Check for updates

RESEARCH ARTICLE

Precision medicine implementation and research-practice partnerships: implications of measurement scale differential item functioning [version 1; peer review: 1 approved, 1 not approved]

John J. O. Mogaka , Moses J. Chimbari

Discipline of Public Health Medicine, University of KwaZulu-Natal, Durban, 4041, South Africa

V1 First published: 28 Jul 2020, 9:782
<https://doi.org/10.12688/f1000research.25010.1>
Latest published: 28 Jul 2020, 9:782
<https://doi.org/10.12688/f1000research.25010.1>

Abstract

Background: Omics-based biomarkers (OBMs) inform precision medicine (PM). As omics-based technologies gradually move into clinical settings, however, a co-occurrence of biomedical research and clinical practice is likely an important variable in the implementation of PM. Currently, little is known about the implications of such research-practice co-occurrence.

Methods: This study used data collected from a pilot study designed to inform a full-scale PM implementation study through the validation of the measurement tool. It applied item response theory (IRT) methods to assess the tool's reliability and measurement invariance across two study subgroups associated with research and practice settings.

Results: The study sample consisted of 31 participants. Measurement invariance assessment was through differential item functioning (DIF) analysis with bootstrapping through Monte Carlo simulation. Overall, 13 out of 22 items that formed the PMI scale had DIF at significance level $\alpha=0.25$. Item response functions (IRFs) revealed how each subgroup members responded to scale items and their attitudes towards factors that influence PM implementation.

Conclusions: Attitudinal similarities and differences towards factors influencing PM implementation amongst those in biomedical research as compared with those in practice were established. Results indicated PM implementation knowledge that is unique and common to both groups. The study established the validity and reliability of the new PM implementation measurement tool for the two subgroups.

Keywords

precision medicine implementation, omics technologies, implementation measurement, differential item functioning, item response theory

Open Peer Review

Approval Status

	1	2
version 1 28 Jul 2020	 view	 view
1. Gerald Mboowa , Makerere University College of Health Sciences, Kampala, Uganda		
2. Ken Asada , Cancer Translational Research Team, RIKEN Center for Advanced Intelligence Project, Tokyo, Japan		
Any reports and responses or comments on the article can be found at the end of the article.		

Corresponding author: John J. O. Mogaka (johnmogaka91@gmail.com)

Author roles: **Mogaka JJO:** Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – Original Draft Preparation; **Chimbari MJ:** Conceptualization, Data Curation, Methodology, Project Administration, Resources, Supervision, Validation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2020 Mogaka JJO and Chimbari MJ. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Mogaka JJO and Chimbari MJ. **Precision medicine implementation and research-practice partnerships: implications of measurement scale differential item functioning [version 1; peer review: 1 approved, 1 not approved]**

F1000Research 2020, 9:782 <https://doi.org/10.12688/f1000research.25010.1>

First published: 28 Jul 2020, 9:782 <https://doi.org/10.12688/f1000research.25010.1>

Introduction

Recent advances in omics technologies offer greater understandings of how gene-environment interactions affect health and pathological processes. Omics-based biomarkers (OBMs) resulting from advances in biomedical technologies form the basis of precision medicine (PM), an approach to medicine that emphasizes predictive, preventive and personalized medical care¹. Deep insights gained in systems biology, molecular evolution and microbial environments is increasingly being applied in clinical settings through OBMs, with huge biomedical, clinical and public health implications. For instance, due to broadening understanding of the complex genotype-lifestyle-environment interactions, how biomedical research evidence is collected and applied is changing to reflect a shift from “average many” to “unique few” approach to health interventions. Evidence-based medicine (EBM)², the widely applied and accepted approach to medicine, is mostly informed by systematic reviews, meta-analyses or group-level studies from which average recommendations for the “average person” are derived. However, this “one-size-fits-all” approach has severe limitations in that it does not adequately cater for uniqueness in individuals, neither does it consider disease heterogeneity in larger populations. Although most people may be accounted for within some average categories, yet many others are ‘outliers’, falling outside recommended mean estimates. Although understanding individual-level underpinnings of health and disease at finer resolution may be overwhelming in practice, PM provides a shift away from the generalized average protocols to personalized strategies in medical care. Moreover, PM, through OBMs and other biotechnologies, enables better disease delineation and patient stratification. In emphasizing individual uniqueness at the molecular level and tailored care based on uniqueness of genotype, PM also enables advanced detection of pre-symptomatic conditions in the population and individuals, while improving surveillance and management of infectious diseases. This results in better-adapted and personalized therapies, significantly delayed disease onset, as well as disease prevention.

However, while interventions associated with PM, including cell-, gene-, and tissue-engineered therapies³ are highly personalized in nature, their clinical validity and utility is generally limited. This lack of adequate clinical utility implies that in PM, the boundary between research and practice is blurred as opposed to the traditionally more pronounced research-practice boundary in EBM. Despite mounting evidence supporting the value of OBMs in improving health outcomes (e.g., Berg AO⁴), the moving-target nature of most genomic-based interventions implies that the evidence base for PM must constantly be updated as more knowledge is gained over time. Implementation of PM therefore inevitably juxtaposes biomedical research and clinical practice elements, implying co-existence of both research and practice settings. Although such co-occurrence settings create novel opportunities to advance medicine through PM, little is currently known about the nature and implications of such research-practice partnerships, especially for PM implementation at health systems level.

We assessed a PM implementation (PMI) measurement model using a latent variable approach informed by item response

theory (IRT). The study constructed four variables (factors) that were hypothesized to have varying influences on PM implementation at health systems level. These were: 1) Characteristics of omics biomarkers; 2) Organizational support in terms of resources required for implementation; 3) Public genomic awareness; and 4) precision medicine implementation outcomes. OBMs are biological disease- or health- status indicators gleaned from genomics (gene structure and sequencing), epigenomics (gene expression via transcriptomics)⁵, proteomics (RNA products)^{6,7}, metabolomics (metabolites) and DNA-based tools that detect bacteria, parasites and viruses that coexist in the human body through metagenomics⁸. Despite apparent advantages associated with omics biomarkers, varying perceptions about their effectiveness persist, largely due to less-conclusive clinical validity. Concerns around OBMs include questions around the biomarkers’ reliability, interpretation of test results and quality standards⁹. Based on existing literature, six items were selected as indicators for the ‘Omics Biomarkers’(OBM) construct as illustrated in [Table 1](#). Organizational support relates to resource investments required to move OBMs from point of discovery to point of clinical application. For instance, material and non-material investments are needed to continuously update existing biomedical knowledge including new algorithms and bioinformatics methodologies, as well as high-velocity and high-capacity computation facilities. Resources are also required in compliance with good practice in genomic data governance through optimized data capturing, storage, transfer and processing. Organizational support, therefore, is a factor that can influence PM implementation outcomes. Six items were selected as its indicators ([Table 1](#)). On the other hand, although use of biomarkers is beneficial in circumstances such as identifying susceptibility for inherited conditions, this often implies genetic profiling, which can lead to anticipation, skepticism and concern at the personal level^{10,11}. Individuals, empowered through publicly available information prevalent in social media and other connected technologies, may advocate for adoption of certain biomedical technologies. They may also manifest willingness to participate in genomic studies, as they appreciate relative importance of doing so not only for themselves but for public good in general. Therefore, the level and quality of public genomic awareness give rise to differing genomic attitudes, perceptions and values among members of the public, both as individuals and as health providers. Six items were selected to be indicative of ‘Public Genomic Awareness’, as seen in [Table 1](#). Finally, the PM implementation outcomes factor captures the long-term goal of PM implementation efforts: improved uptake and routine use of OBMs in healthcare. Four items were selected as measures of the construct ‘Implementation Outcomes’ ([Table 1](#)).

