



Development and validation of a measure of comprehension of genomic screening—negative results (CoG-NR)

Gail E. Henderson¹ · Molly Ewing¹ · Kristine J. Kuczynski¹ · R. Jean Cadigan¹ · Margaret Waltz¹ · Rita M. Butterfield² · Christine Rini³ · Karen Weck¹ · Jonathan S. Berg¹ · Teresa P. Edwards¹

Received: 27 December 2019 / Revised: 3 April 2020 / Accepted: 28 April 2020 / Published online: 26 May 2020
© The Author(s), under exclusive licence to European Society of Human Genetics 2020

Abstract

To realize the promise of population genomic screening for rare medically actionable conditions, critical challenges in the return of normal/negative results must be understood and overcome. Our study objective was to assess the functioning of a new 13-item measure (CoG-NR) of understanding of and knowledge about normal/negative genomic screening results for three highly actionable conditions: Lynch Syndrome, Hereditary Breast and Ovarian Cancer, and Familial Hypercholesterolemia. Based on our prior research and expert review, we developed CoG-NR and tested how well it functioned using hypothetical scenarios in three Qualtrics surveys. We report on its psychometric properties and performance across the three different conditions. The measure performed similarly for the three conditions. Examinations of item difficulty, internal reliability, and differential item functioning indicate that the items perform well, with statistically significant positive correlations with genomic knowledge, health literacy, and objective numeracy. CoG-NR assesses understanding of normal/negative results for each of the conditions. The next step is to examine its performance among individuals who have actually undergone such tests, and subsequent use in clinical or research situations. The CoG-NR measure holds great promise as a tool to enhance benefits of population genomic screening by bringing to light the prevalence of incorrect interpretation of negative results.

Introduction

Interest in population genomic screening as a component of precision medicine has increased significantly [1–4]. Many experts now argue that screening for rare clinically actionable genetic conditions will benefit the general population and reduce overall healthcare costs, especially as data are shared across learning healthcare systems [5]. Accordingly, the U.S. National Academies of Sciences, Engineering, and Medicine (NASEM) Roundtable on Genomics and Precision Health ad hoc working group, Genomics and Population Health Action Collaborative (GPHAC), recently published a white paper outlining an approach for population screening based on

medically actionable Tier 1 genes related to Lynch Syndrome, Hereditary Breast and Ovarian Cancer (HBOC), and Familial Hypercholesterolemia (FH) [6]. When identified before onset of symptoms, action can be taken to prevent and/or treat these conditions.

Although screening unselected populations will identify individuals with important positive results, most screened individuals will have normal/negative genomic screening results; that is, no genetic variants associated with these conditions will be identified. Based on recent studies [7–10] and our own prior research [11], we argue that there may be critical challenges and potential harms in the return of normal/negative population screening results, and addressing them is critical for realizing the promise of population genomic screening. As a first step to assess understanding among those who receive normal/negative screening results, we developed the measure described in this article.

In 2019, we reported lessons learned from returning negative results to individuals in a pilot genomic screening program, “GeneScreen” [11]. This study recruited 262 adults from a hospital-based general medicine clinic at the University of North Carolina, Chapel Hill (UNC)

✉ Gail E. Henderson
gail_henderson@med.unc.edu

¹ University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

² Duke University, Durham, NC, USA

³ Northwestern University, Evanston, IL, USA

and the Kaiser Permanente Northwest research biobank. GeneScreen used a targeted sequencing panel of 17 genes related to 11 medically actionable conditions. The study's screening protocol identified fourteen CLIA confirmed positive results; the remainder were "normal/negative."

