








The development of the Clinician-reported Genetic testing Utility InDEx (C-GUIDE): a novel strategy for measuring the clinical utility of genetic testing

R. Z. Hayeems, ScM, PhD ^{1,2}, S. Luca, MA ¹, W. J. Ungar, MSc, PhD ^{1,2}, A. Bhatt, BMSc(Cand) ¹, L. Chad, MDCM, MHSc ^{3,4}, E. Pullenayegum, PhD^{1,5} and M. S. Meyn, MD, PhD⁶

Purpose: Clinical utility describes a genetic test's value to patients, families, health-care providers, systems, or society. This study aims to define clinical utility from the perspective of clinicians and develop a novel outcome measure that operationalizes this concept.

Methods: Item selection for the Clinician-reported Genetic testing Utility InDEx (C-GUIDE) was informed by a scoping review of the literature. Item reduction was guided by qualitative and quantitative feedback from semistructured interviews and a cross-sectional survey of genetics and nongenetics specialists. Final item selection, index scoring, and structure were guided by feedback from an expert panel of genetics professionals.

Results: A review of 194 publications informed the selection of a preliminary set of 25 items. Feedback from 35 semistructured interviews, 113 surveys, and 11 expert panelists informed the

content and wording of C-GUIDE's final set of 18 items that reflect on the utility of testing related to diagnosis, management, and familial/psychosocial impact. C-GUIDE achieves content and face validity for use in a range of diagnostic genetic testing settings.

Conclusion: Work to establish reliability and construct validity is underway. C-GUIDE will be useful in comparative studies to generate policy-relevant evidence pertaining to the clinical utility of genetic testing across a range of settings.

Genetics in Medicine (2020) 22:95–101; <https://doi.org/10.1038/s41436-019-0620-0>

Keywords: clinical utility; measurement index; outcomes research; mixed methods; clinician perspective

INTRODUCTION

There is much enthusiasm regarding the potential medical benefits of new genetic, genomic, and other “-omic” tests in diagnostic and screening settings. Anticipated benefits include definitive diagnosis and/or presymptomatic risk identification enabling early intervention and a precision medicine approach to management that can reduce morbidity and improve quality of life.^{1–3} Newer generations of genome sequencing may also replace multiple genetic tests and shorten diagnostic odysseys.^{4,5} While evidence is compelling related to the analytic performance of these technologies, much remains unknown about promised health and health care–related outcomes.^{6–10}

The concept of clinical utility is often used to describe a range of benefits associated with genetic testing but specific definitions vary.^{11–14} The American College of Medical Genetics and Genomics (ACMG) defines clinical utility broadly, as a genetic test's “effect on diagnostic or therapeutic management, implications for prognosis, health and

psychological benefits to patients and their relatives, and economic impacts on health-care systems.”¹⁵ While frameworks have helped guide development of this concept since the early 2000s^{16,17} and evidence is accruing to support its individual components,¹⁸ a single validated measure that quantifies clinical utility does not exist. This empiric gap exists alongside policymakers' and payers' requests for evidence that reflects on the value of genetic testing to make funding and policy decisions.^{8–10,13,17,19}

To this end, we have developed a clinician-reported measure of clinical utility for genetic testing. While a wide range of -omic tests are emerging, the Clinician-reported Genetic testing Utility InDEx (C-GUIDE) is designed to assess the post-test utility of diagnostic and secondary variants generated by germline genetic testing. Conceptually, our work draws upon Fryback and Thornbury's hierarchical model of efficacy.^{20,21} While this value framework includes a spectrum of value domains, we focus on the domains of diagnostic thinking and management decision-making. For example,

¹Program in Child Health Evaluative Sciences, The Hospital for Sick Children, Toronto, ON, Canada; ²Institute of Health Policy Management and Evaluation, University of Toronto, Toronto, ON, Canada; ³Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, ON, Canada; ⁴Department of Pediatrics, University of Toronto, Toronto, ON, Canada; ⁵Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada; ⁶Center for Human Genomics and Precision Medicine, University of Wisconsin School of Medicine and Public Health, Madison, WI, USA. Correspondence: R Z. Hayeems (robin.hayeems@sickkids.ca)

