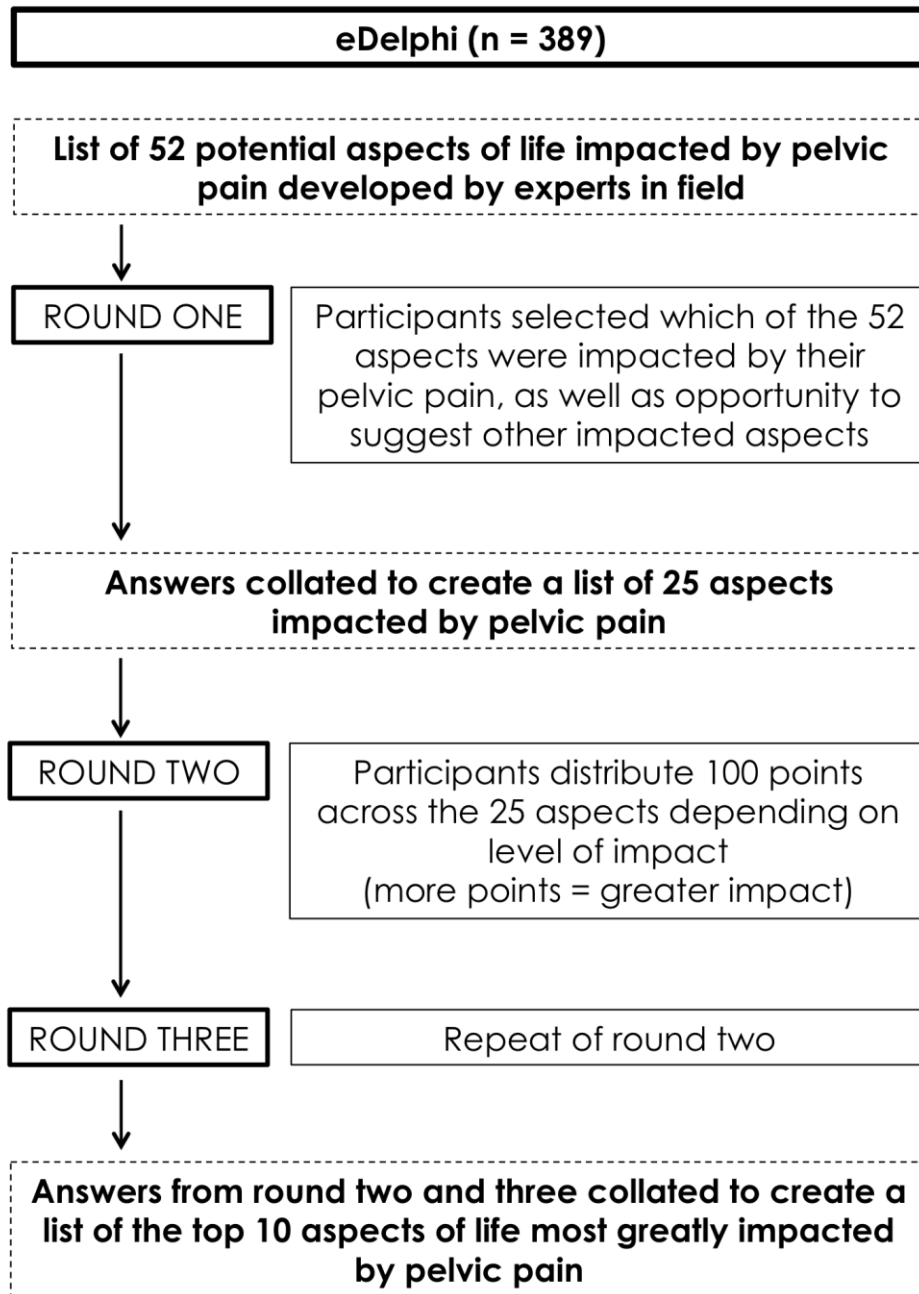
Internal consistency was assessed via the person reliability index [4] and is the Rasch equivalent of Cronbach's alpha.[14] It indicates whether the questionnaire is sensitive enough to distinguish between those greatly impacted and those less impacted. We determined, *a priori*, that values above 0.7 and 0.85 would be used to indicate the appropriateness of the scale for group and individual use, respectively.[14]

Question bias occurs when characteristics other than the variable of interest change the functioning of the scale (i.e. two persons for whom pelvic pain has a similar true impact endorse an question differently because of some other characteristic). Bias was assessed using a Mantelchi-squared test[10]. We assessed whether the participant's age (≤35, >35 years); duration of symptoms (0 – 12 months, >12 months); or diagnosis biased the functioning of the scale. Questions with statistically significant ($p < 0.01$) contrasts in excess of 0.5 logits were further explored.[7, 11] If bias was apparent in a small subset of participants, a sensitivity analysis was conducted. Three random samples of equal size to the bias subset were drawn from the larger sample and compared.

