Article:
Khera, T., Hunt, L.P., Davis, S. orcid.org/0000-0002-6609-4287 et al. (7 more authors) (2022) A clinical tool to identify older women with back pain at high risk of osteoporotic vertebral fractures (Vfrac) : a population-based cohort study with exploratory economic evaluation. Age and Ageing, 51 (3). afac031. ISSN 0002-0729

https://doi.org/10.1093/ageing/afac031

A clinical tool to identify older women with back pain at high risk of osteoporotic vertebral fractures (Vfrac): a population-based cohort study with exploratory economic evaluation

Khera T[1], Hunt LP[2], Davis S[3], Gooberman-Hill R[4], Thom H[5], Xu Y[6], Paskins Z[7], Peters TJ[8], Tobias JH[9], Clark EM[10]

1 Vfrac Research Manager, Bristol Medical School, University of Bristol, Bristol, UK.
2 Senior Research Fellow, Bristol Medical School, University of Bristol, Bristol, UK.
3 Senior Lecturer in Health Economics, School of Health & Related Research, University of Sheffield, Sheffield, UK.
4 Professor of Health and Anthropology/Director of Elizabeth Blackwell Institute, NIHR Bristol Biomedical Research Centre, Bristol, UK; and Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK.
5 Lecturer in Health Economics, Bristol Medical School, University of Bristol, Bristol, UK.
6 PhD Student, Bristol Medical School, University of Bristol, Bristol, UK.
7 Reader in Rheumatology, School of Medicine, Keele University, Staffordshire, UK; and Haywood Academic Rheumatology Centre, Midland Partnership NHS Foundation Trust, Stoke-on-Trent, UK.
8 Professor of Primary Care Health Services Research, Bristol Medical School, University of Bristol, Bristol, UK.
9 Professor of Rheumatology, Bristol Medical School, University of Bristol, Bristol, UK; and MRC Integrative Epidemiology Unit, University of Bristol, UK.
10 Professor of Clinical Musculoskeletal Epidemiology, Bristol Medical School, University of Bristol, Bristol, UK; and Honorary Consultant Rheumatologist, North Bristol NHS Trust, Bristol. UK.

Corresponding author: Prof Emma Clark
Professor of Clinical Musculoskeletal Epidemiology
Bristol Medical School
University of Bristol
Email: emma.clark@bristol.ac.uk

Acknowledgements

The authors would like to thank the following people, without whom the study could not have been completed: Dr Sarah Gowlett, Consultant Radiologist at Sheffield Teaching Hospitals NHS Foundation Trust for her help with the repeatability measures; Pat Mascord and Yvonne Saddler our patient partners who contributed hugely to our Steering Committee; and Charlene Wisdom-Trew the Vfrac administrator. The Vfrac study was funded via an unrestricted Clinical Studies research grant from Arthritis Research UK (Versus Arthritis) No 21507. HT was supported by MRC grant MR/S036709/1 and the National Institute for Health Research (NIHR) Bristol Biomedical Research Centre (BRC). Finally we would like to thank the study participants for giving up their time and sharing their experiences.

Manuscript word count (excluding abstract and refs): 2988

Key words
Vertebral fractures
Back pain
Osteoporosis
Vfrac
Cohort study

Key points

- Vfrac is a clinical tool consisting of 15 questions that can be performed by a practice nurse. The output is a recommendation, or not, for spinal radiographs.
- Vfrac has good sensitivity and specificity for identification of older women with back pain who have OVFs
- Identification of those with OVFs is improved through the addition of self-reported back pain descriptors
- Health economic modelling indicates there is potential value in a future randomised controlled trial to evaluate the Vfrac tool

ABSTRACT (n=241)

Background: Osteoporotic vertebral fractures (OVFs) identify people at high risk of future fractures, but despite this, less than a third come to clinical attention. The objective of this study was to develop a clinical tool to aid healthcare professionals decide which older women with back pain should have a spinal radiograph.

Methods: A population-based cohort of 1635 women aged 65+ years with self-reported back pain in the previous four months were recruited from primary care. Exposure data were collected through self-completion questionnaires and physical examination including descriptions of back pain and traditional risk factors for osteoporosis. Outcome was the presence/absence of OVFs on spinal radiographs. Logistic regression models identified independent predictors of OVFs, with the Area Under the (Receiver Operating) Curve (AUC) calculated for the final model, and a cut-point identified.

Results: Mean age was 73.9 years and 209 (12.8%) had OVFs. The final Vfrac model comprised 15 predictors of OVF, with an AUC of 0.802 (95%CI 0.764-0.840). Sensitivity was 72.4% and specificity 72.9%. Vfrac identified 93% of those with >1 OVF and two-thirds of those with one OVF. Performance was enhanced by inclusion of self-reported back pain descriptors, removal of which reduced AUC to 0.742 (95%CI 0.696-0.788) and sensitivity to 66.5%. Health economic modelling to support a future trial was favourable.

Conclusions: The Vfrac clinical tool appears valid and is improved by the addition of self-reported back pain symptoms. The tool now requires testing to establish real-world clinical and cost-effectiveness.

Osteoporosis and associated fragility fractures are one of the most common musculoskeletal conditions in older people, and approximately three million people in the UK have osteoporosis (www.ons.gov.uk). Osteoporotic vertebral fractures (OVFs) are of particular importance, as they identify people at a high risk of future fracture: within five years of the occurrence, one in four people will have a further vertebral fracture[2], and one in ten will have a limb fracture including hip fracture[3]. In addition, osteoporotic fractures lead to morbidity and disability: more than a third of patients that experience OVFs have difficulties with activities of daily living for the rest of their lives[4]. However, medications are available to reduce the risk of further vertebral fracture by between 31-65% for bisphosphonates[5] and even greater reductions are possible with anabolic treatments[6]. Despite this, less than a third of people with OVFs come to clinical attention[7]. There are many possible explanations for this diagnostic failure. However, a major reason is a need for knowledge and understanding about which clinical features should trigger referral for diagnostic spinal radiographs in people with possible OVFs[10].

To address this, using the MRC framework for development and evaluation of complex interventions[11], we developed a clinical decision tool called 'Vfrac'. The tool helps primary care practitioners decide if an older woman with back pain is at high risk of an OVF and requires a spinal radiograph to confirm the diagnosis.

We carried out four preliminary studies. First, we performed a cross-sectional study of 509 women from primary care[12]and identified four independent predictors of OVF that could be combined into a tool to determine who should have spinal radiographs. A pre-determined cut-point gave a sensitivity of 62% and a specificity of 71%. We investigated this prototype tool to identify if it could have utility for clinical decisions in a randomised controlled trial of 3200 unselected older women from the community[13]. Results showed that allocation to screening approximately doubled the odds of a new prescription for osteoporosis medications (OR 2.24, 95%CI 1.16 to 4.33, P=0.016). However, cost-effectiveness modelling suggested it was unlikely to be cost-effective from the NHS perspective, mainly because of a low prevalence of OVF in this unselected population. We then focussed on the population of older women with back pain[15], and identified six independent predictors of OVF, four of which were newly identified pain descriptors. Finally, to identify if other symptoms, sensations and pain experiences had been missed, we carried out a qualitative focus group study[16] of 19 older adults with OVFs. Results showed that women's experiences of vertebral fractures related to seven sensations, with pain the dominant one.

The aims of this current study were to: (1) enhance the original prototype tool[13] with the newly identified pain descriptors[15] and other sensations[16], to develop an improved clinical decision tool (Vfrac) for use when older women present to primary care with back pain; (2) identify the changes in prediction accuracy when including self-reported back pain descriptors over and above traditional risk factors for OVF; and (3) estimate the tool's potential cost-effectiveness to identify if it is reasonable to conduct further evaluation.

METHODS

The Vfrac cohort study recruited participants from multiple general practices within two areas of the UK: Bristol and Stoke-on-Trent. Research ethics approval was obtained from the National Research Ethics Service (West of Scotland REC 18/WS/0061). All participants provided written consent. The protocol was published before data collection[17]; one substantial amendment was made to allow recruitment of those who had already had a spinal radiograph within the previous 4 months. The study was registered with the ISRCTN Registry (https://doi.org/10.1186/ISRCTN16550671).

Study design and participants

Thirteen general practices from a range of deprivation scores as assessed by the Index of Multiple Deprivation (IMD) were recruited from Stoke-on-Trent and nine from Bristol. Women aged 65 or more with a self-reported episode of back pain in the previous four months were recruited. For more information, see Supplementary Data, Section 1.

Exposure data

Back pain data: A wide range of questions were included in a self-completion questionnaire (see protocol paper for full description[17]) based on previous studies on women with and without OVFs[15,18] plus other back pain questionnaires[19,20]. Findings from the qualitative study[16] were also used to develop questions for quantitative data collection. The Margolis pain diagram[21] was included for participants to mark the anatomical site of their back pain.

Other self-reported data: Data were collected on frailty, traditional risk factors for osteoporosis, concomitant illnesses, health related quality of life, healthcare usage at baseline and three months later and use of pain relieving medication at baseline and three months later using the same question structure as previous studies[15,23]. Fragility fracture was defined as fracture after aged 50 excluding hands, feet, head and excluding high trauma.

Physical examination: Data collected by a trained research nurse were height, weight, chest expansion, waist circumference, rib-to-pelvis distance and wall-tragus distance. Reported height loss was calculated by subtracting the height measured in the research clinic from self-reported height at 25 years of age. For more information see Supplementary Data, Section 1.

Outcome data: OVFs

All participants had lateral thoracic and lumbar radiographs. Radiographs were assessed for the presence or absence of OVFs by EC using the Algorithm-Based Qualitative (ABQ) method[29].

Radiographs were categorised into those with no fracture or with fracture. Those with OVFs were further categorised into mild, moderate or severe fractures based on their 'worst' fracture using the Genant semi-quantitative (SQ) method[30]. Repeatability of the primary outcome was assessed by a random sample of anonymised images reviewed by EC and an independent experienced radiologist (SG) 4 months after completion of initial data collection. Results showed complete agreement for intra-rater reliability by EC. The kappa for agreement between EC and SG was 0.689 indicating substantial agreement. There was 100% agreement between EC and SG for moderate and severe OVFs.

Statistical analysis

Preliminary univariable analysis explored relationships between each predictor variable and OVF using logistic regression. Variables found related to OVF with $P<0.1$ were taken forward to the next stage of the analysis. For this, a series of logistic regression models were carried out using subsets of the predictor variables; this pragmatic approach was adopted as many predictors had missing values. Groups of predictor variables were considered together using backwards stepwise logistic regression analyses to remove those with $P>0.1$. Age was constrained to stay in the model, irrespective of its P value. The reduced subsets of predictor variables were then combined and analysed with a similar backwards stepwise approach. Having determined a 'final' model, the discarded predictors were added back individually to check that none would further improve the model. Regression coefficients needed to calculate the linear predictor, the maximum likelihood R-squared and AUC calculated are reported for the final model obtained. Model validation included calibration-in-the-large (CITL), calibration slope and heuristic shrinkage[31]. Five hundred bootstrapped samples were created and used to estimate shrinkage and adjust the calibration slope and AUC optimism. As the final model was calculated from complete cases, 10 multiply imputed data sets were combined to re-estimate the regression coefficients on the full set. A cut-point of the final linear predictor was identified based on a maximised sum of sensitivity and specificity chosen because this method weights false negatives and positives equally, and is equivalent to minimising Youden's Index. The added benefit of the use of self-reported symptoms was assessed by looking at the proportions of those identified with OVFs using the cut-point before and after removal of these symptoms.