As omics-based technologies gradually move into clinical settings, a co-occurrence of biomedical research and clinical practice is likely to present as an important variable in the implementation of PM. This co-occurrence, however, is complicated by potential differential interpretation of factors that affect or influence PM implementation. Modern psychometric tools offer reliable means of assessing item-level measurement variability that focus on differential item functioning (DIF)¹². DIF has been defined as a measure of different probabilities of

Table 1. Factors influencing precision medicine implementation at health systems level, and their indicators.

Construct	Item code and label
Characteristics of "Omics biomarkers"	<p>X1. The omics test has been used among people similar to the present target population</p> <p>X2. Genetic counselling is part of the procedures in administering the omics test</p> <p>X3. It is easy to obtain the right amount of sample needed for the omics test from participants/patients</p> <p>X4. It is easy to obtain the right quality sample needed for the omics test from participants/patients</p> <p>X5. The time it takes to obtain results for intended use after the omics test is reasonable</p> <p>X6. There are clear instructions on how to obtain samples for the omics test</p>
"Organizational Support"	<p>X7. The amount of money dedicated by our organization for implementing this omics test is enough</p> <p>X8. The amount of physical space dedicated by our organization for the purposes of implementing this omics test is enough</p> <p>X9. As the volume, variety and data availability associated with the implementation of this omics test grows, the organization increases the capacity of its data systems to facilitate analysis and user needs.</p> <p>X10. To match latest data handling needs in the implementation of this omics test, the organization regularly hires people with relevant skills.</p> <p>X11. There exist clear procedures on how to store, access, manage and share available data associated with this omics test in the organization</p> <p>X12. The organization has capacity to handle (e.g. capture, validate., store, and process) data associated with this genetic/omics biomarker</p>
"Public Genomic Awareness/Acceptance": i.e. Acceptability and compatibility of the omics tests among Involved Individuals	<p>X13. Getting buy-in from the public (patients, providers) in carrying out this biomarker testing is easy</p> <p>X14. Participants easily give consent to take samples for biomarker testing.</p> <p>X15. Publicity and free information publicly available about the genetic/omics biomarker cause potential users to willingly ask or look for it</p> <p>X16. Using this genetic/omics test has been regarded by practitioners as an appropriate mechanism for patient management (e.g. aid in drug dosage decisions, in carrying fetuses to term or carry out prophylactic surgery).</p> <p>x17. There is a considerable 'pushback' from practitioners as they feel the genetic/omics test is not consistent with their skills, role, or job expectations.</p> <p>x18. Target individuals feel that the genetic/omics test is in line with their family members' wishes, desires and expectations</p>
"Implementation outcomes"	<p>X19. The genetic/omics test is yet to be used as a routine practice within its intended service setting</p> <p>X20. Practitioners are more willing to order the genetic/omics test more often whenever deemed necessary</p> <p>X21. The number of eligible persons accessing the genetic/omics test is far less than the total number potentially in need of the service</p> <p>X22. So far, the authorities that are supposed to acquire the biomarker testing service have communicated a decision to fully fund its roll out</p>

success or endorsement across construct while controlling for the underlying trait being measured¹³. DIF represents unequal response patterns among groups, which can profoundly threaten research findings, as it may constitute bias. To assess this kind of differential item functioning in PM, this paper sought to establish a PM implementation measurement tool's differential item functioning (DIF) across two study participant sub-groups; one affiliated to academia (representing research) and the other affiliated to industry (representing practice).

In as much as measurement variability due to sub-group membership - irrespective of the construct being measured - has often been underscored in other disciplines¹⁴, it is important to assess the same in the new field of PM. Item-level invariance assessment entails a valuable exercise as it serves to test what may often be implicit assumptions concerning much measurement processes¹⁵. Furthermore, such assessment will help curtail a number of problems including errors in hypothesis testing, flawed population forecasts and policy planning¹⁶ in

relation to PM implementation. The study also confirmed the tool's psychometric reliability and construct validity.

Table 1 shows the four PM implementation factors (latent variables) and items that measure them.

Methods

Study design

This was a cross-sectional study that employed analytical approaches for quantifying factorial relationships. DIF analysis helped to reveal differences in perceptions towards PM implementation between two study participant subgroups: one affiliated to academia and the other affiliated to industry. Item response theory (IRT) is one of the widely used methods for assessing DIF. In this study, IRT allowed for a critical examination of the information that was harvested from the manner study participants endorsed and ranked the various items associated with PM implementation. As noted earlier, DIF examines the probability of endorsing an item conditioned on the latent trait. This method was deemed suitable for examining DIF in this study as IRT examines the monotonic relationship between responses and the latent trait¹⁷. Furthermore, as compared to classical test theory, IRT parameter estimates are not only as confounded by sample characteristics, but statistical properties of items can be expressed with greater precision, thereby increasing the interpretation accuracy of DIF between sub-groups¹⁸.

Members of “academia” were defined as individuals involved in biomedical research that related to translating newly discovered molecular or omics-based biomarkers (OBMs) for purposes of clinical or population health use. Members of “industry” were defined as individuals involved in the clinical use of the biomarkers or those involved in commercialization activities related to OBMs (e.g., Direct-To-Consumers Genetic Testing, DTC-GT)¹⁹. “Precision medicine implementation” was defined as the process of translating newly discovered omics-based biomarkers (OBMs) for clinical or population health use. “OBMs” were defined as candidate genetic biomarkers that are in the process of being clinically validated or those already validated and in clinical use.

Study participants

The snowball sampling method, a non-probability sampling method, was applied in identifying potential study participants from population of interest. Inclusion was based on a participant's affiliation to either an academic institutions or a commercial entity (industry); the institution or organization had to be involved in molecular/genetic testing and/or omics-based biomarkers in Africa; a participant had to be involved in the field of precision medicine and/or biomarker-related biomedical research as indicated by contribution to the field through publications in peer-reviewed journals and/or attendance of related academic conferences. It is well known that snowball sampling poses significant recruitment “community bias” risk whereby recruitment of new members does not branch out from the sample subgroup of individuals initially accessed. In the present study,

this risk of source bias was addressed by properly selecting the initial individuals to ensure proportional inclusivity of subgroups in the seed (initial) sample. Consequently, the seed sample was identified and selected from a precision medicine-related conference; an event we regarded as representative of our study population as it brought together diverse groups from academia and industry involved in precision medicine and its implementation in resource-constrained settings. Therefore, the probability of selection bias for subsequent in-link recruitment waves of study participants was negligible. Besides, it has been shown that with an appropriate bootstrap resampling, selection bias due to the initial seed sample (if known to be representative of the targeted population) is progressively attenuated as the sample expands wave by wave^{20,21}.

