Data in Brief 38 (2021) 107387

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Application of the Rasch measurement framework to mammography positioning data

Patsy Whelehan[a,b,c,*], Maria Pampaka[d], Jennifer Boyd[a], Sarah Armstrong[a], Andy Evans[a,c], Gozde Ozakinci[b]

[a] University of Dundee, United Kingdom
[b] University of St Andrews, United Kingdom
[c] NHS Tayside, United Kingdom
[d] The University of Manchester, United Kingdom

ARTICLE INFO

Article history:
Received 28 June 2021
Revised 26 August 2021
Accepted 15 September 2021
Available online 20 September 2021

Keywords:
Mammography
Breast imaging
Clinical image quality
Rasch model
Measurement theory

ABSTRACT

The purpose of this article is to provide raw data and measure-validation data pertaining to a co-submission published in European Journal of Radiology and entitled: Development and validation of a novel measure of adverse patient positioning in mammography.

This Data in Brief article serves not only to provide greater detail than its companion article but also as an educational worked example of the Rasch measurement framework. Rasch measurement is a form of modern psychometric technique and our articles provide the first known example of its use in the evaluation of clinical radiological image quality.

The data consist of observations of mammographic images, plus limited participant parameters relevant to the measure validation process. Also provided are validation indices produced by subjecting the primary data to Rasch analysis.

An expert observer generated the primary data by reviewing mammographic images to judge the presence or absence of a set of features developed through theory and consultation with other experts. The validation data were generated through Rasch analysis, performed using Winsteps® soft-

DOI of original article: 10.1016/j.ejrad.2021.109747

* Corresponding author at: University of Dundee, United Kingdom.
E-mail address: p.j.whelehan@dundee.ac.uk (P. Whelehan).
Social media: (P. Whelehan), (J. Boyd), (G. Ozakinci)

https://doi.org/10.1016/j.dib.2021.107387

ware, which mathematically models the probability of having a correct response (or a present feature in this dataset) to an item in a given measurement instrument (*e.g.* questionnaire), as a function of the participant's ability/position on the underlying construct under study.
The data can be reused by anyone wishing to learn and practice psychometric validation techniques. They can also form a basis for researchers wishing to build on our preliminary measure for the assessment of mammographic clinical image quality.

Specifications Table

Subject	Radiology and radiography
Specific subject area	Mammography: measurement of positioning quality
Type of data	Table Image Graph Figure
How data were acquired	A questionnaire (supplied) was completed by an observer analysing mammographic images. Direct entry of observations (yes/no – feature present or absent) was performed, using a bespoke database in Microsoft Access. Patients self-reported their height and weight from which body mass index was calculated https://www.nhs.uk/common-health-questions/lifestyle/what-is-the-body-mass-index-bmi/ Patients' ages and mammographer identity were extracted from images' DICOM header using VolparaDataManager® software (Volpara Health Technologies Ltd, Wellington, New Zealand), algorithm version 1.5.2. The ages were then assigned to ten-year age bands and the mammographer identities were anonymised. Validation data were generated through Rasch analysis, performed using Winsteps® software.
Data format	Raw Analysed Filtered
Parameters for data collection	Data were collected from mammograms of women attending for breast cancer screening in the Scottish Breast Screening Programme, United Kingdom.
Description of data collection	Mammograms were viewed by an expert observer and scored for the presence of various features. The observer entered data directly into a Microsoft Access database, responding either Yes or No to whether a given feature was observed on the image.
Data source location	Institution: University of Dundee City/Town/Region: Dundee Country: Scotland, United Kingdom Latitude and longitude (and GPS coordinates, if possible) for collected samples/data: 56.4643° N, 3.0379° W
Data accessibility	The primary data are hosted in a reputable public repository hosted by the University of Dundee and known as "Discovery" https://discovery.dundee.ac.uk/ Data identification number: https://doi.org/10.15132/10000165
Related research article	This article is a companion article to the following: P. Whelehan, M. Pampaka, J. Boyd, S. Armstrong, A. Evans, G. Ozakinci, Development and validation of a novel measure of adverse patient positioning in mammography, Eur. J. Radiol. 140 (2021) 109747. https://doi.org/10.1016/j.ejrad.2021.109747

