



Expert and lay perspectives on burden, risk, tolerability, and acceptability of clinical interventions for genetic disorders

Ryan S. Paquin, PhD¹, Kathleen F. Mittendorf, PhD², Megan A. Lewis, PhD¹,
Jessica Ezzell Hunter, MS, PhD², Kristy Lee, MS, CGC³, Jonathan S. Berg, MD, PhD³,
Marc S. Williams, MD⁴ and Katrina A. B. Goddard, PhD²

Purpose: The Clinical Genome Resource (ClinGen) Actionability Working Group (AWG) developed a semiquantitative scoring metric to rate clinical actionability of genetic disorders and associated genes in four domains: (1) severity of the outcome, (2) likelihood of the outcome, (3) effectiveness of the intervention to prevent/minimize the outcome, and (4) nature of the intervention with respect to burden, risk, tolerability, and acceptability to the patient. This study aimed to assess whether nature of the intervention scores assigned by AWG experts reflected lay perceptions of intervention burden, risk, tolerability, and acceptability given the subjectivity of this domain.

Methods: In July 2017, a general population sample of 1344 adults completed the study. Each participant was asked to read 1 of 24 plain language medical intervention synopses and answer questions related to its burden, risk, tolerability, and acceptability. We conducted three multilevel mixed model analyses predicting the

perceived burden, perceived risk, and perceived overall nature of the intervention.

Results: As AWG nature of the intervention scores increased, lay perceptions of intervention burden and risk decreased, and perceptions of tolerability and acceptability increased.

Conclusion: The findings show alignment between the ClinGen actionability scoring metric and lay perceptions of the nature of the intervention.

Genetics in Medicine (2019) 21:2561–2568; <https://doi.org/10.1038/s41436-019-0524-z>

Keywords: clinical actionability; secondary findings; clinical decision-making; patient perspectives; precision medicine

INTRODUCTION

The Clinical Genome Resource (ClinGen) is a consortium funded by the National Institutes of Health (NIH) that aims to define the clinical relevance of genes and variants for use in precision medicine and research (www.clinicalgenome.org).¹ The ClinGen Actionability Working Group (AWG)—a panel of experts comprising medical geneticists and genetic counselors—developed a systematic framework to synthesize evidence concerning genetic disorders and associated genes with respect to clinical actionability of pathogenic variants identified through genome-scale sequencing in clinical care. In this context, clinical actionability is defined narrowly in terms of the availability of clinical interventions that effectively prevent or delay onset of genetic disorders, reduce clinical burden, or improve clinical outcomes.

To evaluate the clinical actionability of gene-disorder pairs, the AWG created a semiquantitative scoring metric (SQM) to rate the severity and likelihood of the outcome to be

prevented, the effectiveness of the intervention, and the nature of the intervention.² The latter domain is intended to capture how patients experience the intervention. The SQM is designed to be used by the AWG to identify genetic disorders with the greatest potential for clinical intervention when detected in previously undiagnosed adults. This, in turn, allows the scoring system to address a key component of clinical utility and provide medical decision makers with guidance concerning the return of secondary findings in the context of genomic sequencing and their potential for improved outcomes. Details about the development of the SQM and the related clinical actionability curation procedures have been published elsewhere.²

Conceptually, the nature of the intervention domain is an assessment of the acceptability of a procedure or treatment in terms of its intensity, difficulty, or tolerability and “the burdens or risks placed on the individual.”² The AWG scores nature of the intervention as a four-level variable: 0

¹Center for Communication Science, RTI International, Research Triangle Park, NC, USA; ²Center for Health Research, Kaiser Permanente Northwest, Portland, OR, USA;

³Department of Genetics, University of North Carolina School of Medicine, Chapel Hill, NC, USA; ⁴Genomic Medicine Institute, Geisinger, Danville, PA, USA. Correspondence: Ryan S. Paquin (rpaquin@rti.org)

Submitted 11 November 2018; accepted: 15 April 2019

Published online: 26 April 2019

(high risk, poorly acceptable, or intensive intervention), 1 (greater risk, less acceptable, and substantial intervention), 2 (moderate risk, moderately acceptable, or intensive intervention), and 3 (low risk, medically acceptable, and low-intensity intervention). This component of the SQM introduces a subjective aspect to actionability scoring and it is unclear whether expert opinion adequately reflects lay perspectives of this domain.

The subjectivity is attributable to the highly contextual character of intervention risks or burdens when viewed from the perspective of patients, potential patients, or caregivers. For example, an intervention requiring an individual to avoid contact sports may be viewed differently depending on the preference of an individual toward sports. Although the SQM is intended to reflect the views and preferences of people who use or undergo clinical interventions, there is no direct evidence that nature of the intervention scores assigned by the AWG reflect lay perceptions of burden, risk, tolerability, and acceptability.