Guided by general principles of the Nuremberg Code, the Declaration of Helsinki and institutional review board permit obtained from the University of KwaZulu-Natal (BREC Permit Ref No BE513/18), a study package was distributed via email to potential participants between June and July 2019. Since this study contained negligible risk of potential embarrassment or other ethical dilemmas that are usually associated with snowball sampling in many other studies, initial participants were encouraged to forward the email containing the study package to their colleagues. The study package (available as *Extended data*)²² included an invitation letter with study description, consent form and a link to the online card-sort platform. A total of 31 study participants were recruited for the study. Although there is no definitive number for a pilot study sample size according to Hunt, Sparkman and Wilcox²³, to increase statistical power and degrees of freedom for our study, we applied a Monte Carlo simulation (MCS) approach as described by Mooney²⁴. MCS studies are computer-driven experimental investigations in which certain parameters, such as population means and standard deviations that are known a priori, are used to generate random (but plausible) sample data²⁴. This method of generating and analyzing data treats the collected research sample as a “population reservoir” from which a large number of random samples are drawn with continuous replacement such that the probability of selection for any given case remains equal over every random draw²⁵. We requested 2,000 iterations, drawn with replacement from the original data set of 31 cases (our empirical sample).

Procedure

A card-sort task approach²⁶ was used to obtain comparative item-rating data. Respondents were presented with a set of cards that contained constructs clearly defined using everyday language and a set of categories onto which the concepts on the cards were to be mapped, as explained by Angleitner, John, & Lohr²⁷. The respondents were asked to read each item and assign it to the construct or concept it best indicated, in their judgment. Participants were required to rank the cards according to their own understanding and preference, indicating their perception on concepts presented on each card. If participants deemed it necessary to make any changes as to

where a card previously assigned should be reassigned to, they could do so. Although simple in practice, this exercise generated enough data that was used to assess the desired group perception levels, construct validity and scale reliability. The actual sorting of the items was, however, performed using an online platform hosted and supported by Optimal Workshop²⁸. The items are presented in Table 1. Besides offering convenience to study participants, the online platform enhanced data security and confidentiality. A total of 22 items were to be assigned into four categories initially provided for, with participants allowed to create own separate construct categories if they so desired. The results about number of items per category and the items' ranking order (positioning) within their respective categories were then extracted and analyzed. Participants' demographic data were also extracted, including sex, highest educational qualifications and age. Data was subsequently subjected to item response theory (IRT) analysis^{29,30} with logistic regression modelling.

The data

In summary, the data are an output of an exercise where respondents were required to place items arranged as cards into specific categories and ranked in order of perceived importance. The cards related to specific attributes indicative of wider factors thought to influence precision medicine implementation at health systems level. This meant that if an item on the card was regarded as the most important in each category, it was ranked 1st, and the next ranked 2nd and so on. If a card was not placed in a category, it was scored zero in that same category. For compactness and statistical purposes, the data was formatted to have one observation per row (for 31 participants, there were 31*4 = 124 observations).

Statistical considerations

One of the assumptions in item response theory (IRT) model analysis is that factors are generally unidimensional³¹. Therefore, to examine differences in perception between the two sub-groups (academic and industry) related to each item in the PM implementation scale, DIF was tested within each subscale. Each of the four factors that were identified as influencing PM implementation correspond to a subscale. Thus, four sets of tests were conducted with each of the four factors (sub-scales) examined separately for both item response functions (IRFs) and unidimensionality. To determine and confirm sub-scale unidimensionality, we used R package "mokken" version 2.8.11³². In this study, scale reliability referred to the consistency of measurement, while construct validity refers to the extent to which a set of indicators that were devised to gauge the four constructs (OBM, OrG, PGA and ImO) really measured that construct. Differential item functioning (DIF) was defined as a property of an item that shows the extent to which the item might be measuring different dimensions of the variable it was supposed to measure for members of separate subgroups. We applied the ordinal logistic regression modelling function in the package "lordif" (logistic ordinal regression differential item function) using IRT, version 0.3-3³³ in R version 3.6.1³⁴. The package "lordif" provides a logistic regression and IRT framework for detecting various types of differential item functioning³³. We preferred it because it

could adequately handle the item-sort-and-rank data that we used. To analyze DIF in the scale, we devised three models. The first of the three models distinguish DIF and non-DIF items. The second model identifies uniform-DIF items that occur due to an item tapping a dimension of the attribute measured in the scale that manifests differently between the groups. That is, for uniform DIF, whereas the probability of endorsing an item at a certain ranking is greater for one group than the other, this is uniform over all levels of knowledge of the factors under consideration. Finally, the third model identifies non-uniform DIF; that is the probability of endorsing an item at a certain ranking is greater for one group than the other, but this is not uniform over all levels of knowledge of the factors under consideration. This may be because of distracting elements (not necessarily related to the scale) in the measurement process, such as different understandings of a word or phrase used on the item³⁵.

These three models can be summarized as³⁶:

$$\text{logit}(P(U = 1)) = \beta_0 + \beta_1\theta + \beta_2g + \beta_3\theta g \text{ (Model 3: Non-uniform DIF)}$$

$$\text{logit}(P(U = 1)) = \beta_0 + \beta_1\theta + \beta_2g \text{ (Model 2: Uniform DIF)}$$

$$\text{logit}(P(U = 1)) = \beta_0 + \beta_1\theta \text{ (Model 1: No DIF)}$$

where:

$(P(U = 1))$ is the probability of endorsing an item at a particular rank in a category,

θ is the participant's latent perception level on the factor the item is measuring,

g is the group to which the participant belongs (academia or industry).

We used two sets of logistic regression-based criteria for detecting DIF: one set based on statistical significance (likelihood ratio χ^2 tests) and the other on Pseudo R^2 magnitude (Nagelkerke's and McFadden's pseudo R^2). Uniform DIF was examined by comparing the log likelihood of model 1 with 2 (degree of freedom (df)=1) and non-uniform DIF by comparing model 2 with 3 (df=1). The comparison between model 1 and model 3 (df=2) detected the total DIF effect—both uniform and non-uniform DIF^{36,37}. Items that revealed significance in any of the three likelihood ratio χ^2 tests with an alpha of 0.25 (Model 1 versus 2, Model 2 versus 3, and Model 1 versus 3) were flagged as having DIF. To examine the magnitude of DIF, we used various pseudo R^2 ^{238,39} such as Nagelkerke's Pseudo R^2 and McFadden's pseudo R^2 statistics for the defined model comparisons, and applied Zumbo et al.'s guideline to evaluate R^2 ²⁴⁰: effect size was regarded as negligible DIF (pseudo $R^2 < 0.035$), moderate DIF ($0.035 \geq$ pseudo $R^2 < 0.070$), and large DIF (pseudo- $R^2 \geq 0.070$).

For individual participants, the difference between the DIF-adjusted (purified) subscale score and the initial unadjusted score for each subscale. Finally, we obtained item and subscale characteristic curves to show the impact of DIF for

each group, i.e. item parameter estimates were computed and graphically examined via item characteristic curves (ICCs) and other graphs. R code used to calculate DIF is available as *Extended data*⁴¹.