Sample size

Full details are available in the protocol paper[17]. The sample size was calculated as 1633, based on the following assumptions: a prevalence of OVFs between 12-20% based on data from the European

Vertebral Osteoporosis Study[32]; a margin of error of 5%, and sensitivity and specificity of the Vfrac tool between 80 and 95%.

Health economic analyses

Full explanation of the health economic analysis is available in the Supplementary Data, Section 1, Methods. In addition, Supplementary Data, Section 2 describes a within-study analysis, the results of which drove the requirement to move to a modelling-based approach for the economic analysis. The decision tree structure used for modelling is illustrated in Supplementary Data, Section 3. Current standard of care was defined from stakeholder work as consultation with GP for back pain followed by potential referral for radiograph. To compare the cost-effectiveness of the Vfrac tool to this standard of care, the proportions of people diagnosed with OVF by the Vfrac tool and by current standard of care were modelled, as were the life-time costs and quality-adjusted-life-years (QALYs). Simulations were used to estimate expected lifetime costs and QALYs according to whether the individual received treatment with the bisphosphonate alendronic acid, or no treatment, using a previously published bisphosphonate cost-effectiveness model[33]. An NHS and personal social services perspective was adopted for the analysis. For both Vfrac and standard of care groups, life-time net benefits were calculated at a willingness to pay threshold of £20,000/QALY. These were used to calculate the probability that Vfrac or standard of care was most cost-effective (i.e. intervention with greatest net benefit at £20,000/QALY). Expected Value of Perfect Information (EVPI) per person and population EVPI were estimated[34], to measure the value of removing all uncertainty in all parameters.

RESULTS

A total of 1635 participants were recruited (see Supplementary Data, Section 4 for STROBE diagram), with a mean age of 73.9 years (range 65.4-96.8 years). Of these, 209 (12.8%) had VFs: 134 (8.2%) with one and 75 (4.6%) with more than one (range 2-9). Thirty-four participants were excluded from further analysis (33 due to spinal malignancy/metalwork, 1 due to missing baseline questionnaire), leaving 1601 (202, 12.6%, with OVF) for the main analysis.

Full data of all univariable analyses are in the Supplementary Data, Section 5,. Initially univariable analyses were undertaken to look at associations between the individual descriptive words for back pain and presence or absence of OVF. Backwards stepwise logistic regression analysis identified the strongest determinants of OVF (Supplementary Tables 3 and 4). Similarly, univariable analyses were undertaken to look at associations between change in back pain with specific activities (Supplementary Table 5), anatomical site of pain (Supplementary Table 6), change in pain over time (Supplementary Table 7) or posture related pain (Supplementary Table 8) and the presence or absence of OVF. At this stage, the 12 putative pain predictors were combined together, with backwards stepwise analysis used to identify which were the strongest determinants of OVF (Supplementary Table 9). Only back pain described as stinging, described as sharp, described as like toothache, agreement with 'If I'm working in the kitchen, like chopping vegetables or washing, my back pain gets worse and worse to reach a peak – then I have to sit down immediately' and pain marked on the Margolis diagram in the thoracic or low back/buttock area were associated with OVF.

Data were collected on whether specific situations increased back pain, decreased back pain or had no effect (Supplementary Table 10). Eight putative predictors were identified, but a multivariable backwards stepwise analysis removed three, leaving bending, sitting on straight backed chairs, sitting on soft chairs, sleeping and changes in the weather.

Subsequent univariable analyses identified frailty variables, walking distance and use of walking aids were associated with the presence or absence of OVF (Supplementary Table 11). Next, univariable associations between traditional risk factor for osteoporosis and the presence of absence of OVF were assessed, and use of oral steroids for >3 months was identified, along with previous fracture (Supplementary Tables 12 and 13). No association was seen between concomitant illnesses and the presence or absence of OVF (Supplementary Table 14).

Finally, data collected during the physical examination was analysed for associations with OVF (Supplementary Table 15). Backward stepwise analysis removed four variables leaving weight, wall-to-tragus and height loss as independent predictors of OVF, together with age (Supplementary Table 16).

Backwards logistic regression further reduced the important variables from Supplementary Tables 9, 10 and 16 (n=1490). All variables that had been excluded, either in previous steps or preliminary univariable analyses were then re-assessed. Only steroid use for >3 months approached significance (P=0.052) and given the well-recognised clinical association between glucocorticoids and OVF, was added back in to produce the final model. Further backwards stepwise removal and some close scrutiny of the resulting models finally led to the inclusion of two extra variables: pain affected by walking and reclining.

The final Vfrac model is shown in Table 1. The prevalence of OVF in this final data set was 163/1337 (12.2%). The mean value of the linear predictor was -2.50 (SD 1.25), with the mean±SD linear predictor for those without OVFs being -2.68±1.12 and for those with OVFs being -1.18±1.37. This yielded an AUC of 0.802 (0.764-0.840) (Figure 1A). The calibration slope was 1.0 showing no evidence of overfitting or underfitting (Figure 1B). A heuristic estimate of shrinkage was calculated to be 0.925. This was used to estimate a 'shrunken' linear predictor to assess the impact of regression towards the mean in any future real-world use. The mean (SD) of the shrunken linear predictor was -2.43 (SD 1.16). Figure 1C compares the distributions of the linear predictor and with its the shrunken values. Figure 2 illustrates the separations accorded by the linear predictor between participants with none, one or more than one OVFs (Figure 2A) and none, mild, moderate and severe OVFs (Figure 2B). From 500 bootstrapped samples, optimism in the estimate of the AUC was estimated to be 0.019, therefore the optimism-adjusted AUC was 0.783. Finally, as a secondary analysis to check our results, multiple imputation was used to account for the missing data, with results for the imputed model being similar to those seen in Table 1.

The final cut-point of the linear predictor for identification of which older women with back pain should have a spinal radiograph because of a high risk of fracture was -2.00, chosen as this gave a sensitivity of 72.4% and a specificity of 72.9%. Assuming the same prevalence identified in this study cohort, the final model has a positive predictive value of 27.1% and a negative predictive value of 95.0% (Table 2). Without inclusion of back pain symptoms, the Vfrac tool identifies 66.5% of those with OVFs (53.7% with one; 92.5% with more than one), and sensitivity is reduced to 66.5%. Adding

back pain symptoms identifies 72.4% of those with OVFs (62.0% with one OVF; 92.7% with more than one), as shown in Table 2. Removing self-reported symptoms reduces the AUC from 0.802 to 0.742 (95%CI 0.696 to 0.788).

Cost-effectiveness results are presented in Table 3. The lifetime incremental net benefit for Vfrac tool compared to standard of care is £1.47 (95% Credible Interval -£2587, £2456) with 49.4% probability of being cost-effective. The uncertainty translates into a high value in future research with the estimated EVPI being £526 per patient and EVPI per population being £229-458 million, comfortably outweighing the cost of any large scale randomized controlled trial.

DISCUSSION

We now have a clinical tool, Vfrac, to help healthcare practitioners decide which older women presenting to primary care with back pain are at high risk of currently having one or more OVFs and therefore require a diagnostic spinal radiograph. Of these recommended to have radiographs by Vfrac, approximately one third will have an OVF. Furthermore, Vfrac will identify >90% of those with severe OVFs and approximately two-thirds of those with mild or moderate fractures. The output looks robust and valid. Preliminary modelling suggests there is great uncertainty about the cost-effectiveness of implementing Vfrac and these findings strongly support a new trial of Vfrac to establish its real-world clinical and cost-effectiveness.

There are simple clinical tools to guide osteoporosis management decisions currently in use within primary care such as FRAX which estimate future probability of major osteoporotic and hip fracture (https://www.shef.ac.uk/FRAX/). However, FRAX does not give any information on risk of existing (prevalent) OVFs. The Vfrac tool is unique, as the only evidence-based decision tool able to highlight an individual who should have a spinal radiograph because of their risk of an existing OVF.

Compared to the original cross-sectional study[12], sensitivity and specificity for identification of people with OVFs is improved with the Vfrac tool, presumably because more detailed information about back pain has been included. In addition, compared to the previous randomised controlled trial in unselected older women, the Vfrac tool identifies a higher proportion of those with moderate and severe OVFs. Finally, patients' accounts of back pain are necessarily subjective in nature, but our modelling suggests the use of self-reported back pain descriptors in addition to more traditional risk factors for osteoporosis improves the AUC of Vfrac, particularly for those with one OVF.

There are limitations to this study. The recruited study population is unlikely to be fully representative of the background population. Our study has a shortfall in Asian, Black, Mixed and other ethnicities and, *a priori*, no men were included to restrict the development work to those with a high background prevalence of osteoporosis (women). Vfrac was also designed before the coronavirus pandemic, and much more healthcare is now delivered virtually by telephone or video, pointing to the need to develop Vfrac as a remote self-completion tool, which is now being tested. As the Vfrac tool is targeted at those presenting with back pain, it cannot identify those with asymptomatic OVF, and real world testing is required to identify if limiting Vfrac to those with back pain impacts on the clinical- or cost-effectiveness. Finally, to evaluate cost-effectiveness definitively there needs to be a control comparator. The study was not designed to provide this comparison, but

to assess whether there were grounds to conduct further evaluation of the tool's clinical and cost effectiveness.

The Vfrac tool has been designed as a web-based online tool, future-proofed so it will be supportable through NHS IT systems. The source code can be resurrected into any appropriate format such as a mobile website or an app depending on future IT infrastructure. We are now ready to assess real world clinical and cost-effectiveness of Vfrac to improve the detection of older adults with OVFs and improve bone health.

FIGURES

Figure 1: Statistical validation of Vfrac showing: (A) Receiver Operating Characteristic (ROC) curve illustrating the diagnostic ability of Vfrac to identify those with OVF. The Area under the curve (AUC) is 0.802, 95%CI 0.764 to 0.840; (B) a calibration plot over 10 risk groups defined by deciles of the linear predictor; and (C) the original linear predictor and the shrunken linear predictor.

Figure 2: Graphs showing the mean (heavy lighter grey line) and spread of the linear predictor for (A) those with no, one or more than one OVF, and (B) those with no, mild, moderate or severe OVFs.

REFERENCES

1. van Staa TP, Dennison EM, Leufkens HG, Cooper C. Epidemiology of fractures in England and Wales. Bone. Dec 2001;29(6):517-22.

2. Kaptoge S, Armbrecht G, Felsenberg D, et al. When should the doctor order a spine X-ray? Identifying vertebral fractures for osteoporosis care: results from the European Prospective Osteoporosis Study (EPOS). Journal of Bone and Mineral Research. 2004;19(12):1982-1993.

3. Ismail AA, Cockerill W, Cooper C, et al. Prevalent Vertebral Deformity Predicts Incident Hip though not distal Forearm Fracture: Results from the European Prospective Osteoporosis Study. Osteoporosis International. 2001;12(2):85-90.

4. Scane AC, Sutcliffe AM, Francis RM. The sequelae of vertebral crush fractures in men. Osteoporos Int. Mar 1994;4(2):89-92.

5. NICE. TA 464: bisphosphonates for treating osteoporosis. 2017;

6. Estell EG, Rosen CJ. Emerging insights into the comparative effectiveness of anabolic therapies for osteoporosis. Nature Reviews Endocrinology. 2021;17:31-46.

7. Cooper C, Atkinson EJ, O'Fallen WM, Melton Iii LJ. Incidence of clinically diagnosed vertebral fractures: a population-based study in Rochester, Minnesota. Journal of Bone and Mineral Research. 1992;7:221-227.

8. Elliott AM, Smith BH, Penny KI, Smith WC, Chambers WA. The epidemiology of chronic pain in the community. Lancet. 1999;354(9186):1248-1252.