Value of the Data

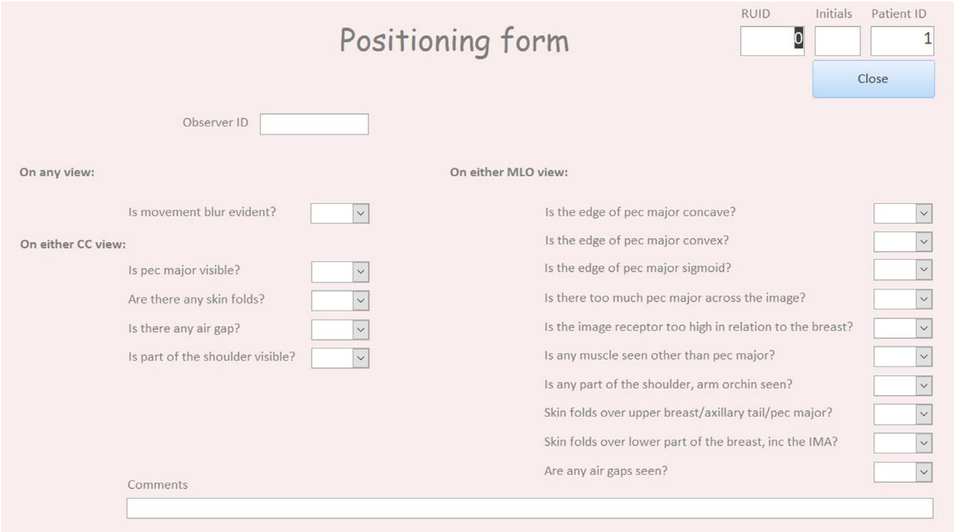
- The data can be used by researchers to learn and practice the techniques of Rasch analysis, comparing their results to ours for verification of correct technique.
- Any researcher wishing to develop skills in modern psychometrics can benefit from these data.
- Researchers can build on these data to develop an improved clinical image quality measure for mammography.

1. Data Description

- The primary raw data are provided on an Excel file stored at the University of Dundee data repository. The file includes one sheet with the data matrix with the first row being the variable name. The 22 variables are described as follows:
  - Columns A to O – the names of the variables (also shown in Table 1 below) match the short names provided with detailed descriptions in Table 1 in the associated manuscript [1], and are listed in full in the abbreviations table at the end of this section: These are the binary responses/scores of the observer of the mammograms on whether the specific features were present or not (coded as 1= present and 0= not present).
  - Column P – "Positioning": the continuous scores on the constructed scale, as produced by the Rasch procedures (described in the associated manuscript and detailed under the "Validation Methodology" section below).
  - Column Q – "Age_Cat": A categorical variable denoting the patients' ages in three categories (coded as 1= 50 to 59 years old, 2= 60 to 69 years old, 3=70 and over).
  - Column R – "BMIcategory": A categorical variable denoting the body mass index of the patients (coded as 1=Underweight (below 18.5), 2= Healthy weight (18.5–24.9), 3=Overweight (25–29.9), 4=Obese (30 and over)).
  - Column S – "MammographerID": a number from 1 to 12 used as the identifier of the person performing the mammogram (anonymised).
  - In all variables the character X was used to denote missing information.

Table 1
Item fit statistics (n = 310 participants; total missing datapoints: n = 8 instances in n = 2 participants).

					Infit		Outfit	
Item	TotalScore	TotalCount	Measure	SE	MNSQ	ZSTD	MNSQ	ZSTD
PecVisCC (1)	79	309	-0.18	0.14	0.95	-0.8	0.93	-0.7
FoldsCC (2)	178	309	-1.77	0.12	0.99	-0.1	0.98	-0.4
AirGapCC (3)	60	309	0.22	0.15	0.86	-1.7	0.76	-1.8
ShoulderCC (4)	21	309	1.50	0.23	1.07	0.4	*1.76	*2.3
CentredHigh (5)	126	309	-0.99	0.12	0.93	-1.6	0.90	-1.5
WidePec (6)	23	310	1.40	0.22	0.95	-0.2	0.75	-0.9
PecConcave (7)	84	310	-0.27	0.14	1.23	3.5	*1.47	*4.1
PecConvex (8)	55	310	0.33	0.16	1.07	0.8	1.20	1.3
PecSigmoid (9)	48	310	0.51	0.16	1.02	0.2	1.10	0.6
FoldsUpper (10)	145	309	-1.28	0.12	0.96	-1.1	0.93	-1.3
FoldsLower (11)	215	310	-2.38	0.13	0.95	-0.9	0.92	-0.9
AirGapMLO (12)	148	310	-1.31	0.12	0.93	-1.8	0.89	-2.1
MuscleOther (13)	61	309	0.18	0.15	1.06	0.7	1.13	0.9
AnatOther (14)	2	309	3.95	0.71	1.00	0.2	0.64	-0.3
Blur (15)	66	310	0.08	0.15	1.03	0.5	1.05	0.4
Mean:			0.00	0.19	1.00	-0.1	1.03	0.0
SD:			1.49	0.14	0.08	1.3	0.28	1.6


Positioning form

RUID: 0 | Initials: | Patient ID: 1

Close

Observer ID

On any view:

Is movement blur evident?