Descriptive statistics that were obtained for the data included analysis of outliers, which was defined as any value that is greater than three standard deviations above or below the mean. Multivariate skewness and kurtosis statistics of the data were assessed using R package ‘psych’ version 1.8.12⁴².

Results

Skew values ranged from 1.21 to 2.41 while kurtosis (k) statistics for 21 items ranged from -0.37 to 2.49, with the 22nd item indicated at $k = 5.87$. **Figure 1** is a quantile-quantile (Q-Q) plot formed by the data and shows the distribution of the data against the expected normal distribution. Since the observations seemed to approximate a straight diagonal line with minimal deviations, we concluded that our data has a general normal distribution. Despite one item being an outlier with $k = 5.87$, its inclusion in the data set did not alter results, and therefore all items were included for further analyses.

The data did not contain any missing values because all the fields of the online item sort tool for data collection were set as “required” to eliminate non-responses. *Underlying data* are available at Zenodo²².

Demographic characteristics of participants

Of the 35 participants invited and who attempted the item-sort exercise, 31 successfully completed it (88% response rate).

There were three more female participants than males (10%). Those with research qualifications (PhDs and equivalent) were the majority (55%), while slightly a quarter of the participants were medical doctor. The rest of participants’ qualities varied greatly and are summarized in **Figure 2**.

Variable correlation

The correlogram in **Figure 3** is an inter-item correlation plot indicating item and scale structure in the data. In the upper triangle, positive correlations are displayed in blue and negative correlations in red colored circles. Color intensity and the size of the circle are proportional to the correlation coefficients, helping to identify “groups” of variables that share a strong relationship with each other (hierarchical clustering). The lower triangular correlation matrix displays the actual correlation values.

Each item seems to fit into one of the four well defined clusters (**Figure 3**), implying four factors defined the data as hypothesized. Items x16 (“*Using this genetic/omics test has been regarded by practitioners as an appropriate mechanism for patient management (e.g. aid in drug dosage decisions, in carrying fetuses to term or carry out prophylactic surgery)*”; and x20 (“*Practitioners are more willing to order the genetic/omics test more often whenever deemed necessary*”) are the only items not showing a strong correlation within the groups they belong.

To confirm the suggested four factors in the variable correlation matrix, we performed further scale analysis using functions in the “moken”⁴³ R package. Four unidimensional scales

Figure 1. Quantile-quantile (Q-Q) plot describing distribution of the data (y-axis) against the expected normal distribution (x-axis).

Figure 2. Demographic characteristics of study participants.

Figure 3. Variable correlation matrix.

were identified as dimensionality solutions via the package's automated item selection algorithm (AISP) performed at factor loading ≥ 0.3 and Type I error level (α)=0.05. Table 2 shows the results of this procedure and various reliability associated with each factor (subscale).

Sub-scales were demarcated by means of scalability coefficients⁴⁴ as indicated by the coefficient H in Table 2. At the lowest scalability threshold of 0.3, four unidimensional subscales were identified through the AISP. AISP is a scaling method of item selection we applied to partition the 22 items into subscales. This implies that there were four latent factors (subscales) measured by the 22 items (i.e., four underlying variables explain the association between latent factors and responses to items). Unidimensionality ranges

between H=0 and H=1: strong unidimensional scale has $H \geq 0.5$, moderate $0.4 \leq H < 0.5$, and weak $0.3 \leq H < 0.4$ ⁴⁴. The four subscales' coefficients H ranged between $H \geq 0.5$ and $H = 0.8$, implying the subscales were strongly unidimensional. This was strongly supported by the reliability indices (Mokken's rho, alpha and Guttman's lambda-2 (λ_2)) which indicated high internal consistency (internal validity/goodness-of-fit) for the four subscales. Cronbach's α , the expected correlation of two tests that measure the same construct, was separately obtained for the four constructs. All four subscales had coefficient alpha (α) ranging between 0.76 and 0.95 (with an average $\alpha = 0.88$). Besides, given that internal scale consistency (intercorrelations among items) is maximized when all items in a subscale measure the same construct, we concluded that the high Cronbach's alpha values for our subscales are an

Table 2. PMI Scale dimensionality solution, subscale reliability indices, likelihood ratio χ^2 test and Nagelkerke's pseudo R².

AISP: $c \geq 0.3$		χ^2 Statistics			Nagelkerke's Pseudo R ²		
G.H/Reliability	Item	Chi12	Chi13	Chi23	R12	R13	R23
OBM	x1*	0.191	0.031	0.041	0.015	0.05	0.04
H = 0.7	x2	0.583	0.478	0.669	0.002	0.006	0.004
MS (rho) = 0.93	x3*	0.131	0.009	0.011	0.016	0.061	0.05
Alpha(α) = 0.92	x4*	0.696	0.002	0.007	0.001	0.062	0.061
Lambda.2 = 0.93	x5*	0.086	0.669	0.209	0.022	0.023	0.001
	x6	0.78	0.668	0.877	0.0004	0.001	0.001
OrG	x7*	0.05	0.025	0.06	0.02	0.04	0.02
H = 0.8	x8	0.91	0.97	0.82	0.0001	0.0005	0.0004
MS (rho) = 0.96	x9*	0.81	0.25	0.09	0.0003	0.017	0.017
Alpha(α) = 0.95	x10*	3E-04	6E-04	0.181	0.06	0.06	0.008
Lambda.2 = 0.96	x11	0.73	0.9	0.78	0.0007	0.0012	0.0005
	x12*	0.024	0.033	0.185	0.04	0.05	0.01
PGA	x13	0.89	0.98	0.94	0.0001	0.0001	0.0001
H = 0.6	x14	0.96	0.88	0.61	0	0.003	0.003
MS (rho) = 0.89	x15	0.61	0.8	0.65	0.001	0.003	0.001
Alpha(α) = 0.88	x16*	0.92	0.12	0.035	0	0.027	0.027
Lambda.2 = 0.89	x17	0.97	0.99	0.95	0	0	0
	x18*	0.01	0.01	0.15	0.09	0.12	0.03
ImO							
H = 0.5	x19*	0.04	0.1	0.7	0.02	0.02	0.001
MS (rho) = 0.78	x20*	0.78	0.1	0.03	0.0004	0.03	0.03
Alpha(α) = 0.76	x21*	0.18	0.09	0.08	0.01	0.02	0.02
Lambda.2 = 0.81	x22	0.8	0.9	0.7	0.0001	0.001	0.0008

Key: MS=Molenaar-Sijtsma statistic (a.k.a. rho); Alpha=Cronbach's alpha; Lambda.2=Guttman's lambda.2; OBM=Characteristics of omics biomarkers construct; OrG=Organizational support construct; PGA= Public genomic awareness construct; ImO = PM implementation outcome construct. 'Chi12', 'Chi13', and 'Chi23' denote the likelihood ratio χ^2 statistic between Models 1 and 2, Models 1 and 3, and Models 2 and 3, respectively; R12, R13, R23 indicate Nagelkerke's pseudo R² from comparing Models 1 and 2, Models 1 and 3, and Models 2 and 3, respectively. Scale dimensionality solution was obtained by the automated item selection algorithm (aisp).

indirect but confirmatory indicator of the degree to which the items in the subscales measure corresponding unidimensional latent construct. However, Cronbach's alpha has been found to have limitations in measuring reliability in some instances^{45–47}. Due to this concern, we obtained other reliability indices for the same purpose as described below.