Submitted 18 April 2019; accepted: 12 July 2019

Published online: 31 July 2019

genetic testing may alter a clinician's thinking about differential diagnosis, strengthen an existing hypothesis, or reassure a clinician and patient that a speculated diagnosis has or has not been confirmed. Beyond this, a genetic test may have decisional impact associated with an alteration to the patient's care plan. When a diagnostic, predictive, or pharmacogenomic variant has been identified, for example, care plans may be tailored to suit prognoses that are better defined by the test result (e.g., subspecialist referrals, surveillance plans, medication implications, family member testing, reproductive planning). When no variant or a variant of uncertain significance has been identified, care plans may be tailored toward more extensive diagnostic investigations (e.g., muscle biopsies, additional genetic analyses) and monitoring. Since clinicians are well-placed to adjudicate the utility of a genetic test characterized in these ways, we developed an index of items to operationalize these components of value for diagnostic genetic testing.

MATERIALS AND METHODS

C-GUIDE was developed by (1) selecting candidate items through a systematic scoping review of the literature,¹⁸ (2) prioritizing and optimizing items using stakeholder interviews and a survey, and (3) establishing index structure and scoring through a series of deliberations with an expert panel and the development team. The study was approved by The Hospital for Sick Children Research Ethics Board. Survey completion constituted informed consent.

Item prioritization and optimization: stakeholder interviews

Based on the content domains identified in the scoping review and other related literature,¹⁸ a preliminary list of 25 items was generated for inclusion in the tool. While not made explicit in the preliminary list, content domains included diagnosis/prediction, patient management, family impact, psychosocial impact, and system impact. For item prioritization and optimization, eligible stakeholders included geneticists and genetic counselors based in the Division of Clinical and Metabolic Genetics at The Hospital for Sick Children in Toronto, Canada as well as tertiary care affiliated subspecialists who routinely order genetic tests (e.g., cardiologists and rheumatologists who order >10 genetic tests in a three-month period, according to institutional records). Participants were also identified through a University of Toronto-based network of community-based developmental pediatricians and by snowball sampling.²² In total, 125 clinicians were invited to participate. Potential participants were invited by email with up to two reminders over a four-week period.

The individual interviews included qualitative and cognitive components to understand the meaning of clinical utility from clinicians' perspectives and to assess the face and content validity of the preliminary items. The qualitative component asked respondents to define the concept of clinical utility. The cognitive component used think-aloud methodology²³ to ascertain respondents' thoughts on the preliminary

list. Specifically, they were asked to provide feedback on each item regarding their interpretation of its meaning, its wording, and its importance to the concept of clinical utility using a 5-point Likert scale ranging from "extremely important" to "not at all important." All interviews were transcribed verbatim and uploaded onto NVivo 11. Using thematic analysis, domains of clinical utility were identified. Item-specific feedback was reviewed to verify item interpretation and to assign items to respondent-derived domains. Feedback on whether additional items were needed or required revision/removal was incorporated into a revised list of 20 items comprised of three domains. Descriptive statistics were performed to summarize the numeric ratings for each item. Quantitative and qualitative analyses informed the development of the next version.

Item prioritization and optimization: stakeholder survey

To ascertain a broader range of perspectives on the index, a cross-sectional survey was developed and administered to a larger group of geneticists, genetic counselors, and nongenetics specialists who routinely order genetic tests. A study invitation letter and survey link was sent to 50 clinical geneticists (via the Ontario Medical Association), 51 cystic fibrosis clinic directors (via Cystic Fibrosis Canada), 101 rare disease specialists (via Care4Rare²⁴), and 21 developmental pediatricians (via the Developmental Pediatrics-Community Section at the University of Toronto). In addition, an invitation letter and survey link were posted to the Genetic Counselors of Ontario listserv ($n = 191$). An additional group of 156 nongenetics specialists was identified through snowball sampling; the survey was sent to them directly. In total, the survey was distributed to 570 individuals. Over a 6-week period, potential participants were reminded 2-3 times to complete the survey.