On either CC view:

Is pec major visible?

Are there any skin folds?

Is there any air gap?

Is part of the shoulder visible?

On either MLO view:

Is the edge of pec major concave?

Is the edge of pec major convex?

Is the edge of pec major sigmoid?

Is there too much pec major across the image?

Is the image receptor too high in relation to the breast?

Is any muscle seen other than pec major?

Is any part of the shoulder, arm orchin seen?

Skin folds over upper breast/axillary tail/pec major?

Skin folds over lower part of the breast, inc the IMA?

Are any air gaps seen?

Comments

Fig. 1. Instrument for rating mammograms, shown on the data entry screen into which the observer directly recorded their observations. Response was Yes or No.

Table 2
Standardized residual variance (in Eigenvalue units).

	Observed		Modelled
	Eigenvalue	Percentage	Percentage
Total raw variance in observations	20.9	100	100
Raw variance explained by measures	5.9	28.4	27.9
Raw variance explained by persons	1.2	5.6	5.6
Raw variance explained by items	4.8	22.7	22.4
Raw unexplained variance (total)	15	71.6–100	72.1
Unexplained variance in 1st contrast	1.8	8.7–12.1	

  ◦ Various tables and figures are then presented in this paper to describe how this raw dataset was constructed (Figs. 1 to 3) and how the ratings can be analysed to construct and validate the measure of positioning introduced in the associated manuscript (Tables 1 and 2; Figs. 4 to 6):
- Fig. 1 Screenshot of the data entry screen for the instrument for rating mammograms. The observer directly recorded their observations. Response was Yes or No.
- Fig. 2 Annotated medio-lateral oblique mammogram image showing examples of some of the features of interest.
- Fig. 3 Annotated cranio-caudal mammogram image showing examples of some of the features of interest.
- Fig. 4 Differential Item Functioning (DIF) according to patients' age group.
- Fig. 5 Differential Item Functioning (DIF) according to patients' BMI category.
- Fig. 6 Differential Item Functioning (DIF) according to which mammographer performed the examination.
- Table 1 Item fit statistics-showing how the observed data fit the predictions of the Rasch Model.
- Table 2 Standardised residual variance (in Eigenvalue units) – results of the principal component analysis of the residuals (i.e. comparing observed values to the ideal Rasch Model).

Abbreviation in dataset	Full name/description of variable
PecVisCC (1)	Pectoralis major muscle visible on cranio-caudal projection
FoldsCC (2)	Skin folds visible on cranio-caudal projection
AirGapCC (3)	Air gap visible on cranio-caudal projection
ShoulderCC (4)	Shoulder visible on cranio-caudal projection
CentredHigh (5)	X-ray beam centred too high in relation to breast (medio-lateral oblique projection)
WidePec (6)	Too much of pectoralis major muscle extending across the field of view
PecConcave (7)	Edge of pectoralis major muscle has a concave outline
PecConvex (8)	Edge of pectoralis major muscle has a convex outline
PecSigmoid (9)	Edge of pectoralis major muscle has a sigmoid outline
FoldsUpper (10)	Skin folds visible overlying the upper part of the breast in the medio-lateral oblique projection
FoldsLower (11)	Skin folds visible overlying the lower part of the breast in the medio-lateral oblique projection
AirGapMLO (12)	Air gap visible on the medio-lateral oblique projection
MuscleOther (13)	Any muscle other than pectoralis major visible on the medio-lateral oblique projection
AnatOther (14)	Any other extraneous anatomical structure visible on the medio-lateral oblique projection
Blur (15)	Motion blur visible
Age_Cat	Age category
BMIcategory	Body Mass Index category
Mammographer ID	Anonymised identity code of the mammography practitioner

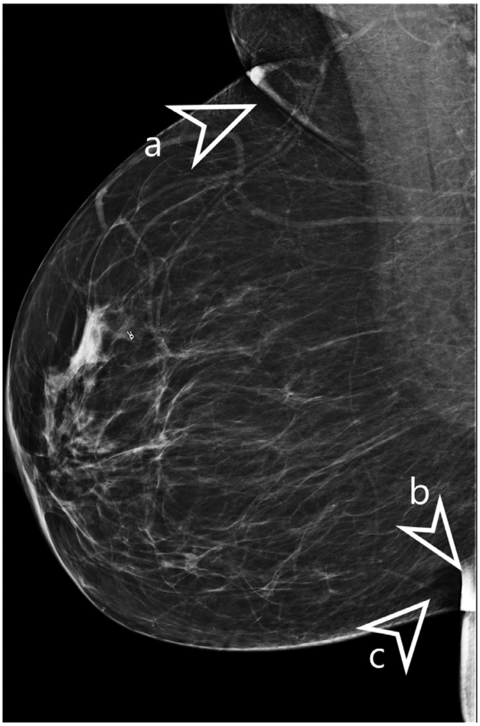


Fig. 2. Example of a medio-lateral-oblique mammogram showing (a) skin fold over the upper part of the breast, (b) skin fold at the lower part of the breast (inframammary angle), (c) air gap associated with the lower skin fold (darker area in front of skin fold).