Our findings raised two concerns about receipt of normal/negative results. First, contrary to expectations that the participants were "healthy" adults, we found disproportionate enrollment of individuals with an elevated prior risk for one of the conditions being screened. Before receiving results, nearly three-quarters of participants reported a personal or family history of at least one condition included on the panel [11]. Having an elevated prior risk increases the likelihood of a falsely reassuring negative screen, particularly when only a subset of relevant genes for a given disease phenotype is being evaluated and the results are limited to pathogenic or likely pathogenic variants. Second, although most participants believed they understood the meaning of their result, 44.3% indicated that their normal/negative result meant they "definitely do not" have a pathogenic variant in a GeneScreen gene. Given the potential for a false negative result, this response indicates a misunderstanding of a negative result, particularly in participants with a prior increased risk of one of the conditions. Over half (51.9%) indicated it was "extremely unlikely" they had a pathogenic variant in a GeneScreen gene, a more accurate response [11]. We conclude that return of normal/negative results in this context may raise nuanced concerns for communicating their meaning [11].

A literature search in PubMed from 2010 to the present, using search terms "understanding," "negative," "genomic," and "screening" found no relevant research or alternative measures capable of assessing whether people understand normal/negative genomic screening results. Consequently, we developed a measure of comprehension of normal/negative screening results. We ultimately chose to create a measure that is applicable to the three highly actionable conditions highlighted in the recent NASEM report as beneficial for screening [6]: Lynch Syndrome, HBOC, and FH. These conditions will continue to be of central importance to public health, but we believe the measure will also be relevant and capable of moving research forward for different conditions. In this paper, we report on the measure's psychometric properties and performance across the three different conditions, and consider next steps in dissemination and use.

Methods

Measure item development

Ideally, we envisioned a measure that would test comprehension of normal/negative results but that would not

be tied to a specific condition (e.g., screening for genetic risk for heart conditions). It became clear, however, that assessing understanding of normal/negative genomic test results across multiple conditions that might be included in a panel would be problematic. We could not ensure that broad reference to "a genomic screening test" would be comprehended and interpreted consistently across conditions and respondents. Therefore, to minimize measurement error, we focused on screening for the three medically actionable conditions identified in the NASEM report.

Based on expert input and concerns identified in our survey of GeneScreen participants who received "normal/negative" results, we developed potential items for the measure. After multiple iterations, making use of expert opinions, testing within our team, and clinical geneticists and public health researchers 13 items were chosen and refined to be below a 7th grade reading level (Fig. 1). They assessed understanding of: (1) features of a genomic screening test, (2) limitations of the test, (3) effect of family history on the possibility of a false negative or falsely reassuring result, and (4) that these diseases are multifactorial.

We administered the resulting "Comprehension of Genomic Screening—Negative Results" (CoG-NR) measure to three Qualtrics panels. Because we would test this measure with the general population rather than individuals with genetic screening experience, we developed an introductory paragraph explaining genetic screening, defining the term "genetic variant," and asking respondents to pretend they had undergone screening for genetic variants related to a particular disease. We present our measure tested for all three conditions.

Participants and procedures

Subjects were participants in the Qualtrics Online Sample, recruited by Qualtrics Research Services [12]. Quotas ensured that aggregate respondent characteristics mirrored the adult U.S. population for age, sex, education, race, and Hispanic origin. The survey was built using Qualtrics web-based survey research software and administered three times, each to a different pool of respondents. The first administration (May 29–June 25, 2018) focused on genetic screening for Lynch Syndrome ("colon cancer"). After examining the results of this initial data collection, we administered the survey to two additional sets of respondents, one focusing on FH ("high cholesterol") (December 12–January 11, 2018) and the other on HBOC ("breast cancer") (females only, December 4–December 18, 2018). In addition to the CoG-NR, subjects were administered additional measures (described below) to use in the validation of the CoG-NR.