The 15-minute online survey asked participants to rank the importance of each item within its respective domain and to rank the importance of each of the three domains. If a domain included six items, respondents were asked to rank each item from 1 to 6, in order of importance (1 = most important; 6 = least important). The survey also asked respondents to rate the clinical sensibility of the index overall (i.e., understandability, comprehensiveness, clarity²⁵) using a 5-point Likert scale and to provide qualitative feedback on the wording of each item, where applicable.

Survey data were analyzed by calculating the average importance ranking (mean and standard deviation) of each item within each domain. Clinical sensibility ratings were summarized with frequency distributions. Qualitative feedback was collated per item. Based on the quantitative and qualitative analyses, the index was further revised (i.e., items were deleted, reworded, combined, separated, or added). The next version of the index included 15 items assigned to one of four domains.

Index structure and scoring: expert panel and team deliberations

An expert panel was assembled to provide final feedback on the items, domains, response options, and scoring strategies.

Table 1 Characteristics of survey participants

Provider type	n (%)
Medical geneticists	24 (21.2)
Genetic counselors	24 (21.2)
Pediatrics	20 (17.7)
Other subspecialties ^a	45 (39.8)
Population^b	
Pediatric/prenatal	61 (54.0)
Adult	99 (87.6)
Practice type^b	
Academic health science center	91 (80.5)
Community-based hospital or private practice (or other)	30 (26.5)
Practice location	
Canada	108 (95.6)
Outside of Canada ^c	5 (4.4)
Years in clinical practice	
0–10	40 (35.4)
>10	73 (64.6)
Years ordering genetic testing	
0–10	51 (45.1)
>10	62 (54.9)
Most frequently ordered genetic test	
Chromosomal microarray	30 (26.5)
Single gene/multigene	69 (61.1)
Exome sequencing/other	14 (12.4)

^aOther specialist groups represented include cardiology, dermatology, endocrinology, hematology/oncology, immunology, nephrology, neurology, obstetrics/gynecology, ophthalmology, otolaryngology, psychiatry, and public health and preventive medicine, respiratory, rheumatology.

^bCheck all that apply.

^cOther countries include United Kingdom, United States.

An 11-member panel was selected, representing expertise in pediatric and adult clinical genetics and genetic counseling, nationally and internationally. All experts had completed the stakeholder survey in the previous step. In a 15-minute online survey, experts were asked to rate the importance of each of the 15 items (using a 3-point Likert scale ranging from “high importance” to “low importance”) and provide final suggestions on item wording. They were also asked to choose their preferred response option (i.e., 3-point or 5-point Likert scale) and to provide feedback on three possible scoring strategies: (1) weighting items by their assigned importance scores, (2) assigning equal weights to each item, or (3) only including items in the index that were rated by experts as highly important. Respondents were asked to choose whether the index score should be calculated as a total score or as domain scores. A mean importance rating was calculated for each item and qualitative feedback was collated and reviewed. Preferences for response options and index scoring strategies were determined with frequency calculations and a review of written comments. Based on the feedback from the expert panel and further deliberations among the development team, a finalized index was constructed for reliability testing and validation.

RESULTS

Participant characteristics

Stakeholder interviews to review and rate candidate items
Of the 125 providers invited, 35 completed interviews (response rate 28%). The sample consisted of a similar number of genetics professionals (42.8%) and nongenetics specialists (57.2%). Hematologists/oncologists (20.0%) and cardiologists (14.5%) made up the largest proportion of nongenetics specialists. The remaining 22.7% of nongenetic specialists practiced in obstetrics/gynecology, otolaryngology, nephrology, ophthalmology, respiratory, or rheumatology. The majority practiced in pediatric settings (80.0%).

Stakeholder survey to rank items within domains

The survey was sent to 156 individuals directly and 414 through a third party. In total, 113 individuals completed the survey (response rate 19.8%). Table 1 describes survey respondent characteristics.