It is also unclear whether the characteristics of intervention burden, risk, tolerability, and acceptability are sufficiently interdependent to warrant representing these dimensions with a single intervention-level summary score. In the SQM coding procedure, interventions with lower risk of harm, that are less burdensome or intensive, and that are more acceptable or tolerable are assigned higher nature of the intervention values. In principle, the dimensions underlying nature of the intervention may vary independently—such as an objectively low-risk, high-burden intervention—even if in practice we would expect these dimensions to be interrelated for most interventions. Smoking cessation is one such intervention that is low risk from a medical perspective but may be viewed as burdensome or unacceptable from the perspective of someone who has been advised to quit. The AWG included smoking cessation in the actionability curation summary for α -1 antitrypsin deficiency because of evidence that it lowers the risk of chronic obstructive pulmonary disease caused by that genetic condition.³ In fact, the AWG downgraded the original nature of the intervention score for this intervention from the most favorable rating (i.e., from a score of 3 to 2) after substantial discussion about how to weigh nicotine dependence and potential withdrawal symptoms with low medical risk. To the extent that perceptions of burden, risk, tolerability, and acceptability do not covary, the significance of nature of the intervention scores as a summary statement of these characteristics would be diminished. If, for example, perceptually low-risk but high-burden interventions are common enough to form the rule rather than an exception, revising the SQM to gather separate scores for burden and risk may be necessary.

To address these challenges, we examined associations between nature of the intervention scores previously assigned by the AWG and the perceived burden, risk, tolerability, and acceptability of these interventions according to a general population sample of research participants. If AWG nature of the intervention scores reflect the perspective of nonexperts, we would expect lay perceptions of burden and perceived risk to decrease as AWG scores increase. Similarly, we would expect perceptions of intervention tolerability and

acceptability to increase along with AWG nature of the intervention scores.

MATERIALS AND METHODS

Participants

The participants were 1344 US adults who consented and completed the study from 21 to 26 July 2017. Qualtrics, a research services company, recruited participants and hosted the online study. Eligible participants were English-speaking adults, at least 18 years of age, living in the United States. Potential participants were selected through third-party, actively managed research panels and invited by email to participate in our study. The panel vendors use various incentive structures and forms of compensation, such as points redeemable for merchandise or cash. For this project, the incentive did not exceed \$6.00 in monetary value. The study protocol received institutional review board (IRB) review and approval from the Committee for the Protection of Human Subjects at RTI International.

Procedure

Participants followed a link in the recruitment email to an online survey consisting of 76 questions. We asked participants demographic and general health-related questions, and then randomly assigned each participant to read 1 of 24 plain language intervention synopses (see Supplementary Materials and Methods, Appendix A). The interventions had been previously scored by the AWG and represented broad intervention categories, such as pharmaceutical treatments as compared with prophylactic organ-removing surgery (Table 1). When selecting interventions for this study, we aimed to include several examples from each category and to cover all possible AWG nature of the intervention scores. The synopses reflected health-care delivery system patient materials and the summary reports used by the AWG as part of their scoring protocol. To develop the synopses, we created an outline template and gathered information fitting broad categories that apply when describing interventions (e.g., procedure/regimen, dosage/schedule, duration, recovery, adverse reactions, etc.). Initial drafts were revised to improve cohesion in writing style, edited to meet plain language standards, and submitted to expert review by members of the AWG to ensure accuracy of the final content. The synopses were written in the second person and participants were instructed to imagine that their doctor had recommended the intervention as the best-available course of action to address a health condition. To avoid foreseeable barriers of participants placing themselves in an irrelevant intervention scenario, only female participants were assigned to synopses describing double mastectomy or mammography, and only current or former smokers⁴ were assigned to the smoking cessation scenario. Female participants and current or former smokers were identified using questions asked prior to synopsis assignment and these data were used to set parameters of the randomization program. After reading the synopsis, participants rated the intervention on multiple scales designed to measure constructs related to

Table 1 AWG nature of the intervention score and sample size by intervention

Category	Intervention	AWG NOI	n	%
Avoidance	Smoking cessation ^a	2	60	4.5
	Sun avoidance	3	54	4.0
	Avoid participation in competitive sports	2	55	4.1
	Phenylalanine restricted diet	2	60	4.5
	Avoid fasting and eat a special high carbohydrate diet	3	58	4.3
	Avoid "triggering" anesthetics	3	54	4.0
Surgery	Total colectomy	1	56	4.2
	Double mastectomy ^b	1	47	3.5
	Total thyroidectomy	1	55	4.1
	Total gastrectomy	0	56	4.2
Surveillance, noninvasive imaging	Mammography	3	49	3.6
	Annual thyroid ultrasound	3	57	4.2
	Cerebral MRI	3	58	4.3
	Periodic renal CT scan	2	56	4.2
Surveillance, noninvasive blood draw	Blood draw	3	56	4.2
Surveillance, invasive	Routine colonoscopy	2	55	4.1
Drug therapy	Daily aspirin	3	58	4.3
	β-blocker	3	57	4.2
	Statins	3	59	4.4
	Sulfonylurea	3	59	4.4
Other	Regular LDL apheresis	2	59	4.4
	Therapeutic phlebotomy	3	55	4.1
	ICD placement	2	56	4.2
	Physical/occupational therapy	3	55	4.1

n = 1344.