The Comprehension of Genomic Screening—Normal/Negative Results (CoG-NR) Measure: Scenario, Items, and Correct Answers

<p>Now, let's pretend you've had genetic screening.</p> <p>Healthcare providers often screen people for diseases. One example is to check blood pressure. Another example is to look at DNA. Most of our DNA is exactly the same as other people's DNA. Only a small percent is different. These differences are called "genetic variants." Most variants do not impact our health. However, a few do increase the risk of certain diseases. Genetic screening is DNA testing that looks for variants related to a high risk of future diseases in healthy people. If these variants are found, doctors can recommend ways to prevent or treat the disease.</p> <p>Pretend that you were screened for genetic variants. These variants cause a much higher than average risk for [colon cancer/high cholesterol/breast cancer]. Your results are "normal/ negative." The test did not find these variants.</p> <p>Based on this pretend "normal/ negative" result, choose a response to each statement. Do not consider any factors from your own true medical history.</p>	
Items	Correct Answer
1. Based on this normal/ negative result, I definitely don't have [colon cancer/high cholesterol/ breast cancer].	False
2. If [colon cancer/high cholesterol/breast cancer] runs in my family, I may still have an increased risk of developing this disease in spite of my normal/ negative result.	True
3. My lifestyle could affect my chances of getting [colon cancer/high cholesterol/breast cancer].	True
4. My normal/ negative result means I will NOT need to have the standard recommended [colon cancer/high cholesterol/breast cancer] screenings.	False
5. My normal/ negative result is 100% accurate.	False
6. Even with this normal/ negative result, I might still develop [colon cancer/high cholesterol/ breast cancer] someday.	True
7. This test can be used to diagnose [colon cancer/high cholesterol/breast cancer].	False
8. This normal/ negative result, means that I am a lot less likely to develop [colon cancer/high cholesterol/breast cancer] than the average person.	False
9. There are causes of [colon cancer/high cholesterol/breast cancer] that are not genetic that could raise my risk.	True
10. There is a small possibility that my result is wrong.	True
11. I may have other genetic variants that this screening test can't detect that cause an increased risk of [colon cancer/high cholesterol/breast cancer].	True
12. My result means that if I develop [colon cancer/high cholesterol/breast cancer] in the future, the cause cannot be genetic.	False
13. Everyone who develops [colon cancer/high cholesterol/breast cancer] has the kind of genetic variant screened for in this test.	False

Fig. 1 The CoG-NR Measure: scenario, items, and correct answers. The comprehension of genomic screening—normal/negative results (CoG-NR) measure: scenario, items, and correct answers.

Per our request, Qualtrics removed responses with poor data quality, defined by straightlining (selecting the same response for all items, implying lack of careful consideration of each item), survey duration <150 s, duplicate responses (identical demographics from same IP address),

and "don't know" responses to more than three-quarters of items in the CoG-NR measure and UNC-GKS scale (described below). These responses were removed prior to meeting quotas that ensured that characteristics mirrored the adult U.S. population.

The CoG-NR measure

Our measure includes the instructions and 13 items shown in Fig. 1. Participants select from three response options: *true*, *false*, and *don't know/not sure*. Our goal was to measure participants' comprehension, similar to a test, so responses were analyzed as correct or incorrect. (We did not aim to measure participants' beliefs or attitudes about their understanding, which would typically be analyzed using the mean or sum of responses provided on a Likert-type scale.) We offered the explicit "don't know" option to minimize guessing. To score, we assigned one point for each correct response and 0 for each incorrect/do not know response and summed the resulting scores to yield scores ranging from 0 to 13.

Genomic knowledge (UNC-GKS)

We measured genomic knowledge with the validated University of North Carolina Genomic Knowledge Scale (UNC-GKS) [13], which includes 19 statements about genes, genetic effects on health, and familial inheritance. Respondents mark each statement as *true*, *false*, or *not sure/don't know* (scored as incorrect). Correct responses were scored as 1 and summed. Possible scores range from 0 (no responses were correct) to 19 (all responses were correct).

Health literacy

We used the five-item subscale from the validated Health Literacy Questionnaire (HLQ) [14] to measure respondents' self-perceived ability to understand health information and know what to do with it. Responses ranged from 1 (Cannot do or always difficult) to 5 (Always easy) on each item (e.g., "Read and understand written health information."). Health literacy scores were calculated by summing responses and could range from 5 to 25 with higher scores indicating stronger perceived understanding of health information.