9. Fechtenbaum J, Cropet C, Kolta S, Verdoncq B, Orcel P, Roux C. Reporting of vertebral fractures on spine X-rays. Osteoporos Int. 2005;16(12):1823-1826.

10. Lentle BC, Brown JP, Khan A, Leslie WD, Levesque J, Lyons DJ. Recognizing and reporting vertebral fractures: Reducing the risk of future osteoporotic fractures. Canadian Association of Radiologists Journal. 2007;58(1):27-36.

11. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions. wwwmrcacuk/complexinterventionsguidance. 2006;

12. Tobias JH, Hutchinson AP, Hunt LP, et al. Use of clinical risk factors to identify postmenopausal women with vertebral fractures. Osteoporosis International. 2007;18:35-43.

13. Clark EM, Gould VC, Morrison L, Ades AE, Dieppe P, Tobias JH. Randomized controlled trial of a primary care-based screening programme to identify older women with prevalent osteoporotic vertebral fractures. Journal of Bone and Mineral Research. 2012;27(3):664-671.

14. National Osteoporosis Society. Life with Osteoporosis: the untold story. Key findings from research into the realities of life with osteoporosis. 2014.

15. Clark EM, Gooberman-Hill R, Peters TJ. Using self-reports of pain and other variables to distinguish between older women with back pain due to vertebral fractures and those with pain due to degenerative change. Osteoporos Int. 2016;27:1459-1467.
16. Drew S, Clark EM, Al-Sari UA, Moore A, Gooberman-Hill R. Neglected bodily senses in women living with vertebral fractures: a focus group study. Rheumatology. 2019:DOI:10.1093/rheumatology/kez249.
17. Khera T, Burston A, Davis S, et al. An observational cohort study to produce and evaluate an improved tool to screen older women with back pain for osteoporotic vertebral fractures (Vfrac): Study Protocol. Archives of Osteoporosis. 2019;OSIN-D-18-00680R1
18. Melzack R. The McGill Pain Questionnaire. Major properties and scoring methods. Pain. 1975;1975(1):277-299.
19. Freynhagen R, Baron R, Gockel U, Tolle TR. PainDETECT: a new screening questionnaire to identify neuropathic components in patients with back pain. Curr Med Res Opin. 2006;22:1911-1920.
20. Hay EM, Dunn KM, Hill JC, et al. An RCT of subgrouping and targeted treatment for low back pain compared with best current care. The start back trail study protocol. BMC Musculoskeletal Disorders. 2008;9(58):doi:10.1186/1471-2474-9-58.
21. Margolis RB, Tait RC, Krause SJ. A rating system for use with patient pain drawings. Pain. 1986;24(1):57-65.
22. Janssen MF, Pickard S, Golicki D, et al. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. Qual Life Res. 2013;22(7):1717-1727.
23. Clark EM, Gould V, Morrison L, Masud T, Tobias JH. Determinants of fracture risk in a UK-population-based cohort of older women: A cross-sectional analysis of the Cohort for Skeletal Health in Bristol and Avon (COSHIBA). Age & Ageing. 2011;41(1):46-52.
24. Clark EM, Ness AR, Tobias JH. Bone Fragility Contributes to the Risk of Fracture in Children, Even After Moderate and Severe Trauma. Journal of Bone and Mineral Research. 2008;23(2):173-179.
25. Debouche S, Pitance L, Robert A, Liistro G, Reychler G. Reliability and reproducibility of chest wall expansion measurement in young healthy adults. J Manipulative Physiol Ther. 2016;39(6):443-449.
26. World Health Organization. Waist circumference and waist-hip ratio: report of a WHO expert consultation. 2008.
27. Siminoski K, Warshawski RS, Jen H, Lee KC. Accuracy of physical examination using the rib-pelvis distance for detection of lumbar vertebral fractures. The American Journal of Medicine. 2003;115(3):233-236.

28. Heuft-Dorenbosch L, Vosse D, Landewe R, et al. Measurement of spinal mobility in ankylosing spondylitis: comparison of occiput-to-wall and tragus-to-wall distance. J Rheumatol. 2004;31:1779-1784.
29. Jiang G, Eastell R, Barrington N, Ferrar L. Comparison of methods for the visual identification of prevalent vertebral fractures in osteoporosis. Osteoporosis International. MM 2004;15(11):887-896.
30. Genant HK, Wu CY, van Kuijk C, Nevitt MC. Vertebral fracture assessment using a semiquantitative technique. Journal of Bone and Mineral Research. 1993;8(9):1137-1148.
31. van Houwelingen JC, le Cessie S. Predictive value of statistical models. Statistics in Medicine. 1990;9(11):1303-1325.
32. O'Neill TW, Felsenberg D, Varlow J, Cooper C, Kanis JA, Silman AJ. The prevalence of vertebral deformity in European men and women: the European Vertebral Osteoporosis Study. J Bone Miner Res 1996;11:1010–1018
33. Davis S, Simpson E, Hamilton J, et al. Denosumab, raloxifene, romosozumab and teriparatide to prevent osteoporotic fragility fractures: a systematic review and economic evaluation. Health Technol Assess. 2020;24(29):1-314.
34. Welton NJ, Sutton AJ, Cooper NJ, Abrams KR, Ades AE. Evidence synthesis for decision making in healthcare. vol DOI: 10.1002/9781119942986. John Wiley & Sons, Ltd; 2012.
35. Gooberman-Hill R, Burston A, Clark EM, et al. Involving patients in research: Considering good practice. Musculoskeletal Care. 2013;11(4):187-190.
36. Melzack R. The short-form McGill Pain Questionnaire. Pain. 1987;30(2):191-197.
37. Schousboe JT, Rosen HR, Vokes TJ, et al. Prediction models of prevalent radiographic vertebral fractures among older women. Journal of Clinical Densitometry. 2014;17(3):378-385.
38. Watson SJ, Jones AL, Oatway WB, Hughes JS. Ionising radiation exposure of the UK population. 2005 Review. 2005;HPA-RPD-001

TABLES

Table 1: Final Vfrac model – Multivariable independent associations between variables and the presence or absence of vertebral fractures (n=1337; 163 with OVFs)

	Multivariable odds ratio per unit change in predictor (95%CI) n=1337	**Coefficient (SE), P**
Age (years)	0.98 (0.94-1.01)	-0.0239 (0.018), P=0.190
Weight (kg)	0.98 (0.96-0.99)	-0.0251 (0.007), P=0.001
Wall to tragus (cm)	1.07 (1.01-1.13)	0.0673 (0.029), P=0.021
Reported height loss (cm)	1.17 (1.10-1.25)	0.1568 (0.032), P<0.001
Pain described as sharp	0.63 (0.40-0.99)	-0.4615 (0.231), P=0.046
Pain described as like toothache	0.49 (0.27-0.91)	-0.7050 (0.311), P=0.024
Agreement with 'If I'm working in the kitchen like chopping vegetables of washing my back pain gets worse and worse to reach a peak – then I have to sit down immediately'	1.97 (1.30-3.00)	0.6799 (0.213), P=0.001
Pain in thoracic area of Margolis diagram	1.66 (1.11-2.49)	0.5073 (0.206), P=0.014
Pain in low back/buttock area of Margolis diagram	0.64 (0.44-0.94)	-0.4433 (0.196), P=0.024
Pain **increased** by walking	0.55 (0.37-0.84)	-0.5918 (0.210), P=0.005
Pain affected by sitting on straight-backed chairs	1.78 (1.16-2.74)	0.5779 (0.220), P=0.009
Pain affected by sitting on soft chairs	0.48 (0.32-0.71)	-0.7431 (0.201), P<0.001
Pain **increased** by reclining	1.93 (1.24-3.02)	0.6588 (0.228), P=0.004
Fracture after aged 50 excluding hands, feet, head and excluding high trauma	3.33 (2.30-4.82)	1.2021 (0.189), P<0.001
Steroids for >3 months	1.37 (0.81-2.32)	0.3124 (0.270), P=0.247
Constant		-1.9355 (1.456)

Predictor variables are put into the regression equation by stating with the coefficient, then the first four variables are entered in the units shown multiplied by their specific coefficient. The remainder variables are all coded zero or one, so the regression coefficients for these reflect the amount added or subtracted if the item is reported

Table 2: Table illustrating the effect of using a cut-point of -2.00 for the linear predictor

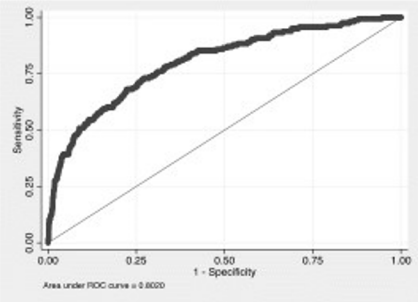
	Linear predictor		
Binary outcome	**< -2.00** **N (%)**	**≥ -2.00** **N (%)**	**total**
No VF	856 (72.9%)	318 (27.1%)	1174 (100%)
VF	45 (27.6%)	118 (72.4%)	163 (100%)
total	901 (67.4%)	436 (32.6%)	1337 (100%)
Number of OVFs	**< -2.00** **N (%)**	**≥ -2.00** **N (%)**	**total**
No VF	856 (72.9%)	318 (27.1%)	1174 (100%)
One VF	41 (38.0%)	67 (62.0%)	108 (100%)
More than one VF	4 (7.3%)	51 (92.7%)	55 (100%)
Severity of OVFs	**< -2.00** **N (%)**	**≥ -2.00** **N (%)**	**total**
No VF	856 (72.9%)	318 (27.1%)	1174 (100%)
Mild VFs	20 (41.7%)	28 (58.3%)	48 (100%)
Moderate VFs	21 (33.3%)	42 (66.7%)	63 (100%)
Severe VFs	4 (7.7%)	48 (92.3%)	52 (100%)

Table 3: Cost-effectiveness analyses (mean and 95% credible intervals)

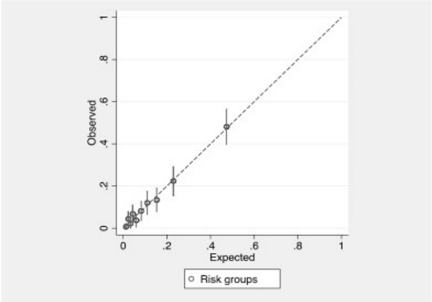
	Standard of care	Vfrac	Incremental Vfrac – standard of care
Proportion cohort with OVF and referred for radiograph	0.025 (0.012, 0.037)	0.091 (0.075, 0.11)	0.066 (0.046, 0.088)
Proportion cohort with OVF and not referred for radiograph	0.10 (0.081, 0.12)	0.034 (0.024, 0.044)	-0.066 (-0.089, -0.046)
Proportion cohort with no OVF but referred for radiograph	0.17 (0.087, 0.25)	0.25 (0.22, 0.27)	0.072 (-0.012, 0.16)
Total costs (£)	315.67 (267.65, 370.99)	322.95 (274.12, 375.08)	7.28 (-58.59, 73.04)
Total QALYs	0.63 (0.53, 0.73)	0.63 (0.53, 0.73)	0.00044 (-0.13, 0.13)
Net benefit (£, at £20,000/QALY)	12,192 (10227, 14208)	12193 (10344, 14209)	1.47 (-2587, 2456)
Probability of Cost Effectiveness	0.506	0.494	NA
EVPI (£)	526		
Population EVPI (£)	229-458 million		

Figure 1

A


Sensitivity
0.00
0.25
0.50
0.75
1.00
0.00
0.25
0.50
0.75
1.00
1 - Specificity
Area under ROC curve = 0.8020