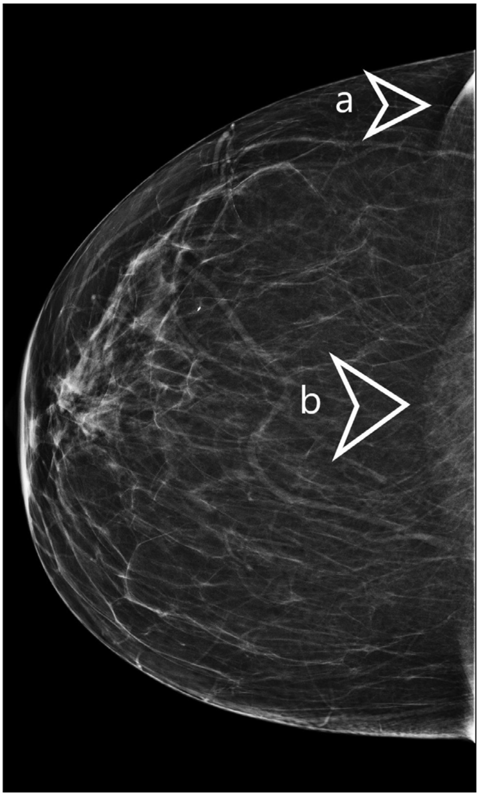


Fig. 3. Example of a cranio-caudal mammogram showing (a) a minor skin fold and (b) the anterior aspect of the pectoralis major muscle.

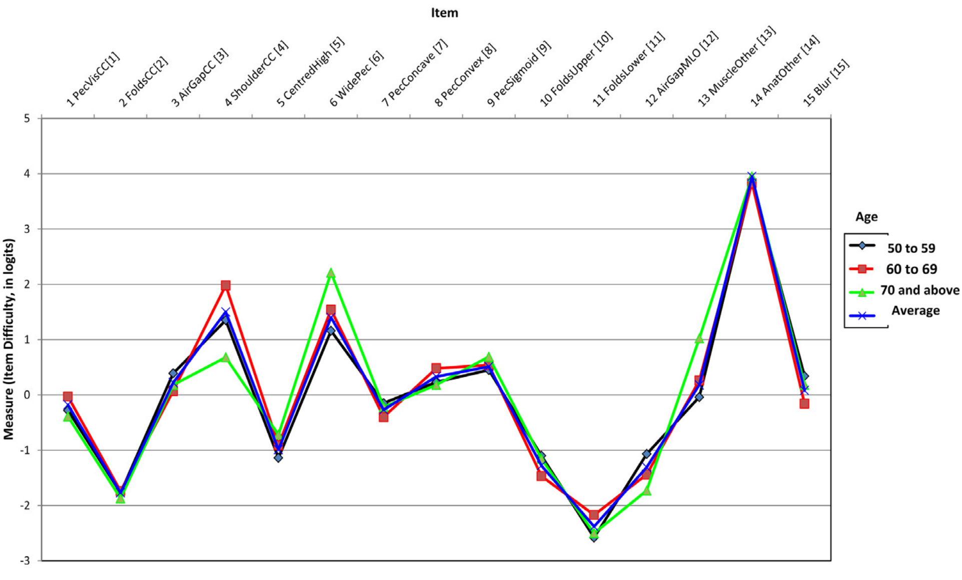


Fig. 4. Differential Item Functioning (DIF) according to patient age group. Age groups are 1: 50–59 years; 2: 60–69; 3: 70 and above.