^aOnly participants who reported smoking at least 100 cigarettes in their entire life were randomized to the smoking cessation intervention.

^bOnly women were randomized to the double mastectomy or mammography arms of the study.

AWG NOI nature of the intervention score assigned by the ClinGen Actionability Working Group, CT computed tomography, ICD implantable cardioverter defibrillator, LDL low-density lipoprotein, MRI magnetic resonance image.

perceived nature of the intervention. At the end of the study, participants were shown a debriefing screen in the online questionnaire reminding them that the interventions do not really apply to them.

Measures

Dependent variables

To align conceptually with how nature of the intervention is defined in the SQM, we used three outcome measures to

assess lay perceptions of intervention burden, risk, and overall nature (i.e., acceptability, tolerability, and favorability).

Perceived burden

We measured perceived burden of the intervention using five Likert-type items with response options ranging from 1 (strongly disagree) to 5 (strongly agree): (a) "This intervention would take a lot of my time," (b) "This intervention would take a lot of effort," (c) "It would be hard for me to do this intervention," (d) "This intervention would get in the way of important things in my life," and (e) "I would have difficulty finding time to do this intervention." The questions were adapted from a portion of the Illness Management Survey (IMS);⁵ some of the questions overlap with the Treatment Satisfaction Questionnaire for Medications (TSQM).⁶ We averaged the five items to create a composite scale ($\alpha = 0.91$, $M = 2.9$, $SD = 1.2$).

Perceived risk

We used six 5-point Likert-type items derived from the IMS, the TSQM, and the Drug Therapy Concerns Questionnaire⁷ to measure perceived risk from intervention side effects. Three of the risk items were measured on the same response scale, with options ranging from 1 (strongly disagree) to 5 (strongly agree): (a) "This intervention has side effects that I really don't like," (b) "The potential side effects of this intervention worry me," and (c) "The potential side effects of this intervention would be embarrassing." The remaining three items had different response scales: (d) "How bothered are you by the potential side effects of this intervention" (1 [not at all bothered] to 5 [extremely bothered]); (e) "If you did this intervention, how likely is it that you would have at least one side effect?" (1 [very unlikely] to 5 [extremely likely]); and (f) "If you did this intervention and had any side effects they would be..." (1 [very minor] to 5 [very serious]). We created a composite perceived risk scale using the average score over the six items ($\alpha = 0.89$, $M = 2.9$, $SD = 1.0$).

Perceived overall nature of the intervention

We measured perceived overall nature of the intervention with three 4-point Likert-type items focusing on intervention acceptability, tolerability, and overall favorability: (a) "In your opinion, how acceptable is this intervention?" (0 [not acceptable] to 3 [highly acceptable]); (b) "In your opinion, how tolerable is this intervention?" (0 [not tolerable] to 3 [highly tolerable]); and (c) "The overall nature of this intervention is..." (0 [extremely bad] to 3 [extremely good]). These items also had a "Don't know" response option, which was treated as missing for this analysis. Because these items had only four response options, we computed the ordinal α coefficient to more accurately estimate interitem reliability ($\alpha = 0.87$) (ref. ⁸). We combined these items into a single scale by computing the mean score for cases having at least one valid response ($M = 1.9$, $SD = 0.7$, $n = 1313$).

Independent variables and covariates

AWG nature of the intervention scores

The primary purpose of this analysis was to assess lay perceptions of intervention burden, risk, and acceptability in relation to nature of the intervention consensus scores assigned by the AWG. The scoring results for all outcome-intervention pairs classified to date are available online at https://actionability.clinicalgenome.org/redmine/projects/actionability_release/genboree_ac/ui. Nature of the intervention scores fall on a 4-point ordinal scale, with values ranging from 0 (high risk, poorly acceptable, or intensive) to 3 (low risk, medically acceptable, or low-intensity). Prior to analysis, we rescaled the scores for the 24 interventions used in this study into three dummy variables, holding the lowest value as the reference category.