Expert panel to review instrument structure and scoring

Of the 14 experts invited to participate, 11 completed the online survey. The experts were clinical geneticists (63.6%) or genetic counselors (36.4%). The majority were Canadian (72.7%) and had been in practice for >10 years (90.9%).

How do clinicians define the concept of clinical utility?

Stakeholder interviews indicated that clinicians define clinical utility broadly. Thematic analysis identified five specific domains: diagnosis and prognosis (identified by 62.8% of respondents), management recommendations (80.0%), family impact (48.6%), psychosocial impact (22.8%), and system impact (17.1%). The idea that achieving clinical utility requires improving *health outcomes*, specifically, did not emerge from the interviews, except when oncologists discussed the impact of molecularly driven therapeutics.

As explained by one clinical geneticist, “*Clinical utility would be about how useful [testing] is...in terms of medical management..., how important is it to the family. We have many patients that are looking for diagnoses...to have that kind of confirmation, even though it might not alter management, is still a very important piece for families; it allows for testing of other family members, making decisions about prenatal testing and having future children.*” (Pediatric Geneticist 02). Furthermore, a pediatric nongenetics specialist emphasizes the relevance of “system impact”: “*Clinical utility for the health-care system really would boil down to stopping the diagnostic odyssey and beginning a much more streamlined and directed patient care.... So, it might cost them more money in the short-term, but you could argue that maybe for many patients it will save the health-care system a lot of money because we’ll be picking up specific things related to that condition rather than just waiting for them to happen.*” (Pediatric Specialist 08 –Otolaryngologist)

Which items best represent clinical utility?

Stakeholder interviews to review and rate candidate items

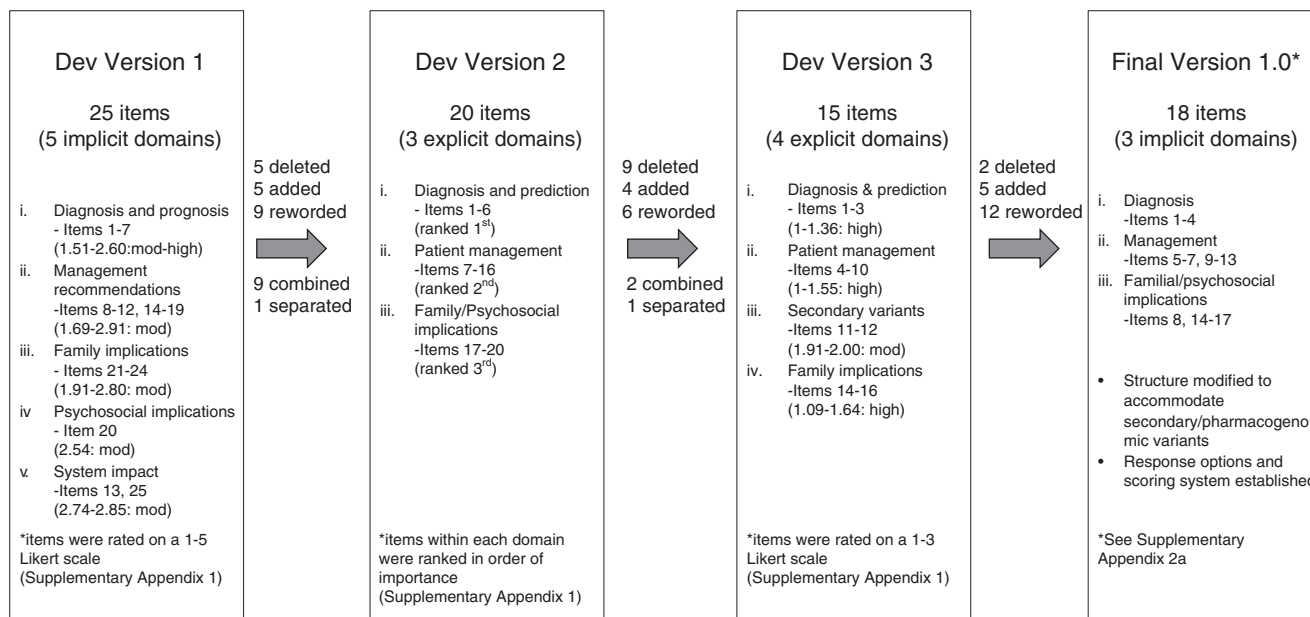


Fig. 1 Flow chart depicting the modifications to Clinician-reported Genetic testing Utility InDex (C-GUIDE) across versions.