B


Observed
0
.2
.4
.6
.8
1
0
.2
.4
.6
.8
1
Expected
Risk groups

C

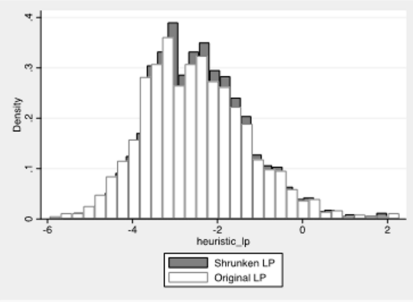
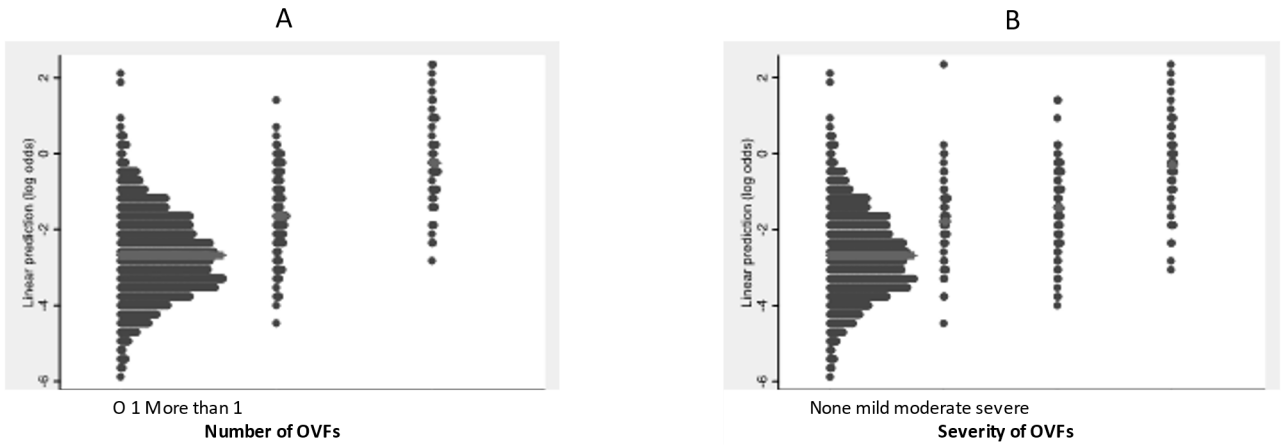

Density
0
.1
.2
.3
.4
-6
-4
-2
0
2
heuristic_lp
Shrunken LP
Original LP

Figure 2


A
Linear prediction (log odds)
2
0
-2
-4
-6
O 1 More than 1
Number of OVFs
B
Linear prediction (log odds)
2
0
-2
-4
-6
None mild moderate severe
Severity of OVFs

Supplementary Data for Vfrac: A clinical tool to identify older women with back pain at high risk of osteoporotic vertebral fractures (Vfrac): a population-based cohort study with exploratory economic evaluation

Contents

SECTION 1: SUPPLEMENTARY METHODS

Study design and participants

Thirteen general practices from a range of neighbourhoods and deprivation scores as assessed by the Index of Multiple Deprivation (IMD) in 2015, were recruited from Stoke-on-Trent and 9 from Bristol. Practices that took part in the original preparatory[12] and pilot studies[13,15] were not recruited. Practices invited all eligible women registered on their system to take part by post between July 2018 and January 2020. Eligibility was solely age 65 or more. Those who reported an episode of back pain in the previous four months completed further data collection through physical examination and spinal radiograph – see Figure 1. Participant ethnicity was compared with the UK census from 2011, where 86.0% were white and 14.0% a mix of Asian, Black, Mixed and other. In the Vfrac study, 78.5% self-reported as white, 20.7% declined to answer, and 0.9% self-reported as non-white, indicating that the cohort has a shortfall in non-white participants.

Exposure data

The majority of exposure data was self-reported; similar studies using identical data collection methods have generally good agreement with electronic GP records[13]. Specifically for this study, the self-reported co-morbidity data were compared against electronic GP records and results showed good agreement for some measures (e.g. self-reported diabetes 95.8% confirmed), but less for others (e.g. 66.0% for chronic lung disease).

Back pain data: A wide range of questions were included in the self-completion questionnaire (see protocol paper for full description[17]) based on previous studies on women with and without OVFs[15,18] plus other back pain questionnaires[19,20]. Findings from the qualitative study[16] were also used to develop questions for quantitative data collection that asked whether participants agreed or disagreed with various statements about how back pain changed with activity and other descriptive statements. The Margolis pain diagram[21] was included to allow study participants to mark the anatomical site of their back pain, and as used in previous studies investigating OVFs[13], marks in thoracic area, waist area and low back/buttock area were used in this analysis. Where participants were asked to indicate whether they agreed or disagreed whether an activity or posture made their back pain worse but neither had been ticked, these types of missing data were recoded as 'disagree'. The implications of this were to reduce the strength of any association seen for univariable associations but increase power overall by reducing missing data.

Other self-reported data: Data were collected on frailty (walking aid use, walking distance, falls, concomitant illnesses), traditional risk factors for osteoporosis (previous fractures, use of steroid tablets, family history of hip fracture, smoking, and alcohol intake), concomitant illnesses (including diabetes, inflammatory arthritis, anxiety and heart disease), health related quality of life (EQ5D-5L[22]), healthcare usage at baseline and three months later (interactions with healthcare professionals in primary and secondary care) and use of pain killing medication at baseline and three months later using the same question structure as previous research studies[15,23]. Previous fractures were categorised according to age at fracture, anatomical site of fracture and level of trauma of the precipitating injury[24] if known. A pragmatic decision was taken, based on ease of data collection/recollection whilst attempting to capture previous fragility fractures, to use the following definition: fracture after aged 50 or over excluding hands, feet, head and excluding high trauma (falls from more than 3 metres, car accidents, being hit by a heavy moving object or crushed in a machine).

Physical examination: Data were collected during the research clinic on all participants, by trained research nurses, chosen from literature review and our previous research[12,13] and were: (1) height without shoes and without headgear measured with a free-standing stadiometer; (2) weight without shoes, hats, coats and cardigans/jumpers on a calibrated weighing machine; (3) chest expansion using a validated method[25]; (4) waist circumference using the WHO method[26]; (5) rib-to-pelvis distance using a validated method[27]; and (6) wall-tragus distance using a validated method[28]. Reported height loss was calculated by subtracting the height measured in the research clinic from that recorded on the baseline questionnaire (self-reported height at 25 years of age). Those whose measured height was taller than self-reported height at age 25 years were censored so reported height loss was recorded as zero.

Outcome data: Osteoporotic vertebral fractures (OVFs)

All participants had lateral thoracic and lateral lumbar radiographs. Antero-posterior (AP) views were not performed for pragmatic reasons to reduce radiation exposure and financial cost. Radiographs were assessed for the presence or absence of OVFs by an experienced clinical researcher (EC) using the Algorithm-Based Qualitative (ABQ) method[29]. Radiographs were categorised into those with no fracture and those with fracture. Site and number of fractures were also noted. As a secondary outcome, those with OVFs were further categorised into mild, moderate or severe fractures based on their 'worst' fracture using the Genant semi-quantitative (SQ) method[30]. Repeatability of the primary outcome was assessed by a random sample of anonymised images reviewed by EC and an independent experienced radiologist (SG) 4 months after completion of initial data collection. Neither EC nor SG knew the previously assigned categorisation. Results showed complete agreement for intra-rater reliability by EC. The kappa for agreement between EC and SG

(inter-rater reliability) was 0.689 indicating substantial agreement. There was 100% agreement between EC and SG for moderate and severe OVFs. EC under-diagnosed mild OVFs compared to SG.

Statistical analysis

Preliminary univariable analysis explored the relationships between each (categorical or continuous) predictor variable in turn using logistic regression. Variables found related to OVF at this stage with P<0.1 were taken forward to the next stage of the analysis. For this, a series of logistic regression models were carried out using subsets of the predictor variables in turn; this pragmatic approach was adopted in part as there was a large number of predictors of which many had missing values and complete case analysis at this stage would have led to a much reduced data set. Moreover, this approach allows greater interpretive control over variable selection, including reducing the risks of multi-collinearity going unnoticed, as well as generally increased epidemiological insights to potentially be gleaned. Groups of predictor variables were considered together using backwards stepwise logistic regression analyses to remove those with P>0.1. Continuous predictor variables were further investigated as to whether the model could be improved by their prior transformation using fractional polynomials up to the second degree. Age was constrained to stay in the model, irrespective of its P value. The reduced subsets of predictor variables were then combined and analysed with a similar backwards stepwise approach. Having determined a 'final' model, the discarded predictors were added back individually to check that none would further improve the model, and those that did were added back using likelihood ratio tests to assess improvement. Regression coefficients needed to calculate the linear predictor, the maximum likelihood R-squared, Brier's score and AUC calculated are reported for the final model obtained. Model validation included calibration-in-the-large (CITL), calibration slope and heuristic shrinkage[31]. Five hundred bootstrapped samples were created and used to estimate shrinkage and adjust the calibration slope and AUC optimism. As the final model was calculated from complete cases, 10 multiply imputed data sets were combined to re-estimate the regression coefficients on the full set of 1,601. Multiple imputation assumed missing at random (MAR). Subgroup analyses were performed to look at results from the final model separately for the two participating centres. A cut-point of the final linear predictor was identified based on a maximised sum of sensitivity and specificity. The added benefit of the use of self-reported symptoms was assessed by looking at the proportions of those identified with OVFs using the cut-point before and after removal of these symptoms. All analyses were carried out using Stata 16.0.

Sample size

Full details are available in the protocol paper[17]. The sample size was calculated as 1633, based on the following assumptions: a prevalence of OVFs between 12-20% based on data from the European Vertebral Osteoporosis Study[32]; a margin of error of 5%, and sensitivity and specificity of the Vfrac tool between 80 and 95%. 1633 will be large enough to encompass and specificity of Vfrac as sample sizes required for specificity are much lower.

Health economic analyses

Within study analysis: In Supplementary Data, Section 2, pg 7 of this document, it is shown that more severe patients are referred or consult with their GP and patients always have higher costs and worse EQ-5D profiles at follow-up, regardless of referral or GP consultation. This finding made it difficult to use the Vfrac study data to construct a counterfactual analysis to show any benefit of Vfrac referral for OVF or GP consultation for back pain. Modelling was therefore instead used for the health economic analysis.

Stakeholder work: This was undertaken to provide a description of current standard of care and to sense check the baseline data. An online survey was sent to clinicians in the field of osteoporosis and primary care, and to patients written in plain language. Results were anonymous, and questions covered estimations of healthcare usage including consultation rates for back pain and getting a spinal radiograph in the presence of back pain, as well as sense-checking the baseline data collected in the study for healthcare use. Seven clinicians and 12 patients took part.