Guttman's λ_2 is a reliability estimate that is more robust than Cronbach's alpha, i.e., more resistant to outliers. The range of λ_2 values for the four subscales were: $0.81 \leq \lambda_2 \leq 0.96$. This implies that more than 80% of variance in each of the subscales is attributable to the true score and less than 20% to error, confirming that the internal construct validity and scale reliability is good. On the other hand, Mokken's rho, considered to be an improvement in relation to Cronbach's alpha and whose values should exceed 0.70 to show adequate scale reliability⁴⁸, was obtained. As shown in Table 2, all the four subscales had their rho (MS) coefficients above 0.78, confirming good scale reliability.

Table 2 also shows DIF detection outcomes using logistic regression approach³³. DIF items revealed from the likelihood ratio χ^2 tests are indicated with asterisks in Table 2. Of the 22 items making up the PM implementation (PMI) scale, 13 were flagged as DIF items at $\alpha=0.25$ with MCS set at 2,000. They all showed statistical significance for total DIF effect test when comparing Models 1 and 3. Of the 13 DIF items however, seven were indicated as mixed DIF items. Their likelihood ratio χ^2 test comparisons between Model 1 and 2, and models 2 and 3 were statistically significant. Moreover, of the remaining six, five had non-uniform DIF (x1, x4, x9, x16 and x20); the comparisons were statistically significant for Models 2 vs 3 as compared to Models 1 vs 2. Only one item had uniform DIF (x19): it showed non-significant comparisons for models 2 vs 3 but significant comparisons for models 1 vs 2. However, using the same statistical significance test, all items were DIF flagged at $\alpha=0.5$ and non at $\alpha=0.01$. This finding indicates the effect of statistical significance levels in DIF detection.

However, in contrast to the likelihood ratio χ^2 tests that detected 13 of the 22 items as DIF-positive, Nagelkerke's Pseudo R2 statistics flagged only five items as DIF (x1, x4, x7, x10 and x12). These were all moderate DIF ($0.035 \geq \text{pseudo R2} < 0.070$). To verify this result, we inspected other pseudo R2 statistics, such as McFadden's and Cox and Snell's R2. The additional pseudo R2 indices also yielded similar values.

Graphical Analyses of DIF Effects

From the above analyses, we generated various graphs. However, due to space constraints, we present only those graphs related to characteristics of omics biomarkers (OBM construct). Diagnostic plots are used below to describe the relationship between the perception latent trait and the probability of responding positively to an item as observed from the data (positively here means endorsing an item in any of the four scales). Participant's perception trait level is signified by theta (θ).

The top left plot in Figure 4a shows item true score functions based on group-specific item parameter estimates. It shows the item characteristic curves (ICCs) for the item for industry vs academia affiliated participants represented with red dashed and black solid lines respectively. The slope of the function for industry group was substantially higher than that for academia group, indicating non-uniform DIF. The upper-right graph shows the absolute difference between the ICCs for the two groups, indicating that the difference is mainly at high levels of perception of omics biomarkers (especially between average ($0 \geq \theta \leq +3$)). The lower-left graph shows the item response functions for the two groups based on the affiliation-specific item parameter estimates (slope and category threshold values by group, as also printed on the graph). It juxtaposes the item response functions for academia and industry groups. The non-uniform component of DIF revealed by the LR χ^2 test can also be observed in the difference of the slope parameter estimates (33.46 vs. 2.75). The lower-right graph shows the absolute difference between the ICCs (the upper-right graph) weighted by the score distribution (size) for the focal group, i.e., industry group, indicating minimal impact.

Figure 4b displays the boxplot and a scatter plot that show DIF items as unique to each group, accounting for DIF in trait estimates. It shows the difference in latent trait level between the initial IRT-based trait estimate with ignored DIF items and the purified trait estimate with DIF items accounted for. Each (black) circle and (red) triangle stand for the latent trait estimate of the OBM domain for a single participant. It also shows the expected amount of change in latent trait estimates when DIF is accounted for. Accounting for DIF items in a multi-item scale is commonly referred to as "scale purification". In both graphs, the y-axis is the difference (initial – purified) and the x-axis on the right graph is the initial latent trait level. The Box-and-Whisker plot on the left shows that median difference was around -0.04, and the middle 50% of respondents, depicted as the box (i.e., the interquartile range) ranged from approximately -0.04 to 0.07, indicating a fairly normal but slight right-skewed distribution with possible outliers at high ends (θ above average). The scatter plot on the right shows the initial (no DIF) vs purified (DIF accounted for) difference against the initial latent trait level that ignored DIF, implying that across the entire latent trait continuum, those participants who were affiliated to industry (red triangles) show more positive difference, further suggesting that accounting for DIF leads to lower scores for them compared to initial (non DIF) scores; for those affiliated to academia, (dark circles), the pattern is in the opposite. Guidelines are placed at 0.0 (solid line), i.e. no difference, and the dotted horizontal reference line is drawn at the mean difference between the initial and purified estimates.

What is suggested in Figure 4b corresponds well with Figure 4c, the test characteristic curves (TCC). The y-axis of Figure 4c indicates all possible scores for domain "OBM". Figure 4c indicates a clear difference in TCC curves, albeit in small margins, of DIF item responses between academia and industry groups. On the other hand, the left graph for all items displayed

Figure 4. Graphical Analyses of DIF Effects. (a) Graphical display of the item “X1: The omics test has been used among people similar to the present target population” which shows non-uniform DIF with respect to affiliation to either academia or industry. (b) Box-and-whisker plot and a scatter plot showing (individual-level) differences between initial and purified scale scores for each respondent. (c) Impact of DIF items on test characteristic curves (TCC) for all items and DIF items. (y-axis = sum of the item scores of the OBM domain; x axis (theta) = knowledge of OBMs). (d) Smoothed histogram showing perception differences towards omics biomarkers between those affiliated to academia and industry.

a much smaller difference than the right graph of Figure 4c. This can be explained as the DIF impact being diluted by the non-DIF items, resulting in the overall score difference between the two groups becoming minimal. It is evident that at average knowledge levels ($\theta = 0$), the scale performs better among those affiliated to industry than academia. But at higher trait levels ($\theta \leq +2$), industry lags behind academia.

Finally, Figure 4d shows smoothed histograms of the omics biomarkers perception (knowledge) levels for those study participants affiliated to industry (dashed red line) and academia (solid black line) as measured by the OBM subscale (theta). Industry affiliated participants on average had lower mean scores than their academia-affiliated counterparts, although there is a broad overlap in the distributions.