Respondents' ratings of the importance of each item are presented as Development Version 1 in the Supplementary Appendix 1. All items were rated as important (or very/extremely important) to the concept of clinical utility. Genetics professionals rated two items as more important than nongenetics professionals but all other items were rated similarly across provider groups. Both importance ratings and qualitative feedback were considered in establishing a reduced item list. For example, item 2 had a mean score of 1.51 (0.95) (denoting high importance) and was qualitatively endorsed as an essential element of clinical utility. It was therefore retained for Development Version 2. Item 12 was rated with a mean score of 2.91 (1.34) (denoting low importance) and not endorsed qualitatively. As such, this item was removed from the index. Where items were identified to be redundant (e.g., items 18 and 19) the item that received the more favorable score and/or was identified to have the clearest wording was retained. In addition to feedback regarding item clarity and wording, item selection was guided by a desire to create a tool that can be applied to a wide range of test types. Figure 1 presents a summary of the modifications made to each version of C-GUIDE throughout its development.

Revised items were then assigned to respondent-derived domains. Where respondents emphasized the importance of specific domains that had not been represented, new items were derived (e.g., emotional response item). Where items could be assigned to more than one domain, these domains were consolidated. For example, "identifies a support group" aligned with the family implications domain as well as the new psychosocial implications domain. As such, these domains were combined. While system impact was identified as a core domain in the thematic analysis, its presence in the index remains more implicit than explicit. For example, the item "avoids further diagnostic testing" is explicitly assigned

to the patient management domain but reflects on system impact implicitly. Final item sorting resulted in three domains: (1) role of genetic testing in diagnosis and prediction, (2) role of genetic testing in patient management, and (3) family/psychosocial implications of genetic testing. Overall, interview feedback led to removing five items, adding five items, rewording nine items, and combining elements of nine items. Development Version 2 included 20 items organized into three domains (Fig. 1).

Stakeholder survey to rank items within domains

Item rankings are presented as Development Version 2 in the Supplementary Appendix 1. Guided by importance rankings and free text comments, the index was further reduced to 15 items. For example, item 2 was among the poorly ranking items for both provider groups in domain 1. Qualitative feedback indicated that the definition of "an extensive search" was subject to interpretation. As such, this item was removed. Different rankings by provider groups on two other items (11 and 16) prompted rewording. Item-specific feedback also suggested that "the identification and management of unexpected health risks" (i.e., items 6 and 15) would be better considered as a separate domain related specifically to secondary variants because these aspects of utility may be viewed differently than those derived from primary diagnostic testing. Overall, nine items were deleted, four items were added, six items were reworded, one item was separated into two, and two items were combined into one. At this juncture, the reduced tool included 15 items organized into four domains. In terms of clinical sensibility, the majority of participants agreed or strongly agreed that C-GUIDE was a comprehensive reflection of the concept of clinical utility (81.4%) and that the items were clear and easy to understand (75.2%). A minority indicated that items were redundant

(21.2%), should be reworded (24.8%), deleted (7.1%), or were missing from the index (8.0%).

Expert panel and developer deliberations to finalize items, structure, and scoring

Experts' ratings of the importance of each item are presented as Development Version 3 in the Supplementary Appendix 1. Overall, all items were rated as important with mean scores ranging from 1.0 to 2.0. Eleven items had mean scores of 1.0–1.5 (i.e., high importance) and four items had mean scores of 1.5–2.0 (i.e., moderate importance). In addition to minor wording changes, expert panelists suggested (1) differentiating recurrence risk implications for patients and family members, (2) capturing psychosocial concern (in addition to psychosocial benefit), (3) differentiating medical actionability and inactionability related to secondary variants, and (4) differentiating utility ratings for primary and secondary variants. As above, final item selection was also guided by a desire to create a tool that can be applied to a wide range of test types.