Numeracy

We measured objective numeracy with a 3-item scale that presents three arithmetic problems testing use of proportions, fractions, and percentages [15]. Objective numeracy scores were calculated by assigning one point for each correct response, and thus could range from 0 (no responses were correct) to 3 (all responses were correct).

Sociodemographic characteristics

Sociodemographic variables were self-reported and included respondents' sex, race/ethnicity, age, and educational attainment.

Data analysis

We required respondents to provide a response to all items to avoid missing data. All analyses were conducted in R version 3.4 except the differential item functioning (DIF) assessment, which was conducted in jMetrik 4.1. The analyses described below were conducted separately for each of the three survey versions, except the analysis of variance (ANOVA) in which the versions were directly compared.

Item-level descriptive statistics

First, we examined the proportion of respondents who correctly answered each question in the CoG-NR measure to identify whether any items might be very easy or difficult (e.g., 90% or more of respondents answering correctly or incorrectly) [13]. Second, we computed an intercorrelation matrix for CoG-NR items to determine whether they were positively associated with one another as we expected. Finally, we calculated the correlation between each item and the total score (minus the item). Positive correlations indicate consistency between the item and the total score (e.g., respondents who do well on the item do well on the measure), whereas negative correlations indicate inconsistency (e.g., respondents who do well on the item do poorly on the measure). Weak or negative correlations would indicate items that need to be reviewed and possibly reworded or discarded.

Classical test theory reliability

We evaluated internal reliability by computing Cronbach's coefficient alpha. A set of items are considered strongly related to one another if alpha values are at least 0.70 [16].

Differential item functioning

DIF analysis allowed us to evaluate whether the measure's items performed similarly across different demographic subgroups (i.e., gender, race/ethnicity). Respondents from different demographic subgroups with equivalent levels of knowledge about normal/negative screening results (operationalized as having the same total scores on the CoG-NR measure) were matched. Matched individuals from different demographic subgroups were then compared regarding how they responded on each item. A DIF: Mantel-Haenszel analysis in jMetrik reports the extent to which items function differently for subgroups using an A, B, C classification system based on a combination of chi-square test p values and common odds ratios [17] in this classification system: A represents no DIF, B represents a moderate amount of DIF, and C represents a large amount of DIF. To ensure enough power in the comparison tests, we examined the number of matched individuals in each subgroup for each total score.

There were cases in which very few members of a subgroup had the lowest scores (i.e., scores of 0–3) and/or highest scores (i.e., 10–13). In these cases, respondents who scored at the low and high ends were collapsed before performing the DIF analysis.

Correlation with related measures

We calculated Pearson correlations between the CoG-NR score and scores on three other measures—the UNC-GKS, the objective numeracy scale, and a subscale of the HLQ—to evaluate the extent to which measures that we hypothesized to be related with each other were, in fact, related. We expected a positive correlation between the CoG-NR measure and UNC-GKS because individuals with more genomic knowledge should be more likely to understand the meaning of a normal/negative genomic screening result. We predicted a positive correlation between the CoG-NR measure and the objective numeracy scale because the ability to reason and apply numerical concepts—particularly regarding probability—would be expected to correlate with an understanding of risk, which is important in understanding screening results. We predicted that the CoG-NR measure would correlate positively with the HLQ subscale because individuals with greater perceived ability to understand health information should better understand the meaning of a normal/negative genetic screening result.

Analysis of variance

We ran an ANOVA to test for significant differences in the mean scores for the colon cancer, high cholesterol, and breast cancer CoG-NR. We expected there would be no differences between these scores (i.e., that participants' understanding of normal/negative results did not vary meaningfully across conditions).

Results

Sample

Demographics for the colon cancer ($n = 506$), high cholesterol ($n = 502$), and breast cancer ($n = 515$) administrations reflected the U.S. population; the breast cancer administration included only females. See Table 1.