Discussion

Items related to public genomic awareness (PGA) appear to have the least DIF. In contrast, all the factors (characteristics of omics biomarkers (OBM), organizational support (OrG) and precision medicine implementation outcomes (ImO)) had a substantial number of items flagged for DIF. There were two types of

DIF items that identified in this study: uniform and non-uniform. The only item shown to have uniform DIF on the scale was item x19 (“The genetic/omics test is yet to be used as a routine practice within its intended service setting”); implying that one subgroup was consistently more likely than another to endorse this item at each level of the measured trait. Five items that showed explicit non-uniform DIF implied that there was inconsistent pattern of response; i.e., one group crossing-over, so that at certain levels of the trait, one group was more likely to endorse the item, while at another level, the other group was more likely to endorse the same item. Public genomic awareness is closely related to social and ethical nuances of research in precision medicine^{49,50}. PGA focuses on ethical questions ranging widely: from social justice issues regarding healthcare equity and access to specific questions regarding consent, safety and public acceptability of genomic technologies. The level of omics technology acceptability among practitioners (physicians) is another aspect of PGA⁵¹. Our study findings agree with other findings in literature that focus on genomic engagement with public and practitioners^{49,50,52,53}. The PGA non-DIF items indicate deep-cutting agreement across study participants on the likelihood of inadequate genomic

knowledge and genetics personnel to deal with issues surrounding ethical, legal and social implications of advancements in precision medicine. The fact that both subgroups (academia and industry) mostly agreed on most PGA aspects suggests that in a research-practice partnership designed to advance PM, both groups are likely to have similar outlooks towards ethics and genomics such as the balance between health gains and possible loss of privacy which may warrant sensitive, ongoing attention. The non-DIF items about PGA among study participants also suggests overwhelming support for community consultation and involvement where clinical implementation of omics technologies is concerned.

Assessment of invariance of chosen measurement models across subgroups is likely to emerge central to policy efforts that assess optimal ways of implementing PM at national health system level. This study has helped show the critical importance of a measurement tools' psychometric evaluation and how DIF analysis can play a vital role in implementation decisions. For instance, depending on the statistical procedure used for DIF detection, differing results may be yielded. Thus, one procedure may indicate DIF for certain items while others do not. The present study's findings, however, have another policy implication: in as much as evidence of DIF may play a vital role, it should by no means be the sole basis for policy decisions and conclusions. Presence of DIF may be an indication of problematic items that need to be revised or omitted and not necessarily an indication of sub-group irrationality. Therefore, DIF analysis can be considered a useful tool for item analysis but is more effective when combined with robust reasoning. Another consideration in the face of DIF in a measurement model is that a sub-group's domain score should be interpreted in a way that corrects for the measurement bias. This can lead to two policy implications: either to develop subgroup norms or to understand why DIF items occur in the first place and then to later revise and improve DIF items of the measurement model.

Conclusion

More generally, precision medicine combines both practice and research components. The relationship between academic research and clinical practice in precision medicine is, however, not easy and straightforward. In as much as public health and clinical practice through PM may benefit from advances in omics technologies and deeper gene-environment insights that flow from rigorous biomedical research, research also benefits from being informed by practice problems and practical knowledge, leading to co-creation of solutions to broader PM issues in both spheres.

This study assessed latent traits (factors) that relate to precision medicine implementation using a newly formulated implementation measurement scale in order to validate the scale's general reliability and construct accuracy. The findings demonstrate general conceptual and psychometric measurement equivalence of the scale across two subgroups involved in PM implementation: academia (research) and industry (practice). Differential item functioning (DIF) analysis was used to study the performance of the 22 items of the PMI measurement scale, examining whether the likelihood of item (categorical)

endorsement is equal across the two subgroups as matched on the measured trait. Even though there were no DIF flagging at statistical significance of $\alpha=0.5$, DIF analysis at $\alpha=0.25$ revealed the presence of some DIF items in the scale. There were two types of DIF items that were pointed out in this study: uniform and non-uniform.

Moreover, the study delved into the assessment of statistical basis of the measurement model as well as providing various indices such as scalability, reliability and unidimensionality. All four subscales showed strong scalability. The a priori subscales were strongly supported by the automated item selection algorithm, AISP, further validating the theory behind the generation of the measurement items.

Apart from highlighting DIF items on a newly developed precision medicine implementation (PMI) measurement tool, this paper also pointed out potential and benign differences between perception on PM implementation held by those affiliated to industry (those in clinical and public health practice) and those in biomedical research. This has both implicit and explicit implications in the field of PM implementation, especially at health systems level. First, it highlighted the kind of information separately held by these two subgroups in the field of PM. Secondly, it also highlighted the kind of information held together in common by these important subgroups in the pursuit of ideals of PM. Thirdly, the findings contained in this paper may inform the kind of information that should be availed to them. With this paper, we hope to have contributed to draw attention to those in PM to the fact that biomedical research and PM practice need deliberative attention from both researchers and practitioners to succeed.

Data availability

Underlying data

Zenodo: Extended data for Manuscript: Precision medicine implementation and research-practice partnerships: implications of measurement scale differential item functioning (DIF). <http://doi.org/10.5281/zenodo.3925276>²².

File 'DIF Analysis Raw.Extended Data' (XLSX) contains the redacted survey responses from the card sort exercise and the results of the card sort exercise itself.

Extended data

Zenodo: Extended data for Manuscript: Precision medicine implementation and research-practice partnerships: implications of measurement scale differential item functioning (DIF). <http://doi.org/10.5281/zenodo.3925276>²².

File 'DIF Study Package.Extended data' (DOCX) contains the study package provided to participants.

Zenodo: R code for DIF. <http://doi.org/10.5281/zenodo.3925336>⁴¹.

This project contains the R code for this study in R and PDF formats.

Data are available under the terms of the [Creative Commons Attribution 4.0 International license \(CC-BY 4.0\)](#).