Intervention-level (level 2) covariates

We statistically controlled for additional intervention-level variables to account for other features of the intervention not captured by the AWG consensus scores that might influence perceptions of burden, risk, and acceptability. Although we limited the length of the intervention synopses to under one single-spaced page of text, some interventions required longer descriptions than others. We included both the word count ($M = 321.1$, $SD = 125.1$) and readability of the intervention descriptions as covariates ($M = 8.2$, $SD = 1.5$). We used the SMOG Index as a measure of readability,^{9, 10} which we assessed using an online calculator developed by the University of Nottingham, UK.¹¹ We subtracted a constant from these scores to return the scale to US grade levels. We also included two variables measuring the percentage of participants in each synopsis group who reported having either personal experience with the intervention or a close friend or family member who had done the intervention before. We grand mean centered these level 2 covariates prior to analysis.^{12, 13}

Participant-level (level 1) covariates

We collected numerous participant-level variables that we included in the models, including age, sex, race/ethnicity, health literacy, attention to the synopsis, familiarity with the intervention, and personal or family/friend experience with the intervention. See Supplementary Materials and Methods, Appendix B for measurement details concerning participant-level covariates. Categorical variables were dummy coded, and all participant-level covariates were grand mean centered for the main analysis.

Statistical analysis

We designed the study so that participants were nested within intervention groups, yielding a cross-sectional multilevel design. We used a mixed modeling approach^{12, 14} to assess variation in perceptions of intervention burden, risk, and overall nature of the intervention in relation to characteristics of the interventions (i.e., level 2) and of participants (i.e., level 1). Multilevel modeling allowed us to estimate how much of

the variance in perceived burden, risk, and acceptability was explained by AWG nature of the intervention scores at the intervention level; that is, how strongly AWG scores are associated with average patient perceptions attributable to the intervention synopsis. The portion of the models focusing on participant-level characteristics allowed us to assess the degree to which variation in the outcome variables was caused by individual differences rather than a response to the intervention synopses. We used a hierarchical model building approach, beginning with an intercept-only model, systematically adding variance components and predictors in subsequent models, and conducting likelihood ratio tests to assess the relative fit of more complex to simpler models at each step.¹² Only the final models are presented in this report. We followed up on the significant main effects of AWG nature of the intervention scores by computing estimated marginal means and conducting two-tailed unadjusted pairwise comparisons across levels. We conducted all analyses using Stata 15.0.¹⁵

RESULTS

Participant characteristics and descriptive statistics

Participant characteristics are summarized in Table 2. Participants ranged in age from 18 to 93 years ($M = 44.8$, $SD = 17.1$). A small majority of participants were women (55.4%, $n = 745$) and about three-quarters of participants identified as non-Hispanic white (75.2%, $n = 1010$). Almost half of participants had attained an associate degree or higher (47.5%, $n = 638$). Nearly a third of participants reported an annual household income between \$25,000 and \$49,999. A majority of participants reported giving at least moderate attention to the intervention synopsis (63.8%, $n = 857$) and being at least moderately familiar with it (66.6%, $n = 895$). Gastrectomy had the smallest proportion of participants who reported medium or high familiarity with the intervention (37.5%, $n = 21$), whereas mammography had the largest proportion of participants who reported medium or high familiarity with the intervention (93.9%, $n = 46$). Nearly two-thirds of participants reported having no direct or indirect experience with the intervention about which they were assigned to read (65.0%, $n = 874$). Mammography had the smallest proportion of participants reporting no experience with the intervention (30.6%, $n = 15$), whereas implantable cardioverter defibrillator (ICD) placement had the largest proportion of participants reporting no experience with the intervention (89.3%, $n = 50$).

Multilevel models predicting perceived burden, risk, and overall nature of the intervention

We found medium to large correlations¹⁶ among perceived burden, risk, and perceived overall nature of the intervention ($r_{\text{burden-risk}} = 0.59$, $n = 1344$, $P < 0.001$; $r_{\text{burden-nature of the intervention}} = -0.43$, $n = 1313$, $P < 0.001$; $r_{\text{risk-nature of the intervention}} = -0.38$, $n = 1313$, $P < 0.001$). To assess whether multilevel modeling was needed for each of these outcomes, we computed the intraclass correlation coefficient

Table 2 Participant characteristics

Variable	<i>n</i>	%
Age		
18–29	327	24.3
30–44	376	28.0
45–59	314	23.4
60 or older	327	24.3
Sex		
Male	599	44.6
Female	745	55.4
Race/ethnicity		
White, non-Hispanic	1010	75.1
Black, non-Hispanic	115	8.6
Hispanic, all races	120	8.9
Other, non-Hispanic	89	6.6
Refused	10	0.7
Educational attainment		
High school or less	376	28.0
Some college or technical school	330	24.6
Associate degree	187	13.9
Bachelor's degree	302	22.5
Advanced or postgraduate degree	149	11.1
Annual household income		
Less than \$25,000	298	22.2
\$25,000–\$49,999	402	29.9
\$50,000–\$74,999	261	19.4
\$75,000–\$99,999	171	12.7
\$100,000 or more	168	12.5
Refused	44	3.3
Attention to the synopsis		
Low	487	36.2
Medium	564	42.0
High	293	21.8
Intervention familiarity		
Low	449	33.4
Medium	577	42.9
High	318	23.7
Intervention experience		
None	874	65.0
Close friend or family member	191	14.2
Self	279	20.8

n = 1344. Percentages may not sum to 100 due to rounding.