CoG-NR Item descriptive statistics

Table 2 shows the percentage of respondents who provided correct, incorrect, or “don't know” responses for each item and each version of the CoG-NR. Looking first at the colon cancer version, the difficulty (proportion of respondents

Table 1 Demographics of respondents in colon cancer, high cholesterol, and breast cancer administration.

	Colon cancer (%)	High cholesterol (%)	Breast cancer (%)
All respondents (n)	506	502	515
Sex			
Female	51	52	100
Male	49	48	0
Race/ethnicity			
Hispanic	19	22	17
White non-Hispanic	60	59	62
Black non-Hispanic	11	11	12
Asian non-Hispanic	6	7	5
Other non-Hispanic	3	1	4
Education			
HS/GED	41	38	41
Some college	21	23	20
College degree	27	28	28
Graduate degree	11	11	12
Age			
18–24	13	12	12
25–34	19	19	18
35–54	34	34	34
55+	35	34	36

answering correctly) ranged from 0.21 to 0.83, with a mean difficulty across the items of 0.56 ($SD = 0.19$). For the high cholesterol version, the range of difficulty was 0.26–0.84 and the mean difficulty across the items was 0.59 ($SD = 0.2$). For the breast cancer version, the range of difficulty was 0.21–0.83 and the mean difficulty across the items was 0.57 ($SD = 0.19$). Table 2 shows responses for each item for each version sorted by items with the lowest number of respondents answering correctly (most difficult) to items with the highest number of respondents answering correctly (least difficult) on the colon cancer version. No items had more than 90% of respondents respond correctly or incorrectly, indicating no items were too easy or difficult.

All correlations among items in the colon cancer version were positive. In addition, all correlations between items and total scores (minus the item) were positive—with a mean of $r = 0.34$ ($SD = 0.09$) and range of 0.15–0.46 for the colon cancer version—meaning that individuals who performed well on the item also performed well on the measure overall, and vice versa. We reviewed two items—item 7 and item 8—that had relatively low item-total correlations (under 0.2), but determined that the items and corresponding responses still provided valuable information and possibly represented common misconceptions, which could explain the relatively low item-total correlations. As such, all items were retained for the high cholesterol and

Table 2 Percentage of respondents who answered correct, incorrect, or “don’t know” for each item and each version.

Item		% Correct	% Incorrect	% DK	
1.	Based on this normal/ negative result, I definitely do not have [colon cancer/high cholesterol/breast cancer].	CC	21	60	19
		HC	28	53	19
		BC	21	61	18
7.	This test can be used to diagnose [colon cancer/high cholesterol/breast cancer].	CC	32	50	18
		HC	26	57	17
		BC	32	46	22
8.	This normal/negative result, means that I am a lot less likely to develop [colon cancer/high cholesterol/breast cancer] than the average person.	CC	33	45	22
		HC	35	42	23
		BC	34	44	22
12.	My result means that if I develop [colon cancer/high cholesterol/breast cancer] in the future, the cause cannot be genetic.	CC	43	29	28
		HC	46	29	25
		BC	46	26	28
5.	My normal/negative result is 100% accurate.	CC	47	25	28
		HC	45	24	31
		BC	51	22	27
13.	Everyone who develops [colon cancer/high cholesterol/breast cancer] has the kind of genetic variant screened for in this test.	CC	47	22	31
		HC	52	21	27
		BC	50	19	31
11.	I may have other genetic variants that this screening test can not detect that cause an increased risk of [colon cancer/high cholesterol/breast cancer].	CC	63	14	23
		HC	68	10	22
		BC	66	10	24
10.	There is a small possibility that my result is wrong.	CC	68	14	18
		HC	67	13	20
		BC	71	13	16
4.	My normal/negative result means I will NOT need to have the standard recommended [colon cancer/high cholesterol/breast cancer] screenings.	CC	70	17	13
		HC	72	15	13
		BC	75	13	12
3.	My lifestyle could affect my chances of getting [colon cancer/high cholesterol/breast cancer].	CC	73	14	13
		HC	84	10	6
		BC	62	21	17
9.	There are causes of [colon cancer/high cholesterol/breast cancer] that are not genetic that could raise my risk.	CC	75	11	14
		HC	80	9	11
		BC	72	11	17
2.	If [colon cancer/high cholesterol/breast cancer] runs in my family, I may still have an increased risk of developing this disease in spite of my normal/negative result.	CC	77	13	10
		HC	77	12	11
		BC	80	11	9
6.	Even with this normal/negative result, I might still develop [colon cancer/high cholesterol/breast cancer] someday.	CC	83	9	8
		HC	85	8	7
		BC	83	8	9