References

[1] What is a Logit? In: NCoSBo Nursing editor. Chicago, IL: National Council of State Boards of Nursing; Examinations Department, 2013.
[2] Bond TG, Fox CM. Applying the Rasch model: Fundamental measurement in the human sciences: Psychology Press, 2013.
[3] Brüggen E, Willems P. A critical comparison of offline focus groups, online focus groups and e-Delphi. International Journal of Market Research 2009;51(3):363-381.
[4] Fox C, Jones J. Uses of Rasch Modeling in Counseling Psychology Research. Journal of Counseling Psychology 1998;45(1):30-45.
[5] Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. Journal of advanced nursing 2000;32(4):1008-1015.
[6] Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. Health Technology Assessment (Winchester, England) 2009;13(12):iii.
[7] Hungi N. Employing the rasch model to detect biased items. In: S Alagumalai, DD Curtis, N Hungi, editors. Applied rasch measurement: A book of exemplars. Dordrecht, The Netherlands: Springer, 2005.
[8] Linacre JM. Comparing "Partial Credit Models" (PCM) and "Rating Scale Models" (RSM). Rasch Measurement Transactions 2000;14(3):768.
[9] Linstone HA, Turoff M. The Delphi method: Techniques and applications, Vol. 29: Addison-Wesley Reading, MA, 1975.
[10] Mantel N. Chi-square tests with one degree of freedom; extensions of the Mantel-Haenszel procedure. Journal of the American Statistical Association 1963;58(303):690-700.
[11] Mantel N. Chi-square tests with one degree of freedom: extensions of the Mantel-Haenszel procedure. Journal of the American Statistical Association 1963;58(303):690-700.
[12] Raiche G. Critical Eigenvalues sizes in standardized residual principal components analysis. Rasch Measurement Transactions 2005;19(1):1012.
[13] Smith EV, Jr. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. Journal of applied measurement 2002;3(2):205-231.
[14] Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis & Rheumatism 2007;57(8):1358-1362.
[15] Wright BD, Linacre JM, Gustafson J, Martin-Lof P. Reasonable mean-square fit values. Rasch measurement transactions 1994;8(3):370.

Appendix 2. The three rounds of the eDelphi. Each round was followed by refining of the data collected and was sent out to participants again. Finally, a list of the top ten aspects of life most greatly impacted by pelvic pain was created and formed the ten questions of the Pelvic Pain Impact Questionnaire.

```
┌─────────────────────────────────────────────────┐
│              eDelphi (n = 389)                   │
└─────────────────────────────────────────────────┘

┌ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ┐
   List of 52 potential aspects of life impacted by pelvic
        pain developed by experts in field
└ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ┘
          │
          ▼
┌──────────────┐   ┌────────────────────────────────┐
│  ROUND ONE   │   │  Participants selected which of the 52 │
└──────────────┘   │   aspects were impacted by their       │
       │           │  pelvic pain, as well as opportunity to│
       │           │   suggest other impacted aspects       │
       │           └────────────────────────────────┘
       ▼
┌ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ┐
   Answers collated to create a list of 25 aspects
              impacted by pelvic pain
└ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ┘
       │
       ▼
┌──────────────┐   ┌────────────────────────────────┐
│  ROUND TWO   │   │  Participants distribute 100 points    │
└──────────────┘   │   across the 25 aspects depending on   │
       │           │            level of impact             │
       │           │   (more points = greater impact)       │
       │           └────────────────────────────────┘
       ▼
┌──────────────┐   ┌────────────────────────────────┐
│ ROUND THREE  │   │        Repeat of round two             │
└──────────────┘   └────────────────────────────────┘
       │
       ▼
┌ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ┐
   Answers from round two and three collated to create a
   list of the top 10 aspects of life most greatly impacted
                     by pelvic pain
└ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ┘
```

Appendix 3. The 52 statements used in the first round of the eDelphi study, ordered from most selected to least selected statements. Symbols indicate which questions were combined to form one question in subsequent rounds two and three.