(ICC) and design effect (DE) of our variance components models to verify that variance occurred both across participants (i.e., level 1) and across intervention groups (i.e., level 2). The variance components model included no predictors at either level and was used to estimate the portion of unexplained variance in the dependent variables remaining at each level, after accounting for the grand mean and the mean in each intervention synopsis group. All three models had a DE greater than 6, which is evidence that multilevel modeling is appropriate.¹⁷ Sixteen percent

of the unexplained variance in perceived burden scores occurred across intervention synopsis groups (ICC = 0.16, SE = 0.04, 95% CI [0.09, 0.26], DE = 9.79); whereas the remaining 84% of the variance in that outcome occurred and can be explained at the participant level. Thirteen percent of the variance in perceived risk can be explained at the intervention synopsis level (ICC = 0.13, SE = 0.04, 95% CI [0.07, 0.22], DE = 7.99), whereas the remaining 87% of the variance in that outcome occurred at the participant level. Ten percent of the unexplained variance in perceived overall nature of the intervention occurred across intervention groups (ICC = 0.10, SE = 0.03, 95% CI [0.05, 0.17], DE = 6.19), whereas the remaining 90% of the variance in that outcome occurred at the participant level.

As predicted, participant ratings of perceived burden, $\chi^2(3, N = 1334) = 26.07, P < 0.001$, and perceived risk, $\chi^2(3, N = 1334) = 27.67, P < 0.001$, decreased significantly with increases in AWG scores, as shown in Table 3. Also as predicted, perceived overall nature of the intervention increased as AWG scores increased, $\chi^2(3, N = 1304) = 36.53, P < 0.001$. This suggests that participant perceptions of burden, risk, tolerability, and acceptability tended to align with expert opinions about nature of the intervention as expressed by the AWG.

Pairwise comparisons of estimated least squares means (Fig. 1) revealed that participant perceptions of intervention burden did not significantly differ for interventions that the AWG assigned a score of 1 (adj M = 3.53, SE = 0.18) as compared with 0 (adj M = 3.79, SE = 0.30), $z = -0.79, P = 0.427$, or a score of 1 as compared with 2 (adj M = 3.16, SE = 0.11), $z = 1.90, P = 0.057$. All other observed differences in perceived burden between AWG nature of the intervention scores were statistically significant. Controlling for other variables in the model, average perceived risk differed significantly between each pair of AWG nature of the intervention scores. This suggests that the distinctions the AWG made when assessing intervention risk are cleanly reflected by incremental differences in participant risk perceptions. Lastly, perceived overall nature of the intervention did not differ significantly for interventions with an AWG nature of the intervention score of 1 (adj M = 1.72, SE = 0.08) as compared with a score of 2 (adj M = 1.82, SE = 0.05), $z = -1.21, P = 0.227$; however, all other pairwise differences were significant. Participant perceptions of overall nature of the intervention increased as expected with the AWG scores but there is no evidence of a distinction between the midrange scores on this outcome.

In addition to the AWG scores, some intervention-level covariates were significant in the final models. Specifically, each 1% increase in the number of participants who reported having a close friend or family member with intervention experience was estimated to result in a small decrease in perceived burden ($B = -0.03, SE = 0.01, z = -3.92, P < 0.001$). Also, the reading level of the intervention synopsis had an impact on perceived risk and acceptability. Each grade-level increase in reading

Table 3 Multilevel mixed model regressions predicting perceived burden, risk, and overall nature of the intervention (NOI)