CC colon cancer, HC high cholesterol, BC breast cancer.

breast cancer versions for which similar patterns were found across all items.

Classical test theory reliability

The internal consistency reliability estimate—Cronbach’s alpha (α)—for the 13-item CoG-NR was similar and

adequate for each of three versions ($\alpha_{\text{Colon cancer}} = 0.72$, $\alpha_{\text{High cholesterol}} = 0.72$; and $\alpha_{\text{Breast cancer}} = 0.73$).

Differential item functioning

We noticed some differences in mean CoG-NR scores based on demographics. Different scores are not necessarily

Table 3 Items that had a DIF classification of B.

Item	Version	Focus group	Reference group	Classification	Valid <i>N</i>
				+ item favors the focal group	
				– item favors the reference group	
2	Colon cancer	White	Black	B+	366
	High cholesterol	White	Hispanic	B–	404
5	Colon cancer	White	Hispanic	B–	384
6	Colon cancer	Female	Male	B+	385
	Colon cancer	White	Hispanic	B+	301
	High cholesterol	Female	Male	B+	422
8	Colon cancer	White	Black	B–	355
	Colon cancer	White	Hispanic	B–	384
	High cholesterol	White	Black	B–	338
10	Colon cancer	White	Hispanic	B+	337
	High cholesterol	White	Hispanic	B+	400
11	Breast cancer	White	Hispanic	B–	409
12	High cholesterol	Female	Male	B–	502

Table 4 Pearson correlations between the CoG-NR score and other measure scores.

	Colon cancer	High cholesterol	Breast cancer
Correlation to genomic knowledge scale score (GKS)	0.49*	0.52*	0.51*
GKS alpha coefficient	0.75	0.79	0.76
Correlation corrected for attenuation	0.66	0.70	0.69
Correlation to health literacy scale score (HLS)	0.15*	0.17*	0.15*
HLS alpha coefficient	0.91	0.88	0.89
Correlation corrected for attenuation	0.19	0.21	0.18
Correlation to objective numeracy scale score (ONS)	0.28*	0.27*	0.32*
ONS alpha coefficient	0.49	0.53	0.47
Correlation corrected for attenuation	0.47	0.43	0.54

* $p < 0.001$

concerning as there may be construct-relevant reasons that one group would outperform another on any particular item. However, it is important that each item functions equally well for different groups of respondents; in other words, items should not favor one group over another. To examine this, we performed a DIF analysis comparing groups for which we noted differences in scores might be construct-irrelevant. Specifically, we explored the differences we saw between males and females, between non-Hispanic white respondents and Hispanic respondents, and between non-Hispanic white respondents and non-Hispanic black respondents within each survey administration. This resulted in 104 tests across the 13 items and 3 administrations. Out of these 104 tests, 91 received a DIF classification of A, meaning the item did not favor one group over another in that test. Thirteen tests were assigned a classification of B, meaning the item moderately favored one group over another. The DIF was unsystematic. Among the seven items that scored a B on any test (shown in

Table 3), three were for only one comparison group in one of the three administrations; others were classified B for two or three tests. The moderate and unsystematic favoring of a limited set of items was not concerning regarding the functioning of the items or measure as a whole. Items with moderate DIF should be replaced if comparable items with no DIF are available, but can remain if no such substitutions exist [17], as was the case for our measure.