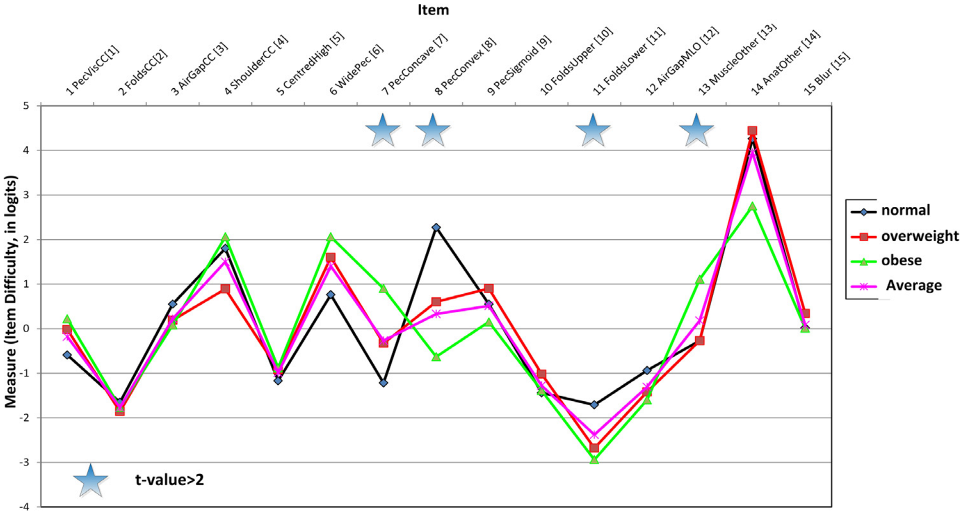


Fig. 5. Differential Item Functioning (DIF) according to patient body mass index (BMI).

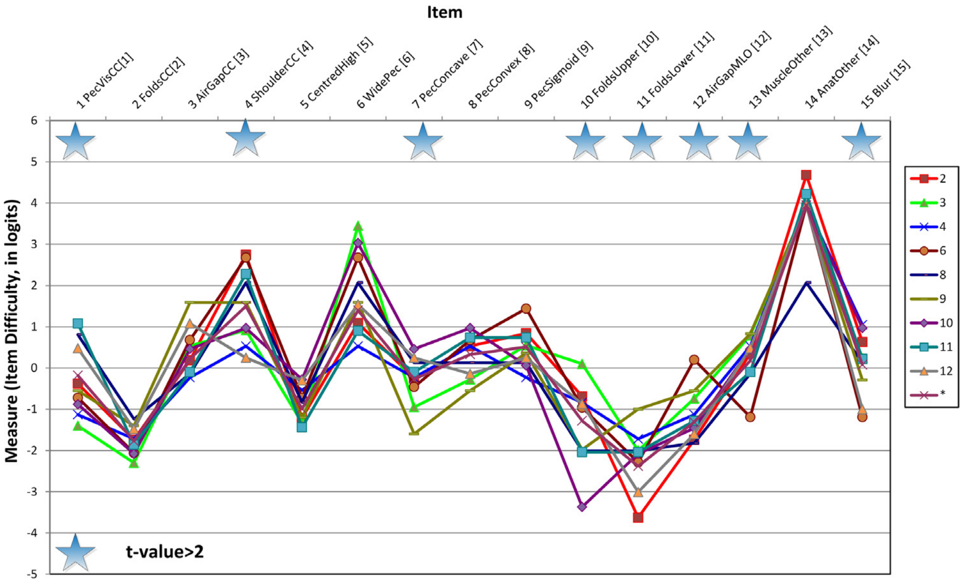


Fig. 6. Differential Item Functioning (DIF) according to which mammographer performed the examination. Mammographers 1, 4 and 7 have been removed because of low frequencies of examinations performed in the study participants.

2. Experimental Design, Materials and Methods

Participant recruitment and study design, materials and methods are described in full in our associated manuscript [1].

Participants consisted of 310 women attending a population-based breast screening service in the UK.

Consenting participants provided their self-reported height and weight, from which body mass index (BMI) was calculated using the formula specified here: https://www.nhs.uk/common-health-questions/lifestyle/what-is-the-body-mass-index-bmi/

Patient age and mammography practitioner identity were extracted from the DICOM header (metadata) of the image files using VolparaDataManager® software (Volpara Health Technologies Ltd, Wellington, New Zealand), algorithm version 1.5.2.

2.1. Mammogram data

The observer viewed the mammograms on a mammography-grade workstation and recorded the presence or absence of the features of interest directly into a database, responding Yes or No to the questions. The instrument used for rating the mammograms is shown in Fig. 1, which is a screenshot of the data-entry screen in a Microsoft Access database designed for the study.

Figs. 2 and 3 show examples of mammograms showing some of the features of interest.

With written informed participant consent, the ratings of the mammograms were matched with patients' age-bands, BMI and mammography practitioner, and this is the resulting raw dataset (provided in the excel file as detailed earlier).

In the following sections we describe the methodology for validating the measure of adverse positioning in mammography, providing considerably more detail than we were able to include in the parent publication [1].