Health economic modelling: The decision tree structure used for modelling is illustrated in Supplementary Data, Section 3. This decision tree was designed in discussion with our clinical team and reflecting on findings of the stakeholder work. The model reflects the consensus that patients with suspected OVF are either referred or not referred for radiograph and that patients diagnosed by radiograph as having OVF are always assigned to treatment. Current standard of care was defined from stakeholder work as consultation with GP for back pain followed by potential referral for radiograph. To compare the cost-effectiveness of the Vfrac tool to this standard of care, the proportions of people diagnosed with OVF by the Vfrac tool and by current standard of care were modelled, as were the life-time costs and quality-adjusted-life-years (QALYs). Model parameters are described here briefly but full details are provided in Supplementary Data, Section 4, Supplementary Tables 1 and 2 on pg. For patients receiving standard-of-care, we assumed all patients to have a GP consultation with the proportion referred being 20% with a 95% credible range of 10-30%, modelled by a Normal distribution, following the stakeholder work. For patients receiving Vfrac, it was important to distinguish not only whether the patients had an OVF, but also whether patients would be recommended for radiograph by the Vfrac tool. Proportions with OVF referred for radiograph, with OVF but not referred for radiograph and without OVF but referred for radiograph were modelled with a Dirichlet distribution with parameters equal to the numbers of patients in these categories in the Vfrac

study. Uncertainty was modelled using 1000 samples from these distributions. The cost of radiograph (£72) was taken from standard NHS costs and no radiograph disutility was applied.

Life-time costs and QALYs for OVF diagnosed (i.e. treated with standard bisphosphonate anti-fracture medication) and OVF undiagnosed (i.e. untreated) were obtained using a previously published bisphosphonate cost-effectiveness model[33]. This long-term model was developed to inform appraisals of osteoporosis treatments by the UK's National Institute for Health and Care Excellence and therefore the methods used comply with the reference case for cost-effectiveness analysis in the UK. It uses a discrete event simulation (DES) framework. The key clinical events modelled are fractures at the hip, vertebrae, wrist or proximal humerus, all-cause mortality, fracture-related mortality, and new admission to long-term residential care following hip fracture. Costs are estimated from an NHS and Personal Social Services perspective and QALYs are estimated using utility values derived from the UK valuation of the EQ-5D. Patient characteristics for simulated patients were set to match the distribution in the 118 patients in the study with a positive Vfrac score and an OVF. These had an average age of 76 years (SD 6.3). Repeated resampling from this cohort of 118 yielded a cohort of 50,000 patients. The model was run 1000 times for this cohort of 50,000 using different parameter samples each time, averaging the heterogeneity. Simulations were used to estimate expected lifetime costs and QALYs according to whether the individual received treatment with the bisphosphonate alendronic acid, or no treatment. Lifetime costs and QALYs from this model are summarised in Supplementary Table 2.

Total costs and QALYs for Vfrac and standard of care were calculated and summarized with mean and Bayesian 95% credible intervals (CrI); these are the Bayesian equivalent of frequentist confidence intervals. For both life-term Vfrac and standard of care groups, net benefits were calculated at a willingness to pay threshold of £20,000/QALY, the current standard for UK healthcare decision making. These were used to calculate the probability that Vfrac or standard of care was most cost-effective (i.e. intervention with greatest net benefit at £20,000/QALY). Expected Value of Perfect Information (EVPI) per person and population EVPI were estimated[34], to measure the value of removing all uncertainty in all parameters (i.e. those listed in Supplementary Table 1 and the life-term costs and QALYs from the bisphosphonate model in Supplementary Table 2). Population EVPI is an upper bound on the value of a randomized controlled trial on the efficacy of the Vfrac diagnostic tool. A time-horizon for the population EVPI of 10 years was applied with discounting at 3.5% per year. The population was that of women aged ≥65 years in the UK consulting with their GP for back pain (further details in Supplementary Table 1).

The key assumptions of the health economic analysis are:

1. The consultation time for nurses is an average of 9.72 minutes.
2. For costing painkillers, if the patient takes painkillers on a few days/week we assume their medication cost is the weighted average daily drug cost multiplied by a random number drawn from a normal distribution ranging from 1 to 3.5 (i.e. incurs day of drug costs 1-3.5 days a week); if the patient takes painkillers at least once/twice per day, we assume their medication cost is the weighted average daily drug cost multiplied by a random number drawn from a normal distribution ranging from 3.5 to 7 (i.e. incurs daily drug cost 3.5 to 7 days per week). These weekly frequencies were multiplied by average daily cost to get 7-day costs and then, assuming 30-day months. (shown in supplementary data table 3)
3. For patients receiving standard of care, we assumed all patients to have a GP consultation with the proportion referred being 20% with a 95% credible range of 10-30%, modelled by a Normal distribution, following the stakeholder work.
4. Distribution assumption in the proportion of different patient types for the Vfrac health economic modelling (Section 5: SUPPLEMENTARY TABLE: Supplementary Table 1.)

Patient and Public Involvement

Patients and the public have been involved since 2012 in planning and delivery of the research projects that led up to development of the Vfrac tool. Two people with OVFs provided input into the steering committee of the study. Our award-winning Patient Experience Partnership in Research (PEP-R)[35] helped to write the patient/public facing paperwork including Patient Information Leaflets for recruitment, an infographic for communication of results, and plain language summaries for the next stages of Vfrac testing.

SECTION 2: WITHIN-STUDY HEALTH ECONOMIC (HE) ANALYSIS

In this appendix we explain why it is not possible to use the Vfrac data to construct a counterfactual analysis to show any benefit of Vfrac diagnosis or GP consultation, as more severe patients are referred or consult with their GP and patients always become worse at follow-up regardless of referral or consultation.

METHODS

We describe here the methods used to calculate the total costs and quality adjusted life years (QALYs) for patients with and without osteoporotic vertebral fracture (OVF) who would or would not have been referred for radiograph, and therefore diagnosed, by the Vfrac tool. We also describe methods to calculate these costs and QALYS for a hypothetical standard of care consisting only of GP consultations of patients with back-pain for lateral thoracic and lateral lumbar radiographs.

Through the Vfrac programme, 1635 older women with back pain in the previous 4 months were recruited from the community and data collection was based on self-reporting and physical examination. The data available for each patient includes clinical characteristics (e.g. whether or not radiograph shows an OVF), whether they would be referred for radiograph by the Vfrac tool, whether they had a GP consultation for back-pain in the past 3 months, resource use and Euroqol 5-dimensions questionnaire 5-level (EQ5D-5L) at baseline, plus 3 month follow up data on resource use and EQ5D-5L. The analysed sample consisted of only 1600 women as 35 patients did not report baseline EQ5D-5L or medication data.

Resource use

Resource use (i.e. costs) recorded were medical professional consultations (i.e. healthcare usage) and painkiller medication taken.

For consultations, we multiplied the unit cost of each consultation by the number of those consultations per patient over a three-month period in the questionnaire, with the sum over all types giving the cost of consultations. Unit costs were obtained from the Personal Social Services Research Unit (PSSRU) 2020 and the NHS website, and these are summarised in HE Table 1. For nurse consultations, consultation time is required. [1, 2] A recent study reported this as being on average 9.72 minutes. [3] We assume that the consultation time for nurse types is an average of 9.72 minutes. In addition, the unit costs of Osteopath and Chiropractor are zero from the NHS perspective as neither service is covered by the NHS.

HE Table 1 Unit costs for different types of consultation and data sources.

Consultation type	Unit cost (£)	Data sources
GP	28	Unit costs health and social care 2020
Nurse	6.156	£38/hour from PSSRU and average 9.72min/consultation from Hobbs, et al. (2016)
Hospital specialist	117	Unit costs health and social care 2020
Physio	55	Unit costs health and social care 2020
Osteopath	0	Not reimbursed by the NHS
Chiropractor	0	Not reimbursed by the NHS
OT	106	Unit costs health and social care 2020
Orthotics	6.156	unit costs health and social care 2020
Pharmacist	0	Community pharmacists are not paid for by the NHS.

Daily medication costs came from the British National Formulary (BNF).[4] Medication costs were a weighted average of daily drug costs, with weights based on the proportion of patients receiving each type of medication in the Vfrac study shows in HE Table 2**.**

HE Table 2 Daily cost of drugs and weighted average daily drug costs from Vfrac study data.

Drugs types	Quantity	Quantity Unit	Assumption on daily dose	Daily Costs(£)	Proportion (n=1634*)
Paracetamol 500mg tablets	100	Tablet	Full standard daily dose	0.0536	36.2%
Ibuprofen 400mg tablets	168	Tablet	Full standard daily dose	0.968	12.1%
Co-codamol 15mg/500mg tablets	100	Tablet	Full standard daily dose	0.074	8.7%
Codeine 30mg tablets	100	Tablet	50% standard daily dose	0.3512	8.7%
Amitriptyline 10mg tablets	56	Tablet	10mg per day	0.0461	7.1%
Citalopram 20mg tablets	28	Tablet	Full standard daily dose	0.0257	6.1%
Aspirin 75mg gastro-resistant tablets	28	Tablet	Full standard daily dose	0.0429	5.3%
Ibuprofen 10% gel	100	Gram	A tube lasts a month /31 days a month	0.1226	5.1%
Weighted average daily drug costs	£0.187				

*Sample size for prescription costs was 1634 as one patient did not complete baseline questionnaire. The 1600 patients used in health economic analysis reported both baseline resource use and EQ5D-5L.

Total cost of medication was calculated using reported frequency of taking painkiller medication. Patients were asked "How often have you taken pain killers for your back pain in the last week?" and given four options: 0=I do not take painkillers, 1= on a few days/week, 2=once/twice per day, 3=3-4 times per day, and 4=4-6 times per day. We categorised these responses into numbers of painkillers taken per week as explained in HE Table 3. If the patient does not take painkillers, then their medication cost is 0; if the patient takes painkillers on a few days/week, then their medication cost is the weighted average daily drug cost multiplied by a random number from a normal distribution ranging from 1 to 3.5 (i.e. incurs day of drug costs 1-3.5 days a week); if the

patient takes painkillers at least once/twice per day, then their medication cost is the weighted average daily drug cost multiplied by a random number from a normal distribution ranging from 3.5 to 7 (i.e. incurs daily drug cost 3.5 to 7 days per week). These weekly frequencies were multiplied by average daily cost (HE Table 2) to get 7-day costs and then, assuming 30-day months, scaled to 3-month medication costs.

HE Table 3 Assumption on weekly drug does costs calculation.

Response category*	Assumed frequency	Distribution
0	0	0
1	Incurs daily drug costs 1 and 3.5 days per week.	Normal mean 2.25 and standard deviation 0.638
≥2	Incurs daily drug costs 3.5 and 7 days per week.	Normal mean 5.25 and standard deviation 0.638

*0=I do not take painkillers, 1= on a few days/week, 2=once/twice per day, 3=3-4 times per day, and 4=4-6 times per day

EQ5D-5L score and QALYs

Baseline and 3-month follow-up EQ5D-5L questionnaire were completed by patients in the Vfrac study. In this case, these were converted to utility scores using the EuroQoL value set dated 19th January 2019.[5] Baseline and follow-up quality adjusted life years (QALYs) were calculated by multiplying their EQ5D-5L score by 0.25 (i.e. 3 months).

Total costs and QALYs on Vfrac and those with GP consultation

Whether a patient would be referred for lateral thoracic and lateral lumbar radiographs by the Vfrac tool and whether they had an OVF was recorded. We could therefore tabulate total cost and QALYs across these categories of patients.

It was also recorded if a patient had a GP consultation in the past 3 months for back pain. The study did not record if a patient was referred for radiograph by GP consultation in the past 3 months so this could not be used as a hypothetical "standard of care" comparator for the Vfrac tool. For the long-term analysis (as described in the main text), we used stakeholder work on the proportion of patients consulting with a GP for back pain who would be referred for radiograph. In the tabulation of total cost and QALYs using Vfrac data, we only categorise patients based on having or not having a GP consultation for backpain, and do not model the hypothetical proportion referred for radiograph.

RESULTS

The estimated costs and QALYs are provided in HE Table 4 and HE Table 5, respectively. The key finding is that costs are always higher at follow-up than at baseline and QALYs are lower at follow-up than at baseline for all categories of patient except those without GP consultation and without OVF. In patients with OVF, those with GP consultation have higher costs and lower QALYs than those without, indicating that more severe patients go to their GPs.