Predictor	Perceived burden				Perceived risk				Perceived overall NOI			
	Coef.	SE	z	P	Coef.	SE	z	P	Coef.	SE	z	P
Level 1												
Age	-0.01	0.00	-4.99	<0.001	-0.00	0.00	-1.36	0.173	-0.00	0.00	-3.74	<0.001
Sex (ref: male)	0.00	0.06	0.07	0.948	0.03	0.05	0.59	0.558	-0.04	0.04	-1.12	0.264
Race/ethnicity (ref: non-Hispanic white)												
Non-Hispanic black	0.03	0.10	0.34	0.735	0.08	0.09	0.89	0.374	0.06	0.07	0.93	0.351
Other, non-Hispanic	0.07	0.10	0.68	0.497	0.20	0.09	2.22	0.027	-0.06	0.06	-1.02	0.310
Hispanic	0.48	0.12	4.11	<0.001	0.34	0.10	3.20	0.001	-0.08	0.07	-1.07	0.285
Health literacy	-0.12	0.04	-3.09	0.002	-0.08	0.03	-2.37	0.018	0.15	0.02	6.04	<0.001
Attention to the synopsis (ref: low)												
Medium	-0.13	0.07	-1.92	0.055	-0.00	0.06	-0.03	0.980	0.05	0.04	1.18	0.236
High	-0.14	0.08	-1.71	0.087	0.07	0.07	0.98	0.326	0.14	0.05	2.69	0.007
Intervention familiarity (ref: low)												
Medium	-0.19	0.07	-2.65	0.008	-0.09	0.06	-1.42	0.156	0.16	0.04	3.58	<0.001
High	0.05	0.10	0.50	0.614	0.07	0.09	0.81	0.416	0.42	0.06	6.75	<0.001
Intervention experience (ref: none)												
Close friend or family	-0.06	0.09	-0.73	0.467	-0.16	0.08	-1.96	0.050	0.15	0.06	2.59	0.010
Self	-0.30	0.09	-3.46	0.001	-0.19	0.08	-2.35	0.019	0.27	0.06	4.93	<0.001
Level 2												
Percent with intervention experience												
Close friend or family	-0.03	0.01	-3.92	<0.001	0.01	0.01	0.98	0.328	-0.00	0.00	-0.12	0.907
Self	0.00	0.01	0.10	0.921	-0.01	0.00	-1.33	0.184	-0.00	0.00	-1.04	0.297
Profile reading level	0.02	0.04	0.43	0.669	-0.06	0.03	-2.04	0.041	0.04	0.02	2.28	0.023
Profile word count	-0.00	0.00	-0.49	0.622	-0.00	0.00	-1.65	0.098	0.00	0.00	1.80	0.071
AWG NOI (ref: 0)												
1	-0.25	0.32	-0.79	0.427	-0.46	0.23	-1.97	0.049	0.49	0.14	3.42	0.001
2	-0.63	0.30	-2.08	0.038	-0.87	0.22	-3.93	<0.001	0.59	0.13	4.40	<0.001
3	-1.20	0.33	-3.67	<0.001	-1.10	0.24	-4.60	<0.001	0.83	0.15	5.74	<0.001
Intercept	3.79	0.30	12.74	<0.001	3.78	0.22	17.38	<0.001	1.23	0.13	9.33	<0.001
Variance components												
Intercept	0.05	0.02			0.02	0.01			0.01	0.00		
Residual	1.05	0.04			0.85	0.03			0.41	0.02		
	Wald $\chi^2(19) = 153.00, P < 0.001$				Wald $\chi^2(19) = 104.48, P < 0.001$				Wald $\chi^2(19) = 349.27, P < 0.001$			
	PRV ₁ = 0.07				PRV ₁ = 0.03				PRV ₁ = 0.17			
	PRV ₂ = 0.76				PRV ₂ = 0.83				PRV ₂ = 0.88			
	N = 1334				N = 1334				N = 1304			

AWG ClinGen Actionability Working Group, PRV₁ proportional reduction in residual variance from level 1 predictors, PRV₂ proportional reduction in intercept variance from level 2 predictors.

difficulty was associated with a reduction in perceived risk (B = -0.06, SE = 0.03, z = -2.04, P = 0.041) and an increase in perceived overall nature of the intervention (B = 0.04, SE = 0.02, z = 2.28, P = 0.023). Taken together, the AWG scores and these covariates accounted for 77% to

88% of the intervention-level variance in these outcomes (perceived burden PRV₂ = 0.76; perceived risk PRV₂ = 0.83; perceived overall nature of the intervention PRV₂ = 0.88), with the greatest contribution in all three models coming from the AWG scores.