3. Validation Methodology

Although the Rasch model has principally been used in education research, it has also been successfully applied in healthcare research over many years [2–6]. The validation process within the Rasch framework involves the accumulation of evidence to establish whether the proposed philosophical/empirical construct exists as a distinct, unidimensional "measure" (or scale), and if not whether there are other relevant or useful dimensions.

Rasch analysis, performed using Winsteps® software [7], mathematically models the probability of having a correct response (or a present feature in this case) to an item in a given measurement instrument (e.g. questionnaire) as a function of the participant's ability/performance on the underlying construct under study. When this item-response data adequately fit the Rasch model, objective and valid measurement has been achieved. Rasch techniques allow ordinal/categorical raw data to be converted to a continuous scale – provided that the data fit the Rasch model adequately. This facilitates further statistical analysis.

In this dataset, "participants" are patients' mammograms, with each consisting of four radiological images. The measurement instrument (or tool) is the list of items presented in Table 1 in the main manuscript [1] (and also shown in Fig. 1 and Table 1 in this article). The measurement data were generated by a human observer responding to the items (questions) in the instrument as to whether or not the item was present in each mammogram. The responses were denoted as 0 (absence of the feature) and 1 (presence of the feature), dictating the direction of the resulting measure: the higher the score, the more adverse the positioning. Given the binary response format used in this study, the dichotomous Rasch model was the most appropriate [8].

Decisions about the validity of the measures are based on a range of statistical indices comparing the observed data to the predictions of the Rasch model. These indices include item fit statistics, item and person separation and reliability, and differential item functioning [9] as detailed next and illustrated with example outputs from the analysis of this dataset to help the reader with the interpretation of such outputs.

3.1. Item fit statistics

Item fit statistics indicate how accurately the data fit the Rasch model and thus provide evidence of whether the unidimensionality assumption has been fulfilled. Unidimensionality means the presence of a single coherent construct captured by the items in the measurement instrument (*e.g*. questionnaire or the scoring instrument in this case).

In a "perfect" model, fit statistics (*i.e*. infit and outfit mean-squares (MNSQ)) should be 1, but an acceptable range is 0.6 to 1.4 depending on the analysis. Higher infit and outfit values indicate more variation in the observed data than those expected from the Rasch model (in an ideal measurement), while lower values indicate less variation in the observed response pattern compared to that predicted [8]. Infit and outfit values above the recommended thresholds (data underfitting the model) indicate that responses are more haphazard than expected.

Infit is affected by unexpected responses to the item by participants whose overall level on the scale is near that of the item's level; outfit is more sensitive to unexpected responses among those whose level on the scale is far from the item's level [5]. Infit/outfit values below acceptable thresholds (data overfitting the Rasch model) indicate item redundancy [6].

For most analyses, such as the example here, we take values for infit and outfit mean squares of 1.4 and above as suggesting cause for concern and requiring further exploration, because values above 1 suggest that data are unpredictable, under-fitting the model.

All infit values (Table 1) were within acceptable ranges, providing evidence for measure validity. Two items (asterisked) show slightly higher than desirable values for outfit, which is associated with outlier response patterns, i.e. responses which do not fit well with the model's expectations. Removal of these two items is not desirable because they are considered important to the measure overall. For example, Item 4 was the least frequently observed item so its removal would reduce the amount of variation captured by the measure.

Item 4 refers to the inclusion on the cranio-caudal image of part of the patient's shoulder. According to clinical experience, this is an uncommon fault and may be more likely in slim and/or elderly women with postural or anatomical concavity of the chest and shoulder area. Item 7 refers to the contour of the pectoralis major muscle being shown on the medio-lateral oblique image as concave. This fault is believed to indicate that the muscle is not lying flat on the detector assembly, and/or is tense. It may also be caused by insufficient displacement of the muscle and breast medially or by the breast being considerably thicker than the muscle mass in the included field in slim women with relatively large breasts. Overall, the acceptable infit values suggest outlier responses as the cause of the observed outfit misfit. Item 4 is infrequently endorsed so it is not surprising for it to be subject to outlier responses.

3.2. Dimensionality checks

Principal component analysis of the residuals produced by comparing the observed data to the Rasch ideal model provides additional evidence of unidimensionality or lack thereof (presence of more than one dimension) [6].

Table 2 shows the results of principal component analysis (PCA) of the model residuals. The closeness of the observed (empirical) and modelled variance percentages indicates that the value for the raw variance explained by the measures is reliable. The low unexplained variance Eigenvalue of 1.8 is further evidence of unidimensionality to add to that provided by the item infit and outfit statistics.

3.3. Item and Person separation and reliability indices

Item separation indices give an estimate of the ordering and spread of items along the continuum of the construct being measured, i.e. indicating the ability of the measure to define a distinct hierarchy of items along the overall variable being measured [8].