Correlation with related measures

As shown in Table 4, the CoG-NR measure correlated positively with genetic knowledge, general health literacy, and objective numeracy, providing evidence of validity based on hypothesized relationships. The pattern of correlations indicates that, as expected, our measure is related to genetic knowledge, health literacy, and numeracy but that it assesses a distinct underlying concept.

Analysis of variance

The mean scores for the colon cancer, high cholesterol, and breast cancer versions of the measure were 7.4, 7.6, and 7.5 out of 13 total points, respectively. These means were not statistically different from one another ($F(2, 1520) = 0.468$, $p = 0.626$). Along with the other results reported which were similar across the three genetic conditions (e.g., similar Cronbach's alphas, similar levels of difficulty by item), these results lend evidence that the CoG-NR performs similarly for the three genetic conditions.

Discussion

Our objective was to assess the functioning of our new 13-item CoG-NR measure. We tested it first pertaining to screening for genetic variants associated with increased risk of colon cancer. After determining that it functioned well for this condition, we tested it for genetic variants associated with increased risk of high cholesterol and breast cancer, respectively. The measure performed similarly for the three conditions. Examinations of item difficulty, internal reliability, and DIF indicated that the 13 items perform well. In a few tests, we found DIF for gender or race/ethnicity subgroups but the DIF was present in only a small subset of the tests, was unsystematic across the three versions, and was of only moderate severity (class B). We do not consider these findings to raise concerns regarding the functioning of the items or the measure as a whole, but in future administrations of the CoG-NR we will repeat all the evaluations of the measure's performance, with particular attention to DIF results.

As expected, all versions of the CoG-NR were positively associated with genomic knowledge, health literacy, and objective numeracy. A perfect correlation would indicate that our measure was capturing exactly the same thing as these other measures, which was not our expectation. Rather, as predicted, respondents who had greater genetic knowledge better understood the meaning of negative screening results, but that understanding of normal/negative results is likely different from understanding genetics, per se. Similarly, persons who self-reported higher health literacy or demonstrated better objective numeracy were more likely to score well on the CoG-NR, consistent with our hypotheses since both of those skills would likely aid someone in understanding the meaning of negative results. Together, our findings indicate that the CoG-NR is helpful to assess understanding of normal/negative results for each of the conditions recommended as part of Tier 1 population screening [6].

In this paper, we address the need for this new measure, as well as decisions made on how to design the measure, and justifications for its scoring. In this inaugural test of CoG-NR,

we used volunteer research subjects who were asked to pretend they had undergone genetic screening. The next step is to examine its performance among individuals who have actually undergone such tests, through debriefing or cognitive interviews. This will permit further evaluation of how the items function, testing each item to ensure it is interpreted as expected. If it performs well in subsequent "real-life" administrations, we envision its use in clinical or research situations where screening test providers may be concerned about whether patients or research subjects correctly understand the implications of their negative screening results.

Based on our own research and concerns from the literature about understanding of normal/negative results [7], we believe there is a need for education about the correct interpretation of negative genetic screening results and the risk of developing disease despite negative results. The content of such education, and how it should be delivered so that people understand risk appropriately, remain to be determined. Our CoG-NR measure may help answer these questions by evaluating the level of comprehension among subjects given alternative educational materials or modalities. It could also be used to develop and evaluate educational materials for patients and research subjects.

An ongoing challenge of using this measure is that genomic screening panels often provide results of tests for variants associated with multiple conditions. As described above, we began with separate evaluations of colon cancer, breast cancer, and high cholesterol because of the complexities related to measuring understanding screening results for multiple diseases at once. After establishing that CoG-NR performs well for the three conditions, we plan to develop a modified version of the measure and conduct cognitive interviews with individuals who receive normal/negative results from a screening panel, to assess whether they are able to understand their implications.

Though still early in its development, we believe the CoG-NR measure holds great promise as a tool to enhance benefits of population genomic screening by bringing to light the prevalence of incorrect interpretation of negative results. By doing so, we hope it will motivate and help evaluate educational efforts to counter misunderstanding among recipients of negative results. In its debut, the measure showed encouraging psychometric results. We welcome its use by others for further testing and evaluation.