References

1. Collins FS, Varmus H: **A new initiative on precision medicine.** *N Engl J Med.* 2015; **372**(9): 793–5.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
2. Del Mar C, Glasziou P, Mayer D: **Teaching evidence based medicine.** *BMJ.* 2004; **329**(7473): 989–90.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
3. Abou-El-Enen M, Hey SP: **Cell and Gene Therapy Trials: Are We Facing an 'Evidence Crisis'?** *EClinicalMedicine.* 2019; **7**: 13–4.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
4. Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group: **Recommendations from the EGAPP Working Group: can tumor gene expression profiling improve outcomes in patients with breast cancer?** *Genet Med.* 2009; **11**(1): 66–73.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
5. Cazaly E, Saad J, Wang W, *et al.*: **Making sense of the epigenome using data integration approaches.** *Front Pharmacol.* 2019; **10**: 126.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
6. Chambers G, Lawrie L, Cash P, *et al.*: **Proteomics: a new approach to the study of disease.** *J Pathol.* 2000; **192**(3): 280–8.
[PubMed Abstract](#) | [Publisher Full Text](#)
7. Hanash S: **Disease proteomics.** *Nature.* 2003; **422**(6928): 226–232.
[PubMed Abstract](#) | [Publisher Full Text](#)
8. Wooley JC, Godzik A, Friedberg I: **A primer on metagenomics.** *PLoS Comput Biol.* 2010; **6**(2): e1000667.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
9. Dancy JE, Dobbin KK, Groshen S, *et al.*: **Guidelines for the development and incorporation of biomarker studies in early clinical trials of novel agents.** *Clin Cancer Res.* 2010; **16**(6): 1745–55.
[PubMed Abstract](#) | [Publisher Full Text](#)
10. Holtzman NA, Marteau TM: **Will genetics revolutionize medicine?** *N Engl J Med.* 2000; **343**(2): 141–4.
[PubMed Abstract](#) | [Publisher Full Text](#)
11. Hubbard R, Lewontin RC: **Pitfalls of genetic testing.** *N Engl J Med.* 1996; **334**(18): 1192–4.
[PubMed Abstract](#) | [Publisher Full Text](#)
12. Swaminathan H, Rogers HJ: **Detecting differential item functioning using logistic regression procedures.** *J Educ Meas.* 1990; **27**(4): 361–70.
[Publisher Full Text](#)
13. Camilli G, Shepard LA: **Methods for identifying biased test items.** Sage; 1994.
[Reference Source](#)
14. Anderson RBW: **On the comparability of meaningful stimuli in cross-cultural research.** *Sociometry.* 1967; **30**(2): 124–36.
[PubMed Abstract](#) | [Publisher Full Text](#)
15. Millsap RE: **Statistical approaches to measurement invariance.** Routledge; 2012.
[Publisher Full Text](#)
16. Perkins AJ, Stump TE, Monahan PO, *et al.*: **Assessment of differential item functioning for demographic comparisons in the MOS SF-36 health survey.** *Qual Life Res.* 2006; **15**(3): 331–48.
[PubMed Abstract](#) | [Publisher Full Text](#)
17. Steinberg L, Thissen D: **Using effect sizes for research reporting: Examples using item response theory to analyze differential item functioning.** *Psychol Methods.* 2006; **11**(4): 402–15.
[PubMed Abstract](#) | [Publisher Full Text](#)
18. Thissen D, Steinberg L, Wainer H: **Detection of differential item functioning using the parameters of item response models.** 1993.
[Reference Source](#)
19. Dandara C, Greenberg J, Lambie L, *et al.*: **Direct-to-consumer genetic testing: To test or not to test, that is the question.** *S Afr Med J.* 2013; **103**(8): 510–2.
[PubMed Abstract](#) | [Publisher Full Text](#)
20. Heckathorn DD: **Respondent-driven sampling: a new approach to the study of hidden populations.** *Soc Probl.* 1997; **44**(2): 174–99.
[Publisher Full Text](#)
21. Heckathorn DD: **Respondent-driven sampling II: deriving valid population estimates from chain-referral samples of hidden populations.** *Soc Probl.* 2002; **49**(1): 11–34.
[Publisher Full Text](#)
22. John M: **Perceptions on precision medicine implementation attributes arranged into categories [Dataset].** *Zenodo.* 2020.
<http://www.doi.org/10.5281/zenodo.3925276>
23. Hunt SD, Sparkman Jr RD, Wilcox JB: **The pretest in survey research: Issues and preliminary findings.** *J Mark Res.* 1982; **19**(2): 269–73.
[Publisher Full Text](#)
24. Mooney CZ: **Monte carlo simulation.** Sage publications; 1997.
[Publisher Full Text](#)
25. Shrout PE, Bolger N: **Mediation in experimental and nonexperimental studies: new procedures and recommendations.** *Psychol Methods.* 2002; **7**(4): 422–45.
[PubMed Abstract](#) | [Publisher Full Text](#)
26. Anderson JC, Gerbing DW: **Predicting the performance of measures in a confirmatory factor analysis with a pretest assessment of their substantive validities.** *J Appl Psychol.* 1991; **76**(5): 732–740.
[Publisher Full Text](#)
27. Angleitner A, John OP, Löhr FJ: **It's what you ask and how you ask it: An itemmetric analysis of personality questionnaires.** *Personality assessment via questionnaires.* Springer; 1986; 61–108.
[Publisher Full Text](#)
28. Optimal-Workshop: **Optimal Workshop.** *Optimalworkshop.com*; 2019; [cited 2019 June 20].
[Reference Source](#)
29. Kean J, Reilly J: **Item response theory.** *Handbook for Clinical Research: Design, Statistics and Implementation.* New York, NY: Demos Medical Publishing. 2014; 195–198.
[Reference Source](#)
30. Stocking ML, Lord FM: **Developing a common metric in item response theory.** *Applied psychological measurement.* 1983; **7**(2): 201–10.
[Publisher Full Text](#)
31. Zampetakis LA, Lerakis M, Kafetsios K, *et al.*: **Using item response theory to investigate the structure of anticipated affect: do self-reports about future affective reactions conform to typical or maximal models?** *Front Psychol.* 2015; **6**: 1438.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
32. Van der Ark LA: **Mokken scale analysis in R.** *J Stat Softw.* 2007; **20**(11): 1–19.
[Publisher Full Text](#)
33. Choi SW, Gibbons LE, Crane PK: **Lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations.** *J Stat Softw.* 2011; **39**(8): 1–30.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
34. Team RC: **R: a language and environment for statistical computing.** *R Foundation for Statistical Computing.* 2016.
[Reference Source](#)
35. Breslau J, Javaras KN, Blacker D, *et al.*: **Differential item functioning between ethnic groups in the epidemiological assessment of depression.** *J Nerv Ment Dis.* 2008; **196**(4): 297–306.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
36. Crane PK, Gibbons LE, Jolley L, *et al.*: **Differential item functioning analysis with ordinal logistic regression techniques: DIFdetect and difwithpar.** *Med Care.* 2006; **44**(11 Suppl 3): S115–S23.
[PubMed Abstract](#) | [Publisher Full Text](#)
37. Jodoin MG, Gierl MJ: **Evaluating type I error and power rates using an effect size measure with the logistic regression procedure for DIF detection.** *Applied measurement in education.* 2001; **14**(4): 329–49.
[Publisher Full Text](#)
38. Hu B, Shao J, Palta M: **Pseudo-R² in logistic regression model.** *Statistica Sinica.* 2006; **16**(3): 847–860.
[Reference Source](#)
39. Smith TJ, McKenna CM: **A comparison of logistic regression pseudo R² indices.** *Multiple Linear Regression Viewpoints.* 2013; **39**(2): 17–26.
[Reference Source](#)
40. Gelin MN, Zumbo BD: **Differential item functioning results may change depending on how an item is scored: An illustration with the Center for Epidemiologic Studies Depression Scale.** *Educational and Psychological Measurement.* 2003; **63**(1): 65–74.
[Publisher Full Text](#)
41. John M: **R code for DIF [r code].** *Zenodo.* 2020.
<http://www.doi.org/10.5281/zenodo.3925336>
42. Revelle W: **Procedures for psychological, psychometric, and personality research.** *Acesso em.* 2012; 9.
[Reference Source](#)
43. Van der Ark LA: **New developments in Mokken scale analysis in R.** *J Stat Softw.* 2012; **48**(5): 1–27.
[Publisher Full Text](#)
44. Palmgren PJ, Brodin U, Nilsson GH, *et al.*: **Investigating psychometric properties and dimensional structure of an educational environment measure (DREEM) using Mokken scale analysis—a pragmatic approach.** *BMC Med Educ.* 2018; **18**(1): 235.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
45. Schmitt N: **Uses and abuses of coefficient alpha.** *Psychol Assess.* 1996; **8**(4): 350–353.
[Publisher Full Text](#)
46. Sijsma K: **On the use, the misuse, and the very limited usefulness of Cronbach's alpha.** *Psychometrika.* 2009; **74**(1): 107–120.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
47. Tavakol M, Dennick R: **Making sense of Cronbach's alpha.** *Int J Med Educ.* 2011; **2**: 53–55.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
48. Sijsma K, Molenaar IW: **Reliability of test scores in nonparametric item response theory.** *Psychometrika.* 1987; **52**(1): 79–97.
[Publisher Full Text](#)
49. Etchegary H, Green J, Parfrey P, *et al.*: **Community engagement with genetics:**