Similarly, patients who would have been referred for radiograph by Vfrac have higher costs and lower QALYs than those who would not have been referred, suggesting the Vfrac tool picks up more severe patients. The consequence of these findings is that it is not possible to use the Vfrac data to construct a counterfactual analysis to show any benefit of Vfrac diagnosis or GP consultation, as more severe patients are referred or consult with their GP and patients always become worse at follow-up regardless of referral or consultation.

*HE Table 4 Mean costs (£) (95% reference range) using Vfrac study data in patients with or without GP consultation, hypothetically referred for radiograph by Vfrac or not, and with or without osteoporotic vertebral fracture (OVF)**

	OVF		No OVF	
	Baseline	**Follow-up**	**Baseline**	**Follow-up**
	£ cost (95% ref range)	**£ cost (95% ref range)**	**£ cost (95% ref range)**	**£ cost (95% ref range)**
All patients	90.58 (0.00, 377.85)	95.61 (0.00, 383.76)	73.85 (0.00, 413.60)	79.11 (0.00, 422.35)
Patients with GP consultation	136.53(0.92, 380.11)	143.26 (11.89, 384.38)	72.85 (14.50, 145.53)	72.85 (14.50, 145.53)
Patients without GP consultation	85.96 (0.00, 363.94)	90.91 (0.00, 368.10)	87.15 (0.00, 366.99)	92.13 (0.00, 377.58)
Patients who would have been referred for radiograph by Vfrac	94.01 (0.00, 383.18)	98.49 (0.00, 395.71)	119.63 (0.00, 775.80)	124.88 (0.00, 786.92)
Patients who would not have been referred radiograph by Vfrac	80.56 (0.00, 390.94)	84.69 (0.00, 391.28)	85.28 (0.00, 365.51)	90.29 (0.00, 371.95)

*95% reference range is an interval containing 95% of samples.

*Table 5 Mean QALYs (95% reference range) using Vfrac study data in patients with or without GP consultation, hypothetically referred for radiograph by Vfrac or not, and with or without osteoporotic vertebral fracture (OVF)**

	OVF		No OVF	
	Baseline	**Follow-up**	**Baseline**	**Follow-up**
	Cost (95% ref range)	**Cost (95% ref range)**	**Cost (95% ref range)**	**Cost (95% ref range)**
All patients	0.15 (0.02, 0.25)	0.14 (0.01, 0.24)	0.16 (0.03, 0.23)	0.16 (0.05, 0.25)
Patients with GP consultation	0.13 (0.02, 0.24)	0.12 (0.02, 0.25)	0.19 (0.15, 0.21)	0.12 (0.01, 0.25)
Patients without GP consultation	0.16 (0.04, 0.25)	0.15 (0.02, 0.25)	0.15 (0.01, 0.25)	0.15 (0.01, 0.25)
Patients who would have been referred for radiograph by Vfrac	0.16 (0.04, 0.25)	0.15 (0.04, 0.23)	0.15 (0.04, 0.24)	0.14 (0.04, 0.20)
Patients who would not have been referred radiograph by Vfrac	0.17 (0.00, 0.22)	0.16 (0.00, 0.25)	0.15 (0.00, 0.23)	0.14 (0.00, 0.23)

*95% reference range is an interval containing 95% of samples.

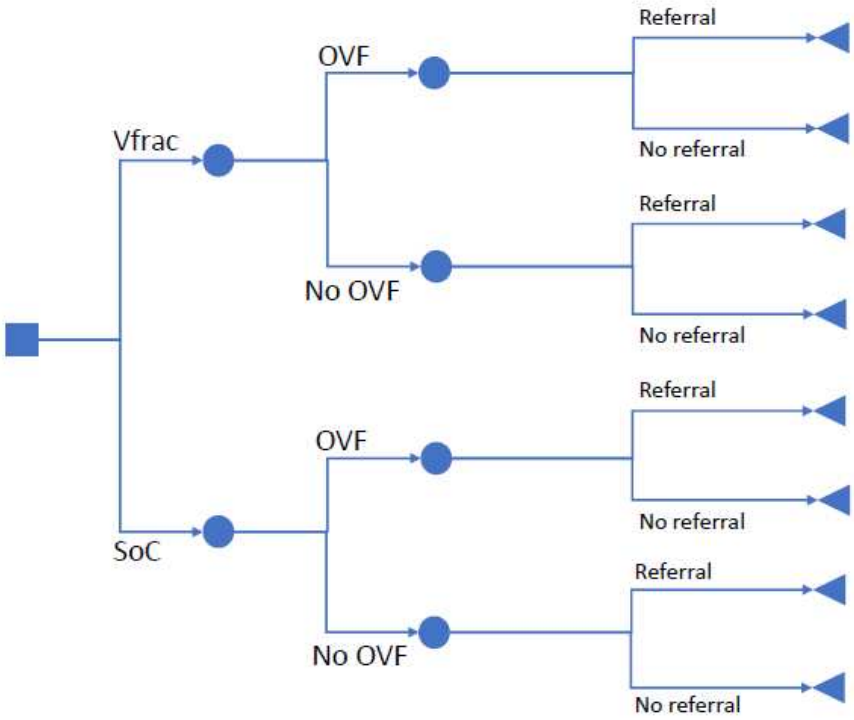
References

1. NHS. NHS Choices. www.nhs.uk/conditions/
2. PSSRU. Unit Costs of Health and Social Care 2020. www.pssru.ac.uk/project-pages/unit-costs/unit-costs-2020/
3. Hobbs FDR, Bankhead C, Mukhtar T, et al. Clinical workload in UK primary care: a retrospective analysis of 100 million consultations in England, 2007-14. *Lancet*. Jun 4 2016;387(10035):2323-2330. doi:10.1016/s0140-6736(16)00620-6
4. BNF. drug. https://bnf.nice.org.uk/drug/
5. EQ-5D. EuroQoL Value Set 19Jan2019. euroqol.org/eq-5d-instruments/eq-5d-5l-about/valuation-standard-value-sets/

SECTION 3: SUPPLEMENTARY FIGURE

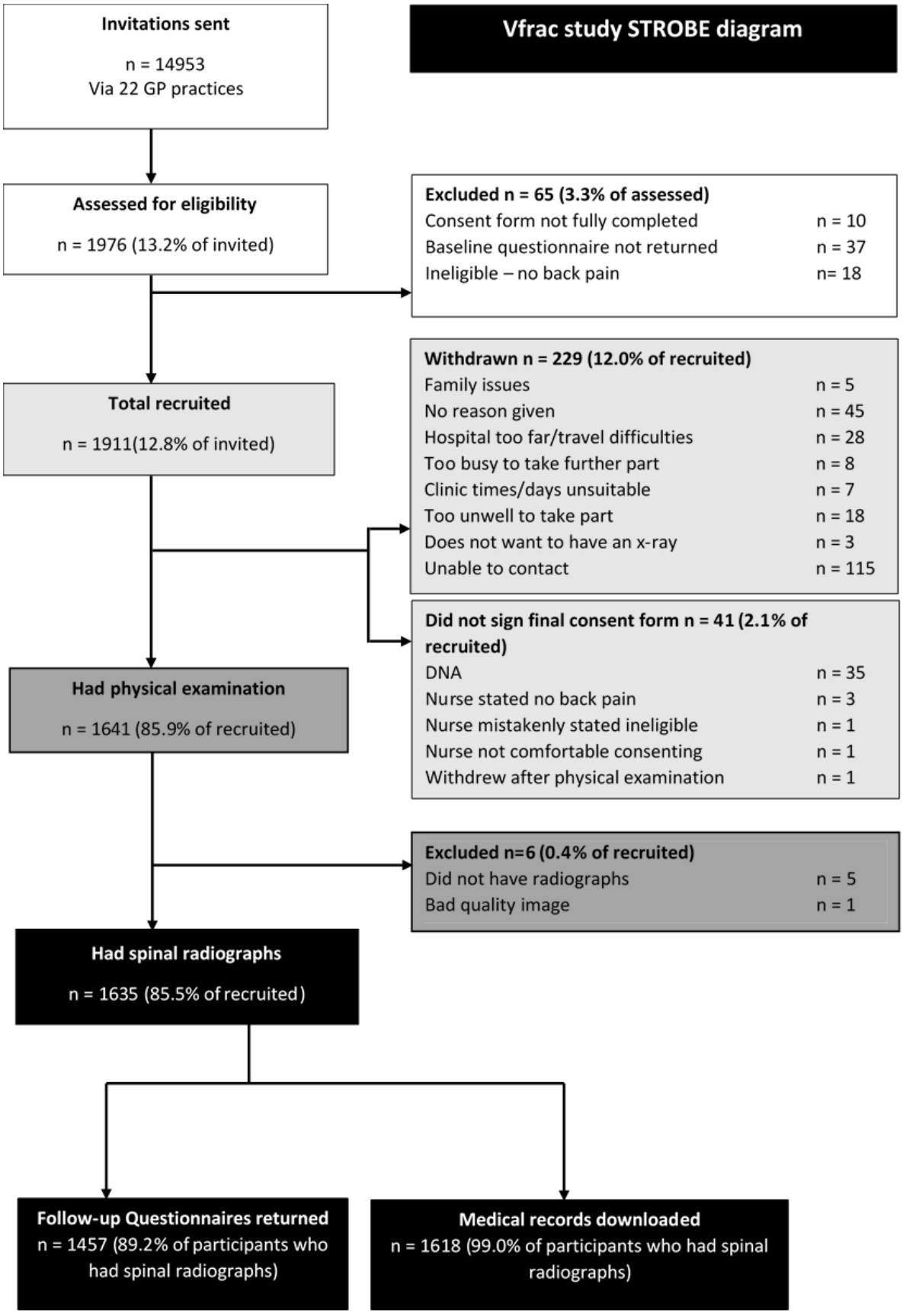
Supplementary Figure: Illustration of decision tree model used for health economic comparison of Vfrac and Standard of Care (SoC)*

* Cohort is assigned to either Vfrac or SoC. They either have underlying osteoporotic vertebral fracture (OVF) or not. Vfrac and SoC have some probability of referring patients for radiograph; for Vfrac these probabilities come from the Vfrac study while for SoC these come from stakeholder work. All referred patients incur a radiograph cost of £72. OVF patients who are referred receive alendronate/bisphosphonate while OVF patients without referral receive no treatment. A lifetime horizon for patients with starting age 76 (chosen to match study data) was adopted. Lifetime QALYs and costs for OVF patients on alendronate or no treatment come from a previously published bisphosphonate model. Patients without OVF do not incur costs or QALYs beyond the radiograph and are assumed zero for calculation (this has no impact on incremental results and thus conclusions). Mean costs and QALYs are presented only for illustration as the model is fully probabilistic.