Acknowledgements The authors wish to acknowledge: James P. Evans, MD, PhD, Professor Emeritus, UNC School of Medicine, for his significant contributions to this work, and Gregory Cizek, PhD, Professor of Educational Measurement and Evaluation, UNC School of Education, for his methodological guidance.

Funding This work was supported by the ELSI Research Program of the National Human Genome Research Institute of the National Institutes of Health, Grant Number P50HG004488.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Williams MS, Buchanan AH, Davis FD, Faucett WA, Hallquist MLG, Leader JB, et al. Patient-centered precision health in a learning health care system: geisinger's genomic medicine experience. *Health Aff.* 2018;37:757–64.
- May T, Nakano-Okuno M, Kelley WV, East K, Moss IP, Sodeke S, et al. Return of raw data in genomic testing and research: ownership, partnership, and risk-benefit. *Genet Med.* 2020;22:12–4.
- National Institutes of Health. 2019. <https://allofus.nih.gov/about/about-all-us-research-program>.
- Evans JP, Powell BC, Berg JS. Finding the rare pathogenic variants in a human genome. *JAMA* 2017;317:1904–5.
- Feero WG, Wicklund CA, Veenstra D. Precision medicine, genome sequencing, and improved population health. *JAMA* 2018;319:1979–80.
- Murray MF, Evans JP, Angrist M, Chan K, Uhlmann W, Doyle DL, et al. A Proposed Approach for Implementing Genomics-Based Screening Programs for Healthy Adults. NAM Perspectives. Discussion Paper, National Academy of Medicine, Washington, DC. 2018. <https://doi.org/10.31478/201812a>.
- Nightingale BM, Hovick SR, Brock P, Callahan E, Jordan E, Roggenbuck J, et al. Hypertrophic cardiomyopathy genetic test reports: A qualitative study of patient understanding of uninformative genetic test results. *J Genet Couns.* 2019;28:1087–97.
- Guan Y, Condit CM, Escoffery C, Bellcross CA, McBride CM. Do women who receive a negative BRCA1/2 risk result understand the implications for breast cancer risk? *Public Health Genom.* 2019;22:102–9.
- Wittman AT, Hashmi SS, Mendez-Figueroa H, Nassef S, Stevens B, Singletary CN. Patient perception of negative noninvasive prenatal testing results. *AJP Rep.* 2016;6:e391–e406.
- Michie S, Smith JA, Senior V, Marteau TM. Understanding why negative genetic test results sometimes fail to reassure. *Am J Med Genet Part A.* 2003;119a:340–7.
- Butterfield RM, Evans JP, Rini C, Kuczynski KJ, Waltz M, Cadigan RJ, et al. Returning negative results to individuals in a genomic screening program: lessons learned. *Genet Med* 2019;21:409–16.
- Qualtrics. 2019 ed. (Qualtrics, Provo, UT). 2019.
- Langer MM, Roche MI, Brewer NT, Berg JS, Khan CM, Leos C, et al. Development and validation of a genomic knowledge scale to advance informed decision making research in genomic sequencing. *MDM Policy Pract.* 2017;2:2381468317692582. <https://doi.org/10.1177/2381468317692582>.
- Osborne RH, Batterham RW, Elsworth GR, Hawkins M, Buchbinder R. The grounded psychometric development and initial validation of the health literacy questionnaire (HLQ). *BMC Public Health.* 2013;13:658.
- Schwartz LM, Woloshin S, Black WC, Welch HG. The role of numeracy in understanding the benefit of screening mammography. *Ann Intern Med.* 1997;127:966–72.
- Streiner D, Norman G, Cairney J. Health measurement scales: a practical guide to their development and use. Oxford, United Kingdom: Oxford University Press; 2015. p. 399.
- Clauser B, Mazor K. Using statistical procedures to identify differentially functioning test items. *Educ Meas Issues Pract.* 2005;17:31–44.