- public perceptions and expectations about genetics research. *Health Expectations*. 2015; **18**(5): 1413–25.
[Publisher Full Text](#)
50. Samuel GN, Farsides B: **Genomics England's implementation of its public engagement strategy: Blurred boundaries between engagement for the United Kingdom's 100,000 Genomes project and the need for public support.** *Public Understanding of Science*. 2018; **27**(3): 352–64.
[Publisher Full Text](#)
51. Wonkam A, Hurst S: **A call for policy action in sub-Saharan Africa to rethink diagnostics for pregnancy affected by sickle cell disease: differential views of medical doctors, parents and adult patients predict value conflicts in Cameroon.** *OMICS*. 2014; **18**(7): 472–80.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
52. Tekola-Ayele F, Rotimi CN: **Translational Genomics in Low- and Middle-Income Countries: Opportunities and Challenges.** *Public Health Genomics*. 2015; **18**(4): 242–7.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
53. Tiffin N: **Unique considerations for advancing genomic medicine in African populations.** *Per Med*. 2014; **11**(2): 187–96.
[PubMed Abstract](#) | [Publisher Full Text](#)

Open Peer Review

Current Peer Review Status:

Version 1

Reviewer Report 03 August 2021

<https://doi.org/10.5256/f1000research.27587.r90224>

© 2021 Asada K. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Ken Asada

Cancer Translational Research Team, RIKEN Center for Advanced Intelligence Project, Tokyo, Japan

Authors focused on the gap between research and practice settings of OBM and stated on the importance in health sciences, which I agreed with. Additionally, DIF is probably a suitable method to use in their study. However, manuscript is difficult to fully understand. Authors need to reconstruct the structure and amend the manuscript to convey their purpose and study smoothly to readers.

Here are my comments.

1. Study design and Procedure should explain more clearly. If needed, please consider using a flowchart.
2. I have a concern about the data size author used for their analyses.
3. Table 1: Please revise carefully if the statements are widely accepted. For example, is it truly "easy to obtain"?
4. Authors stated that "based on existing literatures, six items were selected as indicators for the 'Omics Biomarkers' (OBM) construct as illustrated in Table 1." What literatures do authors refer? Please provide and cite as references. Additionally, please use other abbreviations since OBM has been already used as "Omics based biomarkers" in the early paragraph. OBM and OBMs are quite similar terminology and making confusion. On top of that, I see the following sentences (1) six items were selected as indicators for the 'Omics Biomarkers' (OBM) construct as illustrated in Table 1. (2) Six items were selected as its indicators (Table 1). (3) Six items were selected to be indicative of 'Public Genomic Awareness', as seen in Table 1. (4) Four items were selected as measures of the construct 'Implementation Outcomes' (Table 1). It is of your interest but I suggest that authors should make more clear explanation for Table 1.
5. Abbreviations: Some sentences use abbreviation without explanation before use. On top of that, abbreviations are repeatedly showing up in the manuscript. For example, "differential

item functioning (DIF)" shows up in page 3, page 4, page 6, and page 12. So does IRT. "item response theory (IRT)" has repeatedly appeared, which once in page 5, and twice in page 6. I assume it could be more in other pages. Again, explanation of the abbreviation once is enough.

6. Authors focused on the importance and gap between biomedical research and clinical practice. DIF is the one authors claim to fill in the gap between biomedical research and clinical practice. Actually, some papers have been already reported the advantage of DIF in health sciences. Therefore, please discuss other related works in the discussion section. E.g.; Scott et al. *Health Qual Life Outcomes*. 2010 Aug 4;8:81. doi: 10.1186/1477-7525-8-81¹. Hagquist C. *BMC Medical Research Methodology* 19, 185 (2019). <https://doi.org/10.1186/s12874-019-0828-3>².
7. Research Ethics: Please provide an approved research ethics number for your study.
8. Percentage (%) could be better representation rather than using actual number in Figure 2.
9. It would be better representation if authors invert color palette. Author is now using blue for positive and red for negative but I think blue for negative and red for positive is more familiar with.
10. Please omit following sentence. "From the above analyses, we generated various graphs. However, due to space constraints, we present only those graphs related to characteristics of omics biomarkers (OBM construct)."
11. In page 10 to page 11, authors stated as " Figure 4c indicates a clear difference in TCC curves, albeit in small margins, of DIF item responses between academia and industry groups. On the other hand, the left graph for all items displayed a much smaller difference than the right graph of Figure 4c." In my eyes, the results of all Items and DIF Items are similar.
12. In page 10 to page 11, authors stated as " Figure 4c indicates a clear difference in TCC curves, albeit in small margins, of DIF item responses between academia and industry groups. On the other hand, the left graph for all items displayed a much smaller difference than the right graph of Figure 4c." In my eyes, the results of all Items and DIF Items are similar.

References

1. Scott NW, Fayers PM, Aaronson NK, Bottomley A, et al.: Differential item functioning (DIF) analyses of health-related quality of life instruments using logistic regression. *Health Qual Life Outcomes*. 2010; **8**: 81 [PubMed Abstract](#) | [Publisher Full Text](#)
2. Hagquist C: Explaining differential item functioning focusing on the crucial role of external information – an example from the measurement of adolescent mental health. *BMC Medical Research Methodology*. 2019; **19** (1). [Publisher Full Text](#)

Is the work clearly and accurately presented and does it cite the current literature?

No

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

I cannot comment. A qualified statistician is required.

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Multi-omic analysis, Precision medicine, Machine learning, genome and epigenomics, Cancer research.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Reviewer Report 02 December 2020

<https://doi.org/10.5256/f1000research.27587.r74152>

© 2020 Mboowa G. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Gerald Mboowa

Department of Immunology and Molecular Biology, Makerere University College of Health Sciences, Kampala, Uganda

This article provides insightful background on precision medicine implementation and research-practice partnerships, and suggests implications of measurement scale differential item functioning. The authors describe the methods for precision medicine implementation by employing analytical approaches for quantifying factorial relationships as well as presenting graphical analyses of differential item functioning (DIF) effects.

The authors clearly discuss factors (characteristics of omics biomarkers (OBM), organizational support (OrG) and precision medicine implementation outcomes (ImO)) for DIF. This research

article is scientifically sound.

Minor comment for the authors:

Authors should discuss the limitations of the study design and the future of Omics-based biomarkers (OBMs).

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

I cannot comment. A qualified statistician is required.

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genomics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research