With respect to item response options, six respondents favored a 5-point Likert scale and five respondents favored a 3-point Likert scale. Respondents further indicated that clarity of response option and ease of completion were essential. As such, item-specific, categorical response options rather than generic Likert scale response options were developed. While the majority of respondents (72.7%) stated a preference for a weighted scoring strategy, high importance ratings assigned to each item precluded the development of item weights. The majority of respondents (70.0%) preferred that the index generate a total score rather than domain scores and indicated that a total score would be simpler to calculate and interpret. They also questioned the complexity associated with domain scores if the number of items per domain varied or if some items were applicable to only some scenarios. As such, in conjunction with a revised approach to item response options, a method of calculating a total score for the index was devised. The face validity of the revised response options and the final scoring strategy were further vetted by the development team.

The final index consists of 18 items for assessing the clinical utility of diagnostic genetic testing (Supplementary Appendix 2a, Fig. 1). Item 18 represents a global item of utility, and will be used for validity testing. Items have two to four response options and each is scored from 0 to 2. Where the utility of a genetic test that only analyzes primary variants is rated, the maximum total C-GUIDE score is 32. Where the utility of a test that also includes the potential to identify secondary variants is rated, the maximum total C-GUIDE score is 48. The index is followed by a set of nine descriptive questions to enable the respondent to provide clinical context for the case to assist with the interpretation of utility ratings (Supplementary Appendix 2b).

DISCUSSION

Genetic tests are rapidly evolving into a diverse set of powerful tools to aid in diagnosis and clinical management.

Because of their variable potential for health and informational impacts, understanding the value of genetic tests to practitioners in the form of clinical utility becomes critically important. Herein we present the development of a clinician-reported outcome measure of clinical utility that aims to assess the informational value of genetic testing on a per-case basis. According to C-GUIDE, utility relates to the ability of a genetic test to contribute to (1) understanding diagnosis and prognosis; (2) management decision-making related to subspecialist care, investigations for diagnostic or surveillance purposes, medication use, or surgery; (3) awareness and actionability of current and future reproductive and health risks for index patients and family members; and (4) psychosocial well-being. Importantly, while these dimensions of value emerged through the analysis of responses during the development process, they are not characterized as independently scored domains in the final tool; rather, item scores are summed to calculate a total clinical utility score. In addition, while items and response options attend to potential long-term utility of test results, the scoring system assigns more weight to elements of utility that manifest in the short term. It is also important to note that with the exception of item 17, all items are positively framed and scored. As well, item response options accommodate the scenario whereby a given benefit is *not* achieved by a particular test, allowing for the absence of utility to manifest. In this way, the index attends to both favorable and unfavorable informational impacts of genetic testing (Supplementary Appendix 2a).

In its current form, the strengths of C-GUIDE lie in its specificity to diagnostic germline genetic testing, which constitutes the majority of current genetic testing applications,²⁶ its attention to a wide range of informational impacts that clinicians feel well positioned to adjudicate, and its capacity to attend to emerging applications of genetic testing (i.e., secondary, pharmacogenomic variant testing, RNA-seq, methylome testing). With respect to the scope of the index, it is noteworthy that endorsed items extend beyond the diagnostic thinking and therapeutic efficacy domains that underpin this work conceptually, and into the patient outcome efficacy domain, by attending to psychosocial and familial benefits and concerns. While some classify psychosocial benefits and concerns as elements of clinical utility,^{27,28} we opted initially to exclude these elements, intending to maintain focus on diagnostic thinking and medical management. We also speculated, initially, that clinicians' assessment of patient/family well-being would be indirect and therefore better assessed by patients/families themselves. However, influenced by respondents' perspectives on the importance of psychosocial impact to the concept of clinical utility as well as their expressed confidence that they are well positioned to judge these impacts, we included these elements in the final version.