Final health condition	QALYs (longterm only)	Costs	
		Longterm	Radiograph
OVF, treatment	5.041	2394.18	72.00
OVF, no treatment	5.037	2436.43	0.00
No OVF	0.00	0.00	72.00
No OVF	0.00	0.00	0.00
OVF, treatment	5.041	2394.18	72.00
OVF, no treatment	5.037	2436.43	0.00
No OVF	0.00	0.00	72.00
No OVF	0.00	0.00	0.00

SECTION 4: VFRAC STROBE DIAGRAM



SECTION 5: SUPPLEMENTARY TABLES

Supplementary Table 1: Distribution assumption in the proportion of different patient types for the Vfrac health economic modelling

Parameter	Distribution	Mean and 95% credible interval
Proportion of patients referred for radiograph following GP consultation	Normal (mean=0.2, standard deviation=0.05)	0.2 (0.11, 0.30)
Proportion of patients with OVF	Beta (147, 1183)*	0.11 (0.09, 0.13)
Proportion of patients with OVF referred for radiograph by Vfrac tool	First element of Dirichlet (107, 40, 290, 746)**	0.091 (0.075 ,0.11)
Proportion of patients with OVF not referred for radiograph by Vfrac tool	Second element of Dirichlet (107, 40, 290, 746)**	0.034 (0.024 ,0.044)
The proportion patients without OVF referred for radiograph by Vfrac tool	Third element of Dirichlet (107, 40, 290, 746)**	0.25 (0.22 ,0.27)
Size of population to benefit for value of information analysis population		
Number of people aged 65+ in England and Wales[1]	12,390,000	
Percentage female of population 65+ in England and Wales[2]	51%	
Prevalence of back pain consultation among women aged ≥65 years[3]	800 to 1600 per 100,000	
Total population over 10 years discounted at 3.5%	435,129 to 870,258	
	costs	disutility
Radiograph§	£72	0

*Alpha and beta of the beta distribution given by numbers of patients in each category in Vfrac study data.

** Parameters given by numbers of patients in each category in Vfrac study data, with common Dirichlet distribution across categories to capture correlation. 1183 patients in total with 746 remaining in category without OVF and without referral by Vfrac.

§ Standard NHS costs

Supplementary Table 2: Cost and QALYs estimated by the bisphosphonate model. (Mean with 95% credible interval)

	Costs (£)	**QALYs**
No treatment	2436.43 (2277.28, 2601.36)	5.037 (4.882, 5.186)
Alendronate	2394.18 (2225.38, 2597.55)	5.041 (4.884, 5.189)
Incremental Alendronate - no treatment	-42.26 (-123.14, 86.23)	0.00405 (0.000152, 0.00748)

Supplementary Table 3: Back pain descriptives (words) based on McGill pain questionnaire and qualitative focus group work used by participants according to presence of any OVF (n=1596)

Word	Word not used		Word used		(Univariable) Odds ratio (95%CI) for 'word used' (vs 'not used' as ref.)
	Number not using	Number (%) with OVF	Number using	Number (%) with OVF	
Crushing	1,513	190 (12.6%)	83	12 (14.5%)	1.18 (0.63-2.21) P=0.613
Heavy	1,350	168 (12.4%)	246	34 (13.8%)	1.13 (0.76-1.68) P=0.551
Dull	1,016	141 (13.9%)	580	61 (10.5%)	0.73 (0.53-1.00) P=0.053
Aching	289	35 (12.1%)	1,307	167 (12.8%)	1.06 (0.72-1.57) P=0.758
Sharp	1,158	168 (14.5%)	438	34 (7.8%)	0.50 (0.34-0.73) P<0.001
Agonizing	1,354	171 (12.6%)	242	31 (12.8%)	1.02 (0.67-1.53) P=0.938
Unbearable	1,399	171 (12.2%)	197	31 (15.7%)	1.34 (0.89-2.03) P=0.166
Gnawing	1,048	135 (12.9%)	548	67 (12.2%)	0.94 (0.69-1.29) P=0.708
Excruciating	1,444	184 (12.7%)	152	18 (11.8%)	0.92 (0.55-1.54) P=0.751
Intense	1,170	151 (12.9%)	426	51 (12.0%)	0.92 (0.65-1.29) P=0.620
Stinging	1,526	189 (12.4%)	70	13 (18.6%)	1.61 (0.87-3.00) P=0.131
Cold	1,567	201 (12.8%)	29	1 (3.5%)	0.24 (0.03-1.79) P=0.165
Tingle	1,491	189 (12.7%)	105	13 (12.4%)	0.97 (0.53-1.77) P=0.930
Icy	1,588	201 (12.7%)	8	1 (12.5%)	0.99 (0.12-8.05) P=0.989
Toothache	1,328	182 (13.7%)	268	20 (7.5%)	0.51 (0.31-0.82) P=0.006
Burning	1,375	174 (12.7%)	221	28 (12.7%)	1.00 (0.65-1.53) P=0.995
Brief	1,504	191 (12.7%)	92	11 (12.0%)	0.93 (0.49-1.78) P=0.835
Tiring	826	98 (11.9%)	770	104 (13.5%)	1.16 (0.86-1.56) P=0.324
Niggling	1,035	133 (12.9%)	561	69 (12.3%)	0.95 (0.70-1.30) P=0.752
Continuous	1,095	142 (13.0%)	501	60 (12.0%)	0.91 (0.66-1.26) P=0.580
Annoying	922	128 (13.9%)	674	74 (11.0%)	0.77 (0.56-1.04) P=0.085
Radiating	1,274	164 (12.9%)	322	38 (11.8%)	0.91 (0.62-1.32) P=0.605
Pins&needles	1,420	184 (13.0%)	176	18 (10.2%)	0.77 (0.46-1.28) P=0.305

Supplementary Table 4: Multivariable relationships between back pain descriptives and presence or absence of OVF (n=1596)

Word	Odds ratio (95%CI) for 'word used' (vs 'not used' as ref.)
Dull	0.72 (0.52-0.99) P=0.044
Sharp	0.48 (0.32-0.70) P<0.001
Stinging	1.86 (0.99-3.52) P=0.056
Toothache	0.55 (0.34-0.89) P=0.016

Supplementary Table 5: Variables describing change in back pain with specific activities and the presence or absence of OVFs (n=1601)

	Disagree/ Didn't indicate		Agree		(Univariable) Odds ratio (95%CI) (vs 'disagree' as ref.)
	Total n	Number (%) with OVF	Total n	Number (%) with OVF	
When I start doing an activity the pain builds, and builds until it's agony and I have to stop	801	85 (10.6%)	800	117 (14.6%)	1.44 (1.07-1.94), P=0.016
Walking generally makes my back pain better whilst I'm walking	950	128 (13.5%)	651	74 (11.4%)	0.82 (0.61-1.12), P=0.213
Walking generally makes my back pain better once I've finished walking	1139	148 (13.0%)	462	54 (11.7%)	0.89 (0.64-1.24), P=0.476
If I'm working in the kitchen, like chopping vegetables or washing, my back pain gets worse and worse to reach a peak – then I have to sit down immediately	670	54 (8.1%)	931	148 (15.9%)	2.16 (1.55-3.00), P<0.001
If I have to stand for a long time I just know my back pain is going to get worse and worse	354	37 (10.5%)	1247	165 (13.2%)	1.31 (0.90-1.91), P=0.165
Generally my back pain is better with activity	1009	138 (13.7%)	592	64 (10.8%)	0.77 (0.56-1.05), P=0.096
Generally my back pain builds with activity	743	80 (10.8%)	858	122 (14.2%)	1.37 (1.02-1.86), P=0.039

Supplementary Table 6: Anatomical site of back pain and the presence or absence of OVFs (n=1593)

	Not marked		Yes, marked		(Univariable) Odds ratio (95%CI) (vs 'not marked' as ref.)
	Total n	Number (%) with OVF	Total n	Number (%) with OVF	
Thoracic area (either left or right or both)	1211	134 (11.1%)	382	68 (17.8%)	1.74 (1.27-2.39), P=0.001
Waist area (either left or right or both)	442	52 (11.8%)	1151	150 (13.0%)	1.12 (0.80-1.57), P=0.496
Low back/buttock area (either left or right or both)	467	85 (18.2%)	1126	117 (10.4%)	0.52 (0.38-0.71), P<0.001

Supplementary Table 7: Change in pain over time and the presence or absence of OVFs

	Not ticked		Yes, ticked		(Univariable) Odds ratio (95%CI) (vs 'not ticked' as ref.)
	Total n	Number (%) with OVF	Total n	Number (%) with OVF	
(A) How pain changes with time (n=1,575)					
Continuous, steady, constant	910	112 (12.3%)	665	86 (12.9%)	1.06 (0.78-1.43), P=0.712
Rhythmic, periodic, intermittent	966	122 (12.6%)	609	76 (12.5%)	0.99 (0.73-1.34), P=0.930
Brief, momentary, transient	1438	177 (12.3%)	137	21 (15.3%)	1.29 (0.79-2.11), P=0.309
Other pattern	1388	177 (12.8%)	187	21 (11.2%)	0.87 (0.54-1.40), P=0.556
(B) Pictorial description of course of pain (n=1,582)					
Persistent pain with slight fluctuations (diagram)	1130	137 (12.1%)	452	62 (13.7%)	1.15 (0.84-1.59), P=0.388
Persistent pain with pain attacks (diagram)	1315	172 (13.1%)	267	27 (10.1%)	0.75 (0.49-1.15), P=0.184
Pain attacks without pain between (diagram)	1084	137 (12.6%)	498	62 (12.5%)	0.98 (0.71-1.35), P=0.916
Pain attacks with pain between (diagram)	1425	185 (13.0%)	157	14 (8.9%)	0.66 (0.37-1.16), P=0.148
Pain increases and increases	1442	176 (12.2%)	140	23 (16.4%)	1.41 (0.88-2.27), P=0.152
Other pattern	1499	188 (12.5%)	83	11 (13.3%)	1.07 (0.55-2.05), P=0.849

Supplementary Table 8: Posture-related back pain and the presence or absence of OVFs (n=1601)

	I generally disagree /Didn't indicate		I generally agree		(Univariable) Odds ratio (95%CI) (vs 'disagree as ref.)
	Total n	Number (%) with OVF	Total n	Number (%) with OVF	
It feels as though my upper body is being tugged forward	1119	131 (11.7%)	482	71 (14.7%)	1.30 (0.95-1.78), P=0.095
It feels as though I am being pulled over all the time	1324	157 (11.9%)	277	45 (16.3%)	1.44 (1.01-2.07), P=0.046
It feels as though my head is too heavy	1206	156 (12.9%)	395	46 (11.7%)	0.89 (0.62-1.26), P=0.503
It feels as though there is nothing to hold my head or upper body up	1307	167 (12.8%)	294	35 (11.9%)	0.92 (0.63-1.36), P=0.684

Supplementary Table 9: Multivariable relationships between pain variables and the presence or absence of OVFs (n=1588)

	Odds ratio per unit change in predictor, and 95%CI	**P value**
Back pain described as stinging	1.72 (0.90-3.30)	P=0.100
Back pain described as sharp	0.50 (0.33-0.74)	P=0.001
Back pain described as like toothache	0.53 (0.33-0.87)	P=0.011
Agreement with 'If I'm working in the kitchen, like chopping vegetables or washing, my back pain gets worse and worse to reach a peak – then I have to sit down immediately'	2.13 (1.53-2.98)	P<0.001
Pain in thoracic area	1.55 (1.11-2.16)	P=0.010
Pain in low back/buttock area	0.59 (0.43-0.82)	P=0.001

Supplementary Table 10: Variables describing situations that affect back pain and the presence or absence of OVFs (n=1601)