Item reliability reports how reproducible the ordering of items along the measure is. Higher item reliability indices imply greater confidence in reproducibility of item ordering across different samples. Item reliability indices perform a similar function to the Cronbach's Alpha statistic used in classical test theory approaches to psychometric validation. In the Rasch measurement framework, if item separation and reliability indices are below recommended thresholds, a larger sample size may be necessary.

Person separation indices indicate the ability of the measure to differentiate participants (participants' mammograms in our example) into different groups. Person reliability refers to the reproducibility of the differentiation afforded by the measure across different samples of participants. Poor person separation and reliability values indicate that more items may be required in the measurement instrument, and/or response formats with more categories may be needed.

The values for item and person separation and their reliability are as follows for this example:

- Item separation: 6.17; Reliability: 0.97
- Person separation: 0.65; Reliability: 0.30.

Interpreted according to guidance from Wright and Stone [10], item separation and reliability indices are very good. These indices suggest that the sample of examinations was sufficient to produce a reliable item hierarchy map, i.e. they provide further evidence of measure validity.

The person separation and reliability indices, which should be > 2 and > 0.8 respectively, are less satisfactory. This suggests that the instrument may not be sensitive enough to distinguish between high and low scoring mammograms on the adverse positioning measure, and more items or the use of multi-category response formats, rather than binary (yes/no) may be needed.

3.4. Differential item functioning (DIF)

Along with person reliability indices, differential item functioning relates to the reliability of group differentiation of the constructed measure, which is an important aspect of validity when an instrument is to be used with different groups of participants or on different occasions. For a measure to be unidimensional, and the variable to be linear, the scale values of the items have to work invariantly across individuals and groups [11]. Lack of invariance among sample groups, for example according to gender or country, is known as differential item functioning or DIF. However, DIF may indicate genuine, relevant differences between groups, so items demonstrating DIF do not necessarily need to be resolved or eliminated.

The line graphs in Figs. 4 to 6 show the differences in items' measures based on different calibrations per group, and the average. The figures also indicate (asterisks) items with statistically significant differences, *i.e*. DIF. As shown in Fig. 4, there are no significant differences in item functioning based on patient age-group but four items exhibit significant DIF based on BMI group (Fig. 5). For mammographers (Fig. 6), most of the items exhibit significant DIF.

In the absence of an existing evidence base, clinical experience suggests that three of the four items with significant DIF for BMI group can be explained. The observation "PecConcave" (Item 7) is subjectively perceived to be more common in women with low BMI whereas "PecConvex" (Item 8) seems more common in high BMI. "FoldsLower" is also considered more likely in women with high BMI, where the abdominal wall may frequently intrude on the image, overlapping with the lower part of the breast posteriorly. The fourth BMI DIF item ("muscle other") is less easy to explain, pertaining as it does to the inclusion in the field of a muscle other than pectoralis major, usually pectoralis minor . It is not immediately obvious whether this would be observed at different frequencies in either high or low BMI patients. Overall, the fact that the DIF can mostly be reasonably explained by BMI indicates that it is likely resulting from substantive differences rather than biased items.

While mammographers undergo extensive specialist training, and while mammographic positioning is ideally standardised, these are difficult examinations to perform unvaryingly. Experience in clinical and training contexts indicates that individual mammographers' practice varies and that certain practitioners more frequently produce images with particular features. Research evidence suggests that the amount of compression force applied to the breast during mammography varies according to mammographer [12]. Such variability in practice may extend to positioning and there is the additional likelihood of interplay between compression and positioning.

3.5. Person-item map for the "adverse positioning" measure

Person-item maps and the item difficulty hierarchy provide evidence for substantive, content and external validity. These aspects can be defined as follows: substantive validity is the extent to which the theoretical foundation underlying the construct of interest is sound; content validity is whether the test items appear to be measuring the construct of interest; external validity is whether the test has convergent, discriminant and predictive qualities [13].

Using Winsteps® software [7], a "map" can be produced that displays the locations of both participants and items on a single logit (log odds unit) scale, produced through log transformation of the raw categorical scores during the analytical process [8] (please see Fig. 3 in the associated manuscript [1]. This is an interval scale, i.e. the gradations are of equal magnitude to each other. Traditionally, because of the educational research origins of Rasch analysis, the terms "person ability" and "item difficulty" are used in these person-item maps. In our example, person ability translates to the level of adverse positioning pertaining to each mammogram while item difficulty indicates where each individual feature of adverse positioning sits on the overall adverse positioning scale. The resulting scores for each mammogram in this logit scale are included in the excel file (under column Q, named "Positioning") and were used in further analysis presented in the associated manuscript [1].