We acknowledge several limitations to our measure development approach and to the current version of C-GUIDE. First, our response rates for the stakeholder interviews and survey were low and our reach to nongenetics

specialists was not comprehensive. Low response rates may reflect an ascertainment bias and limited reach to all subspecialist types may reduce the face validity of this version of the index. Our finding that few differences on item importance ratings existed between genetics and nongenetics providers suggests that other specialist types might rate items similarly. However, since we were not powered to detect statistically significant differences between provider groups, further vetting of the index with additional provider groups is warranted. Similarly, since our sampling frame favored pediatric and Canadian providers, further work is required with specialists who work with adult populations and in other jurisdictions. Second, the index is structured to gauge the utility of one genetic test at a time, not to gauge the utility of a series of tests. Hence, in diagnostic testing scenarios where a series of tests may be ordered, the index will address the utility of only one test at a time. As genome-wide testing begins to replace serial testing, this limitation will diminish. Until this point however, we attempt to mitigate this limitation by asking respondents to provide independent ratings for each test result disclosed in a defined time period. This strategy will enable us to account for the utility of each test in a defined series of tests. While serial testing may diminish with increased use of genome-wide strategies, the latter will present the complexity of identifying multiple clinically reportable genomic variants simultaneously. While rating the utility of a combination of variants can be accommodated by the index, attributing utility to one variant over another will be challenging. Since the index was designed to assess the utility of “the test” following the receipt of results, rather than to assess “a variant” per se, the attribution of utility to a particular variant is perhaps less important.

Limitations notwithstanding, our stakeholder-engaged, iterative, mixed methods approach to establishing this first index of its kind aligns with core elements of the ACMG definition of clinical utility, related empiric literature, and Fryback and Thornbury’s hierarchical model of efficacy for assessing the value of diagnostics. Recognizing that the informational impacts of genetic testing will vary by parameters such as test type and indication for testing, we developed this index with input from genetics professionals and genetics-inclined subspecialists as a starting point. Planned further work will establish C-GUIDE’s inter/intrarater reliability and construct validity in a range of clinical settings with a range of provider types. Future work will allow subsequent iterations of this tool to remain responsive to new informational impacts that emerging test capabilities will bring to fruition. While traditional clinical effectiveness studies warrant concerted attention as genotype-driven therapeutics are discovered and adopted by health systems, C-GUIDE can serve to fill evidentiary gaps related to the informational value of genetic testing. Faced with a growing demand for genetic testing by patients and clinicians, funding and policy decision makers in the United States, Canada, and internationally are grappling with how to evaluate and fund costly genetic tests such as genome

sequencing that are now emerging into practice. Payers review evidence on the clinical and analytic validity, safety, and cost-effectiveness of emerging genetic technologies in the form of health technology assessment. The availability of a valid tool to quantify clinical utility can greatly facilitate funding and policy decisions. As such, we will leverage a fertile policy environment to disseminate this tool to academic researchers, professional societies, and decision makers to spur additional validation studies, comparative effectiveness research, and evidence-informed policy.

SUPPLEMENTARY INFORMATION

The online version of this article (<https://doi.org/10.1038/s41436-019-0620-0>) contains supplementary material, which is available to authorized users.

ACKNOWLEDGEMENTS

This work was funded by the Canadian Institutes of Health Research (CIHR).