	Total n	Number (%) with OVF	(Univariable) Odds ratio (95%CI) (vs 'no effect' as ref.)
Walking			
No effect	414	61 (14.7%)	1
Decrease	480	50 (10.4%)	0.67 (0.45-1.00)
Increase	707	91 (12.9%)	0.85 (0.60-1.21)
			Overall P (likelihood ratio) P=0.144
Bending			
No effect	363	54 (14.9%)	1
Decrease	82	5 (6.1%)	0.37 (0.14-0.96)
Increase	1156	143 (12.4%)	0.81 (0.58-1.13)
			Overall P (likelihood ratio) P=0.065
Standing			
No effect	279	31 (11.1%)	1
Decrease	63	7 (11.1%)	1.00 (0.42-2.39)
Increase	1259	164 (13.0%)	1.20 (0.80-1.80)
			Overall P (likelihood ratio) P=0.632
Stretching			
No effect	734	95 (12.9%)	1
Decrease	415	45 (10.8%)	0.82 (0.56-1.19)
Increase	452	62 (13.7%)	1.07 (0.76-1.51)
			Overall P (likelihood ratio) P=0.409
Cold			
No effect	993	136 (13.7%)	1
Decrease	30	1 (3.3%)	0.22 (0.03-1.61)
Increase	578	65 (11.3%)	0.80 (0.58-1.09)
			Overall P (likelihood ratio) P=0.073
Damp			
No effect	998	137 (13.7%)	1
Decrease	14	1 (7.1%)	0.48 (0.06-3.73)

Increase	589	64 (10.9%)	0.77 (0.56-1.05) *Overall P (likelihood ratio) P=0.198*
Lifting			
No effect	309	45 (14.6%)	1
Decrease	9	0	
Increase	1286	157 (12.2%)	0.82 (0.57-1.17), P=0.271
Sitting on straight-backed chairs			
No effect	558	55 (9.9%)	1
Decrease	524	69 (13.2%)	1.39 (0.95-2.02)
Increase	519	78 (15.0%)	1.62 (1.12-2.34) *Overall P (likelihood ratio) P=0.032*
Slouching			
No effect	623	88 (14.1%)	1
Decrease	104	13 (12.5%)	0.87 (0.47-1.62)
Increase	874	101 (11.6%)	0.79 (0.58-1.08) *Overall P (likelihood ratio) P=0.340*
Sitting on soft chairs			
No effect	653	101 (15.5%)	1
Decrease	239	27 (11.3%)	0.70 (0.44-1.10)
Increase	709	74 (10.4%)	0.64 (0.46-0.88) *Overall P (likelihood ratio) P=0.017*
Pulling shoulders back			
No effect	937	116 (12.4%)	1
Decrease	429	57 (13.3%)	1.08 (0.77-1.52)
Increase	235	29 (12.3%)	1.00 (0.65-1.54) *Overall P (likelihood ratio) P=0.889*
Sleeping			
No effect	711	99 (13.9%)	1
Decrease	467	64 (13.7%)	0.98 (0.70-1.38)
Increase	423	39 (9.2%)	0.63 (0.42-0.93) *Overall P (likelihood ratio) P=0.041*
Lying down			
No effect	513	74 (13.9%)	1

Decrease	652	87 (13.3%)	0.95 (0.68-1.33)
Increase	418	41 (9.8%)	0.67 (0.45-1.01)
			Overall P (likelihood ratio) P=0.115
Heat			
No effect	776	107 (13.8%)	1
Decrease	782	92 (11.8%)	0.83 (0.62-1.12)
Increase	43	3 (7.0%)	0.47 (0.14-1.54)
			Overall P (likelihood ratio) P=0.233
Using a hot water bottle or electric blanket in bed			
No effect	814	104 (12.8%)	1
Decrease	761	94 (12.4%)	0.96 (0.71-1.30)
Increase	26	4 (15.4%)	1.24 (0.42-3.67)
			Overall P (likelihood ratio) P=0.888
Reclining			
No effect	796	96 (12.1%)	1
Decrease	497	59 (11.9%)	0.98 (0.70-1.39)
Increase	308	47 (15.3%)	1.31 (0.90-1.91)
			Overall P (likelihood ratio) P=0.312
Changes in the weather			
No effect	1049	143 (13.6%)	1
Decrease	43	8 (18.6%)	1.45 (0.66-3.18)
Increase	509	51 (10.0%)	0.71 (0.50-0.99)
			Overall P (likelihood ratio) P=0.063
Twisting			
No effect	539	73 (13.5%)	1
Decrease	65	8 (12.3%)	0.90 (0.41-1.95)
Increase	997	121 (12.1%)	0.88 (0.65-1.20)
			Overall P (likelihood ratio) P=0.731

Supplementary Table 11: Frailty variables and the presence or absence of OVFs

	Total n	Number (%) with OVF	(Univariable) Odds ratio (95%CI)
Walking distance (n=1,337)			
>400 yards	985	113 (11.5%)	1
≤400 yards	352	53 (15.1%)	1.37 (0.96-1.94), P=0.081
Use of walking aid (n=1,584)			
No	1114	121 (10.9%)	1
Yes	470	78 (16.6%)	1.63 (1.20-2.22), P=0.002
Falls (n=1,555)			
Rarely	1332	165 (12.4%)	1
Few times per year or more	223	27 (12.1%)	0.97 (0.63-1.50), P=0.906

Supplementary Table 12: Traditional risk factors for osteoporosis and the presence or absence of OVFs

	Total n	Number (%) with OVF	(Univariable) Odds ratio (95%CI)
Oral steroids for >3 months (n=1,502)			
No	1344	154 (11.5%)	1
Yes	158	32 (20.3%)	1.96 (1.29-3.00), P=0.002
Mother or father had hip fracture (n=1,428)	1179	147 (12.5%)	1
No	249	35 (14.1%)	1.15 (0.77-1.71), P=0.495
Yes			
Smoking (n=1,589)			
Never	874	114 (13.0%)	1
Gave up	650	74 (11.4%)	0.86 (0.63-1.17)
Current	65	12 (18.5%)	1.51 (0.78-2.91)
			Overall P (likelihood ratio) P=0.239
Alcohol intake (n=1,583)			
<3 units or less per day	1436	177 (12.3%)	1
≥3 units per day	147	23 (15.7%)	1.32 (0.82-2.12), P=0.250

Supplementary Table 13: Previous fractures and the presence or absence of OVFs

	No		Yes		(Univariable) Odds ratio (95%CI) (vs 'no' as ref.)
	Total n	Number (%) with OVF	Total n	Number (%) with OVF	
Any fracture after aged 50	1007	82 (8.1%)	512	103 (20.1%)	2.84 (2.08-3.88), P<0.001
Fracture after aged 50 excluding hands, feet, head	1090	88 (8.1%)	437	96 (22.0%)	3.21 (2.34-4.39), P<0.001
Osteoporotic fractures* only after aged 50	1436	134 (9.3%)	120	53 (44.2%)	7.69 (5.14-11.49), P<0.001
Any fracture after aged 50 excluding high trauma	1022	85 (8.3%)	483	99 (20.5%)	2.84 (2.08-3.89), P<0.001
Fracture after aged 50 excluding hands, feet, head excluding high trauma[∞]	1106	90 (8.1%)	411	93 (22.6%)	3.30 (2.41-4.53), P<0.001
Osteoporotic fractures* only after aged 50 excluding high trauma[∞]	1442	137 (9.5%)	110	49 (44.6%)	7.65 (5.05-11.59), P<0.001
Low trauma[§] fracture after aged 50	1081	90 (8.3%)	424	93 (21.9%)	3.09 (2.26-4.24), P<0.001
Low trauma[§] fracture after aged 50 excluding hands, feet, head	1158	96 (8.3%)	357	86 (24.1%)	3.51 (2.55-4.84), P<0.001
Low trauma[§] osteoporotic fractures only after aged 50	1459	144 (9.9%)	92	42 (45.7%)	7.67 (4.92-11.97), P<0.001

*Osteoporotic fractures were defined according to anatomical site (hip, vertebral, humeral, forearm)

[∞]High trauma was defined according to the modified Landin description i.e. falls from more than 3 meters, car accidents, being hit by a heavy moving object or crushed in a machine

[§]Low trauma was defined according to the modified Landin description i.e. fall from standing height or less

Supplementary Table 14: Concomitant illnesses and the presence or absence of OVFs (n=1594)

	No		Yes		(Univariable) Odds ratio (95%CI) (vs 'no' as ref.)
	Total n	Number (%) with OVF	Total n	Number (%) with OVF	
Poor vision	1422	179 (12.6%)	172	22 (12.8%)	1.02 (0.63-1.64), P=0.940
Poor balance	1034	135 (13.1%)	560	66 (11.8%)	0.89 (0.65-1.22), P=0.466
Menopause before 45	1206	160 (13.3%)	388	41 (10.6%)	0.77 (0.54-1.11), P=0.164
Inflammatory arthritis	1190	153 (12.9%)	404	48 (11.9%)	0.91 (0.65-1.29), P=0.610
Depression	1181	157 (13.3%)	413	44 (10.7%)	0.78 (0.55-1.11), P=0.165
Memory problems	1347	167 (12.4%)	247	34 (13.8%)	1.13 (0.76-1.68), P=0.552
Anxiety	1039	139 (13.4%)	555	62 (11.2%)	0.81 (0.59-1.12), P=0.207
Diabetes (type 1 or 2)	1429	185 (13.0%)	165	16 (9.7%)	0.72 (0.42-1.24), P=0.236
COPD	1485	187 (12.6%)	109	14 (12.8%)	1.02 (0.57-1.83), P=0.939
Heart disease	1424	177 (12.4%)	170	24 (14.1%)	1.16 (0.73-1.83), P=0.531

Supplementary Table 15: Age and physical examination measurements of participants according to presence of any OVF.

	No OVF			OVF			Mann-Whitney U-test to compare the two groups
	Mean (SD)	**Median (IQR)**	**n**	**Mean (SD)**	**Median (IQR)**	**n**	**P value**
Age (years)	73.6 (5.6)	72.5 (69.2-76.8)	1400	76.1 (6.5)	75.1 (71.2-80.8)	202	P<0.001
Height (cm)	159.0 (6.2)	158.9 (154.8-163.4)	1392	156.9 (6.6)	157.5 (152.6-161.8)	202	P<0.001
Weight (kg)	72.7 (15.1)	70.6 (61.9-80.8)	1389	68.2 (15.4)	65.3 (57.1-75.4)	202	P<0.001
Chest expansion (cm)	3.4 (1.7)	3.0 (2.1-4.3)	1388	3.1 (1.6)	3.0 (2.0-4.0)	201	P=0.011
Waist circumference (am)	93.4 (13.4)	92.5 (83.5-102.1)	1391	91.4 (13.4)	89.5 (81.0-101.0)	202	P=0.028
Rib to pelvis distance (number of fingers)	2.4 (0.9)	2 (2-3)	1392	2.1 (1.0)	2 (1-3)	202	P<0.001
Wall to tragus distance (cm)	14.8 (3.5)	14.0 (12.1-16.8)	1393	16.5 (4.1)	16.0 (13.4-19.0)	202	P<0.001
Reported height loss (cm)	4.0 (3.0)	3.6 (1.8-5.7)	1314	6.4 (4.1)	5.4 (3.5-9.3)	193	P<0.001

Supplementary Table 16: Multivariable relationships between physical examination variables and the presence or absence of OVFs (n=1501)

	Odds ratio per unit change in predictor, and 95%CI	**P value**
Age (years)	1.00 (0.97-1.03)	P=0.965
Weight (kg)	0.98 (0.96-0.99)	P<0.001
Wall to tragus distance (cm)	1.07 (1.02-1.13)	P=0.003
Reported height loss (cm)	1.16 (1.10-1.22)	P<0.001

1. Clark D. Population of the UK 2019, by age group. https://www.statista.com/statistics/281174/uk-population-by-age/.
2. https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/demographics/male-and-female-populations/latest

3. Jordan, K.P., et al., International comparisons of the consultation prevalence of musculoskeletal conditions using population-based healthcare data from England and Sweden. Annals of Rheumatic Diseases,2014. 73: p. 212-218.