Ethics Statement

This study was carried out with ethical approval from the relevant authorities and in line with the Declaration of Helsinki. (East of Scotland NHS Research Ethics Committee ref: 16-ES-0083; University of St Andrews Ethics Committee ref: MD12255). All participants gave written informed consent. Applicable research governance and information governance principles were followed.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have or could be perceived to have influenced the work reported in this article.

CRediT Author Statement

Patsy Whelehan: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization, Project administration, Funding acquisition; **Maria Pampaka:** Methodology, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization; **Jennifer Boyd:** Investigation, Data curation, Writing – review & editing; **Sarah Armstrong:** Investigation, Data curation, Writing – review & editing; **Andy Evans:** Investigation, Writing – review & editing; **Gozde Ozakinci:** Methodology, Writing – review & editing, Supervision, Funding acquisition.

Acknowledgments

We gratefully acknowledge the contributions to this work of the study participants and the administrative and radiographic staff in the screening clinics. We are also grateful to Dr Violet

Warwick for database design, to Dr Kulsam Ali for help with participant recruitment and data management, Dr Rachel Nutt for data management, and to Dr Sarah Vinnicombe for comments on the draft manuscript.

Funding sources

This work was supported by the School of Medicine Research Investment Fund at the University of St Andrews, and by charitable donations from Mrs Fiona Edwards (no grant numbers).

VolparaDataManager® software was supplied free-of-charge in exchange for expert advice over several years. The company was not involved in conducting the work and none of the authors has received any payment from the company.

References

[1] P. Whelehan, M. Pampaka, J. Boyd, S. Armstrong, A. Evans, G. Ozakinci, Development and validation of a novel measure of adverse patient positioning in mammography, Eur. J. Radiol. 140 (2021) 109747, doi:10.1016/j.ejrad.2021.109747.

[2] P.G. Conaghan, A. Tennant, C.G. Peterfy, T. Woodworth, R. Stevens, A. Guermazi, H. Genant, D.T. Felson, D. Hunter, Examining a whole-organ magnetic resonance imaging scoring system for osteoarthritis of the knee using Rasch analysis, Osteoarthr. Cartil. 14 (2006) 116–121, doi:10.1016/j.joca.2006.03.011.

[3] J. Kean, E.F. Bisson, D.S. Brodke, J. Biber, P.H. Gross, An Introduction to item response theory and rasch analysis: application using the eating assessment tool (EAT-10), Brain Impair. 19 (2018) 91–102, doi:10.1017/BrImp.2017.31.

[4] M. Wirtz, M. Boecker, T. Forkmann, M. Neumann, Evaluation of the “consultation and relational empathy” (CARE) measure by means of Rasch-analysis at the example of cancer patients, Patient Educ. Couns. 82 (2011) 298–306, doi:10.1016/j.pec.2010.12.009.

[5] Y. Ko, N.N. Lo, S.J. Yeo, K.Y. Yang, W. Yeo, H.C. Chong, J. Thumboo, Rasch analysis of the oxford knee score, Osteoarthr. Cartil. 17 (2009) 1163–1169, doi:10.1016/j.joca.2009.04.004.

[6] H. Court, K. Greenland, T.H. Margrain, Measuring patient anxiety in primary care: rasch analysis of the 6-item spielberger state anxiety scale, Value Heal. 13 (2010) 813–819, doi:10.1111/j.1524-4733.2010.00758.x.

[7] J.M. Linacre, Winsteps® Rasch Measurement Computer Program, Winsteps.com, 2018 (accessed April 19, 2020).

[8] T. Bond, TG; Fox, Applying the Rasch Model: Fundamental Measurement in the Human Sciences, 3rd Ed., Routledge, Hove, 2015.

[9] E.W. Wolfe, E.V. Smith, Instrument development tools and activities for measure validation using rasch models: part II - validation activities, J. Appl. Meas. 8 (2007) 204–234.

[10] B. Wright, M. Stone, Separation statistics, in: Meas. Essentials, 2nd Ed., Wide Range Inc, Wilmington, 1999, pp. 151–155.

[11] C. Hagquist, D. Andrich, Recent advances in analysis of differential item functioning in health research using the Rasch model, Health Qual. Life Outcomes 15 (2017) 181, doi:10.1186/s12955-017-0755-0.

[12] C.E. Mercer, P. Hogg, K. Szczepura, E.R.E. Denton, Practitioner compression force variation in mammography: A 6-year study, Radiography 19 (2013) 200–206 http://libproxy.dundee.ac.uk/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=jlh&AN=104199679&site=ehost-live&scope=site.

[13] S. Messick, Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning, Am. Psychol. 50 (1995) 741–749, doi:10.1037/0003-066X.50.9.741.