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

- Johansen Taber KA, Dickinson BD, Wilson M. The promise and challenges of next-generation genome sequencing for clinical care. *JAMA Intern Med.* 2014;174:275–280.
- Manolio TA, Chisholm RL, Ozenberger B, et al. Implementing genomic medicine in the clinic: the future is here. *Genet Med.* 2013;15:258–267.
- Wade CH, Tarini BA, Wilfond BS. Growing up in the genomic era: implications of whole-genome sequencing for children, families, and pediatric practice. *Annu Rev Genomics Hum Genet.* 2013;14:535–555.
- Oei K, Hayeems RZ, Ungar WJ, Cohn RD, Cohen E. Genetic testing among children in a complex care program. *Children.* 2017;4:42.
- Phillips KA, Trostan JR, Kelley RK, Pletcher MJ, Douglas MP, Weldon CB. Genomic sequencing: assessing the health care system, policy, and big-data implications. *Health Aff.* 2014;33:1246–1253.
- Bowdin S, Ray PN, Cohn RD, Meyn MS. The genome clinic: a multidisciplinary approach to assessing the opportunities and challenges of integrating genomic analysis into clinical care. *Hum Mutat.* 2014;35:513–519.
- Bowdin SC, Hayeems RZ, Monfared N, Cohn RD, Meyn MS. The SickKids Genome Clinic: developing and evaluating a pediatric model for individualized genomic medicine. *Clin Genet.* 2016;89:10–19.
- Caulfield T, Evans J, McGuire A, et al. Reflections on the cost of “low-cost” whole genome sequencing: framing the health policy debate. *PLoS Biol.* 2013;11:e1001699–e1001699.
- Ferrante di Ruffano L, Hyde CJ, McCaffery KJ, Bossuyt PM, Deeks JJ. Assessing the value of diagnostic tests: a framework for designing and evaluating trials. *BMJ.* 2012;344:e686.
- Lin JS, Thompson M, Goddard KAB, Piper MA, Heneghan C, Whitlock EP. Evaluating genomic tests from bench to bedside: a practical framework. *BMC Med Inform Decis Mak.* 2012;12:117.
- Bossuyt PM, Reitsma JB, Linnet K, Moons KG. Beyond diagnostic accuracy: the clinical utility of diagnostic tests. *Clin Chem.* 2012;58:1636–1643.
- Grosse SD, Khoury MJ. What is the clinical utility of genetic testing? *Genet Med.* 2006;8:448–450.
- National Academies of Sciences, Engineering, and Medicine. An evidence framework for genetic testing. Washington, DC: The National Academies Press; 2017.

14. Burke W, Laberge AM, Press N. Debating clinical utility. *Public Health Genomics*. 2010;13:215–223.
15. ACMG Board of Directors. Clinical utility of genetic and genomic services: a position statement of the American College of Medical Genetics and Genomics. *Genet Med*. 2015;17:505–507.
16. Pitini E, De Vito C, Marzuillo C, et al. How is genetic testing evaluated? A systematic review of the literature. *Eur J Hum Genet*. 2018;26:605–615.
17. Sun FBW, Erinoff E. Addressing challenges in genetic test evaluation: evaluation frameworks and assessment of analytic validity. Rockville, MD: US Agency for Healthcare Research and Quality; 2011.
18. Walcott S. *Measuring the construct of clinical utility in the context of genetic testing: A scoping review*. Master's thesis. University of Toronto; 2019.
19. Nelson B. Ensuring quality in genomic medicine: amid the rise in complex laboratory-developed tests, regulatory officials are seeking the right balance on quality assurance. *Cancer Cytopathol*. 2014;122:855–856.
20. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Mak*. 1991;11:88–94.
21. Hayeems RZ, Luca S, Pullenayegum E, Meyn MS, Ungar W. Genome diagnostics: Novel strategies for measuring value. *J Manag Care Spec Pharm*. (In press).
22. Coleman J. Relational analysis: the study of social organizations with survey methods. *Hum Organ*. 1958;17:28–36.
23. Hak T, Kvd Veer, Jansen H. TheThree-Step Test-Interview (TSTI): An observation-based method for pretesting self-completion questionnaires. *Survey Research Methods*. 2008;2:8.
24. Care4Rare. 2019. <http://care4rare.ca/>.
25. Rowe BH, Oxman AD. An assessment of the sensibility of a quality-of-life instrument. *Am J Emerg Med*. 1993;11:374–380.
26. National Center for Biotechnology Information. GTR: Genetic Testing Registry. 2019. <https://www.ncbi.nlm.nih.gov/gtr/>. Accessed 3 April 2019.
27. Kohler JN, Turbitt E, Biesecker BB. Personal utility in genomic testing: a systematic literature review. *Eur J Hum Genet*. 2017;25:662–668.
28. CanMEDS. 2015 Physician Competency Framework. Ottawa, ON: Royal College of Physicians and Surgeons of Canada; 2015.



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, and provide a link to the Creative Commons license. You do not have permission under this license to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2019