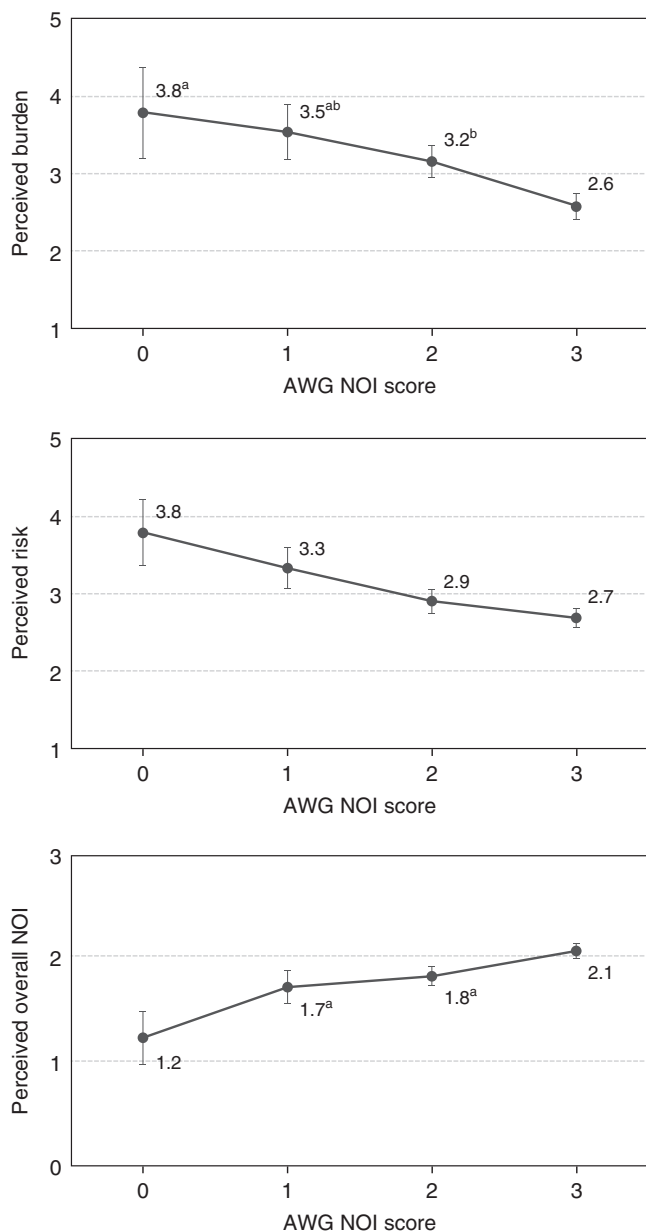


Fig. 1 Estimated perceived intervention burden, risk, and overall nature of the intervention (NOI) by Actionability Working Group (AWG) NOI scores. Values are estimated least squares means. Error bars show the 95% confidence interval for each value. Means sharing a superscript in common by outcome are not significantly different at the $\alpha = 0.05$ level.

DISCUSSION

A major aspect of clinical actionability is whether there is an acceptable intervention to mitigate the impact of the disease outcome(s). In this study, we examined the nature of the intervention dimension of the AWG SQM in relation to lay perspectives. Specifically, we evaluated how well AWG nature of the intervention scores assigned by experts correspond to conceptually related lay perceptions.

In developing the SQM, the AWG aimed to be responsive to patient perceptions and preferences, but the AWG was also keenly aware that nonexpert perspectives of nature of the

intervention may be highly contextual and individuated. Different patients may view the same risks, potential side effects, or procedural burdens differently from one another, and an intervention that is acceptable to some people may be viewed as wholly unacceptable to other people. For example, avoiding the sun during daylight hours is an intervention to reduce cancer risk for people with basal cell nevus syndrome. Conceivably, such an intervention would be more disruptive from the perspective of an individual who spends a lot of time in the sun for work or personal activities. If the influence of these individual differences on perceptions was strong enough, then intervention characteristics would have minimal influence and a summary score capturing nature of the intervention would have little meaning in relation to patient perceptions and experience.

A strength of the multilevel modeling approach used in this study is that it allowed us to directly estimate the amount of variance in perceptions of burden, risk, tolerability, and acceptability available for explanation by characteristics of the intervention as compared with the traits and idiosyncrasies of individual participants. In fact, we found substantial variation in perceptions of burden, risk, tolerability, and acceptability among participants assigned to read about the same intervention. This is represented in the ICCs from the variance components models, which showed that most of the unexplained variance in each outcome occurred at the participant level and that the expected correlation between participants in the same group was fairly small. Nonetheless, the ICCs also show that over our three outcome variables, 10% to 16% of the unexplained variance in these perceptions was due to differences attributable to the interventions. In other words, characteristics of the interventions had at least a modest impact on how participants rated them. Additionally, the AWG nature of the intervention scores accounted for a majority of intervention-level variance. This supports the AWG's assumption about the potential for nature of the intervention scores to represent consolidated lay perceptions of burden, risk, tolerability, and acceptability.

The participants in this study were drawn from the general US population and expressed their perceptions of burden, risk, tolerability, and acceptability in response to a brief narrative summary of the intervention. We recognize that a limitation of this method is that participant ratings should not be taken to represent the lived experience of patients who have undergone the intervention. Instead, the results reflect participants' ability to imagine how they might respond to a clinical recommendation for the intervention. We would anticipate that people living with the condition or who have experienced the intervention may have different responses than the general population. Supporting this assumption, self-reported familiarity and experience with the intervention were significant in most of our models. In a similar vein, perceptions may be more similar among patients who have all had the same intervention, leaving proportionately more intervention-level variation for the SQM score to capture. Logistical challenges aside, an interesting path for future

research would be to solicit perspectives from a community of patients with some of the more common genetic disorders who have actually undergone these interventions.