Statement	Percentage selected
It makes me less energetic*	45.1
Having sex^	39.5
It makes me feel anxious§	37.5
It makes me more sad§	33.2
It makes me feel depressed§	33.2
It makes me feel tense§	33.0
It makes me feel helpless§	32.7
Jogging or running	32.5
I have broken sleep cycles at night¤	30.5
Wear tight fitting clothes◊	29.3
Sitting for a long period of time (greater than 20 minutes)˜	25.7
I am not a spiritual person	23.5
Walking long distances (taking more than 20 minutes) Δ	23.3
Household activities such as cooking, cleaning, and laundry#	22.8
My pelvic pain does not affect my appetite	22.6
Go out with friends socially	22.1
I have trouble getting to sleep at night¤	21.9
I find it hard to sit for as long as my work requires□	21.4
I find it hard to use tampons	19.9
My pelvic pain does not affect my spiritual life	17.8
I find it hard to do the manual labor that my work requires□	17.4
It makes me more sleepy during the day¤	16.7
My pelvic pain does not affect my working life	16.5
It makes me need to have a sleep during the day¤	15.6
Masturbating^	15.3
It makes me less hungry&	15.1
I find it hard to travel to work□	14.7
Team sports (please elaborate below) Δ	14.7
I have trouble waking up in the morning¤	14.2
My pelvic pain does not affect my daily activities	11.7
My pelvic pain has no effect on my sleep	11.3
Walking short distances (taking less than 20 minutes)Δ	11.3
Wear pants◊	11.3
It makes me eat less healthy food&	10.2
Sitting for a short period of time (less than 20 minutes)˜	9.5
I get cravings&	9.3
My pelvic pain has no effect on my energy levels	9.0
I am unable to have a gynaecology exam that involves using a speculum	8.6
It causes me to question my faith∞	7.4
I find it hard to take baths□	7.0
Wear underwear	6.8
I find it hard to shower□	5.9
I find it hard to use sanitary pads□	5.9
It makes me eat more healthy food&	5.9
It causes me to miss church/mass/temple/shul etc.∞	5.6
It makes me more hungry&	5.4
My pelvic pain does not affect my leisure activities	5.4
I find it hard to clean my pelvic/vaginal region□	5.0
It causes me to question my deity∞	3.8
My pelvic pain does not affect my mood	3.4
It makes me more awake during the day¤	2.3
It makes me more energetic*	0.5

*Energy levels; ^Levels of intimacy or sexual relationships; §Mood changes; ¤ Sleep disturbances; ◊ Wearing tight fitting clothes; □Performing and functioning normally at home/work/school/university; & Changes in diet or appetite; ~Sitting for longer than 20 minutes; Δ Taking part in sport activities; ∞My religion/spirituality; Qualitative responses from participants lead to the addition of eight items to the final list of 25 used in Rounds Two and Three (standing, lifting objects, concentration and memory, making plans and completing them, medication side effects, engaging in social relationships, walking uphill, and climbing stairs/ladders).

Appendix 4. The 25 most commonly reported aspects of life impacted by pelvic pain in the lives of participants, as identified by participants in Round One of the eDelphi study. Aspects are ordered from most selected to least selected.

Levels of intimacy or sexual relationships (e.g. having sex, masturbating)
Stomach and intestinal symptoms (e.g. abdominal pain, discomfort, nausea)
Mood changes (including feelings of anxiety, depression, anger, sadness, frustration)
Using tampons
Sitting for longer than 20 minutes
Taking part in sport activities (e.g. bike riding, swimming, yoga, pilates, horse riding)
Jogging/running
Performing and functioning normally at home/work/school/university
Energy levels
Sleep disturbances (e.g. broken sleep cycles, ability to fall asleep)
Wearing tight fitting clothes
Standing for longer than 20 minutes
Problems with medications (e.g. side effects or dependency on medications)
Changes in diet or appetite
Lifting objects (including children)
Household activities (e.g. cooking, cleaning, laundry)
Taking part in social activities (e.g. going out with friends, going out for dinner, going to the movies)
Engaging in social relationships (with friends, work colleagues, etc.)
My pelvic pain does not affect any of these aspects of my life
Making plans, accomplishing goals, and completing tasks
Wearing underwear
Walking uphill
Concentration and memory
Climbing stairs or ladders
My religion/spirituality

Appendix 5 attached separately

Appendix 6. Trial protocol

The protocol for this trial was published on a public blog, Body in Mind, and is available at:
http://www.bodyinmind.org/resources/protocols/pelvic-pain-impact-questionnaire-ppiq/