Importantly, our findings show that consolidated lay perceptions aligned as expected with the expert viewpoint expressed in the SQM. Our approach can also be applied in future research to examine lay perspectives relating to the severity dimension of the SQM. In this study, an ordinal increase in AWG nature of the intervention scores corresponded to decreases in perceived burden and risk, and to an increase in perceived overall nature of the intervention. The association was most clearly represented in the model predicting perceived risk, where the estimated marginal means at each level of AWG nature of the intervention were statistically different from one another. In comparison, the average perceived burden for interventions at the second-lowest point on the AWG nature of the intervention scale was not statistically different from the lowest point or second-highest point. Similarly, perceived overall nature of the intervention was indistinct at the two intermediate points of the AWG nature of the intervention scale. Nonetheless, we found that all three outcomes were monotonically related to the AWG nature of the intervention scores: perceived burden and risk decreased or remained unchanged across adjacent levels, and perceived overall nature of the intervention did the opposite. These findings demonstrate the conceptual fidelity of the nature of the intervention portion of the SQM for genetic conditions in relation to lay perceptions regarding those same interventions.

SUPPLEMENTARY INFORMATION

The online version of this article (<https://doi.org/10.1038/s41436-019-0524-z>) contains supplementary material, which is available to authorized users.

ACKNOWLEDGEMENTS

Funding for this research was provided by National Human Genome Research Institute (NHGRI) grant 4U01HG007437-04 to Jonathan Berg and James Evans at the University of North Carolina at Chapel Hill, and Michael Watson at the American College of Medical Genetics and Genomics, and Marc Williams at Geisinger. The findings and conclusions are those of the authors and do not necessarily reflect the views of NHGRI. The funder had no role in the design of the study, collection and analysis of the data, or the decision to publish the findings.

DISCLOSURE

The authors declare no conflicts of interest.

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

REFERENCES

1. Rehm HL, Berg JS, Brooks LD, et al. ClinGen—the Clinical Genome Resource. *N Engl J Med*. 2015;372:2235–2242.
2. Hunter JE, Irving SA, Biesecker LG, et al. A standardized, evidence-based protocol to assess clinical actionability of genetic disorders associated with genomic variation. *Genet Med*. 2016;18:1258–1268.
3. Clinical Genome Resource. Actionability curation actionability report: SERPINA1 and Alpha-1 antitrypsin deficiency gene-condition pair. 2 May 2018. <https://actionability.clinicalgenome.org/ac-adult/ui/stg2SummaryRpt?version=2356&doc=AC093>. Accessed June 2018.
4. Centers for Disease Control and Prevention. State-specific secondhand smoke exposure and current cigarette smoking among adults—United States, 2008. *MMWR Morb Mortal Wkly Rep*. 2009;58:1232–1235.
5. Logan D, Zelikovsky N, Labay L, Spergel J. The Illness Management Survey: identifying adolescents' perceptions of barriers to adherence. *J Pediatr Psychol*. 2003;28:383–392.
6. Atkinson MJ, Sinha A, Hass SL, et al. Validation of a general measure of treatment satisfaction, the Treatment Satisfaction Questionnaire for Medication (TSQM), using a national panel study of chronic disease. *Health Qual Life Outcomes*. 2004;2:12.
7. Blalock SJ, Patel RA. Drug therapy concerns questionnaire: initial development and refinement. *J Am Pharm Assoc* (2003). 2005;45:160–169.
8. Gaderman AM, Guhn M, Zumbo BD. Estimating ordinal reliability for Likert-type and ordinal item response data: a conceptual, empirical, and practical guide. *Pract Assess Res Eval*. 2012;17:1–13.
9. McLaughlin GH. SMOG grading—a new readability formula. *J Reading*. 1969;12:639–646.
10. Wang LW, Miller MJ, Schmitt MR, Wen FK. Assessing readability formula differences with written health information materials: application, results, and recommendations. *Res Social Adm Pharm*. 2013;9:503–516.
11. University of Nottingham. How to use the SMOG calculator. <http://www.learningandwork.org.uk/SMOG-calculator/smogcalc.php>. Accessed June 2018.
12. Peugh JL. A practical guide to multilevel modeling. *J Sch Psychol*. 2010;48:85–112.
13. Enders CK, Tofighi D. Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. *Psychol Methods*. 2007;12:121–138.
14. Bickel R. *Multilevel analysis for applied research*. New York: Guilford; 2007.
15. StataCorp. Release 13.0 [computer program]. College Station, TX: StataCorp LP; 2013.
16. Cohen J. *Statistical power analysis for the behavioral sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
17. Lai MHC, Kwok O. Examining the rule of thumb of not using multilevel modeling: the “design effect smaller than two” rule. *J Exp Educ*. 2015;83:423–438.