JOURNAL
OF THE CANADIAN ACADEMY OF CHILD AND ADOLESCENT PSYCHIATRY / DE L'ACADÉMIE CANADIENNE DE PSYCHIATRIE DE L'ENFANT ET DE L'ADOLESCENT

RESEARCH ARTICLE

Reliability and Concurrent Validity of the GAIN Short Screener Among Youth Utilizing Integrated Health Services

Raymond Khanano, MSc[1]; Skye Barbic, BScOT, PhD[2,3,4,5]; Joanna Henderson, PhD[6,7]; Steve Mathias, MD[3,4,5,8]; Christopher G. Richardson, PhD[1,4,5]

Abstract

Objectives: There is increasing interest in the identification of mental disorders among youth through routine screening in integrated health services. One tool currently being used in Canada is the Global Appraisal of Individual Needs Short Screener (GAIN-SS). The aims of this study were to (1) estimate the internal consistency of the GAIN-SS and its internalizing disorder screener (IDScr) (2) examine concurrent validity of the GAIN-SS and IDScr in an integrated youth health service centre, and (3) identify clinical cut-points for youth aged 17-24 years. **Method:** Participants [n=201, gender=44% women, median age 21 (min,max: 17,24) years] were recruited from an integrated youth health service in Vancouver, British Columbia. Participants completed the GAIN-SS and three reference measures: Kessler Psychological Distress Scale (K10), Patient Health Questionnaire (PHQ-9), and Generalized Anxiety Disorder Scale (GAD-7). Cronbach's alpha, sensitivity, and specificity of the GAIN-SS and IDScr were examined using the K-10, PHQ-9 and GAD-7 as reference measures. Receiver operator characteristic curves were generated to identify optimal cut-points for the GAIN-SS and IDScr. **Results:** A cut-point of seven for the GAIN-SS optimized sensitivity (90%) and specificity (42%) with Cronbach's alpha of 0.91. A similar pattern of results was found for the IDScr and the reference measures it was tested against. **Conclusion:** The results indicate that the GAIN-SS and IDScr have acceptable sensitivity but poor specificity that could be improved via the optimal cut-points identified in this study. This low specificity may be acceptable within an integrated youth health service that provides follow-up diagnostic assessments by a clinician.

Key Words: *adolescent; mental health; depression; anxiety; sensitivity and specificity; primary health care*

[1]UBC School of Population and Public Health (SPPH), University of British Columbia, Vancouver, British Columbia
[2]UBC Department of Occupational Science and Occupational Therapy, University of British Columbia, Vancouver, British Columbia
[3]Foundry, University of British Columbia, Vancouver, British Columbia
[4]Centre for Health Evaluation & Outcome Sciences (CHÉOS), Vancouver, British Columbia
[5]Providence Health Care Research Institute (PHCRI), St. Paul's Hospital, Vancouver, British Columbia
[6]Margaret and Wallace McCain Centre for Child, Youth and Family Mental Health, Centre for Addiction and Mental Health (CAMH), Toronto, Ontario
[7]Department of Psychiatry, University of Toronto, Toronto, Ontario
[8]UBC Department of Psychiatry, University of British Columbia, Vancouver, British Columbia

Corresponding E-Mail: khananr@alumni.ubc.ca

Submitted: June 21, 2019; Accepted: January 9, 2021

Résumé

Objectifs: Il y a un intérêt croissant pour l'identification des troubles mentaux chez les jeunes au moyen d'un dépistage routinier dans les services de santé intégrés. Un outil présentement en usage au Canada est l'Évaluation globale des besoins individuels - Filtrage court (GAIN-SS). La présente étude avait pour buts (1) d'estimer la cohérence interne du GAIN-SS et son dépistage de trouble internalisant (IDScr) (2) d'examiner la validité concourante du GAIN-SS et du IDScr) dans un centre de service de santé intégré pour les jeunes et (3) d'identifier les seuils d'inclusion cliniques pour les jeunes de 17 à 24 ans. **Méthode:** Les participants [n = 201, sexe = 44 % de femmes, âge moyen 21 (min, max, : 17 à 24) ans] ont été recrutés dans un centre de service de santé intégré pour les jeunes de Vancouver, Colombie-Britannique. Les participants ont répondu au GAIN-SS et à trois mesures de référence : l'échelle de détresse psychologique de Kessler (K10), le questionnaire sur la santé du patient (PHQ-9), et l'échelle du trouble d'anxiété généralisée (GAD-7). L'alpha de Cronbach, la sensibilité, et la spécificité du GAIN-SS et d'IDScr ont été examinés à l'aide de K-10, PHQ-9 et GAD-7 comme mesures de référence. Des courbes caractéristiques de fonctionnement du récepteur ont été générées pour identifier les seuils d'inclusion optimaux pour les GAIN-SS et IDScr. **Résultats:** Un seuil de sept pour le GAIN-SS optimisait la sensibilité (90 %) et la spécificité (42 %) avec un alpha de Cronbach de 0,91. Un modèle de résultats semblable a été constaté pour l'IDScr et les mesures de référence contre lesquelles il a été testé. **Conclusion:** Les résultats indiquent que le GAIN-SS et l'IDScr ont une sensibilité acceptable mais une spécificité médiocre qui pourrait être améliorée par les seuils d'inclusion optimaux identifiés dans cette étude. Cette faible spécificité peut être acceptable dans un service de santé intégré pour les jeunes qui offre des évaluations diagnostiques de suivi par un clinicien.

Mots clés: *adolescent; santé mentale, dépression; anxiété; sensibilité et spécificité; soins de santé de première ligne*

Introduction

Mental Health of Youth

It is estimated that 1 in 5 Canadians are living with a mental illness and approximately one million are youth between the ages of 9 and 19 years (Mental Health Commission of Canada, 2013). Additionally, the highest incidence of new cases of mental illnesses tends to be observed between the ages of 12 and 24 (Gore et al., 2011). In the literature, there have been distinctions made between the terms "adolescence" and "young adulthood" as developmental periods. Among these, "youth" has been used to refer to any individual between the ages of 12 and 24 (Hetrick et al., 2017). In addition to the nomenclature confusion in this population, a further barrier to research, practice, and policy is understanding the needs of young people, from the perspective of young people themselves. For example, in Canada, the National Longitudinal Survey of Children and Youth (NLSCY, 2010) collected data on emotional and behavioural health outcomes of children and youth (up to age 25). These data have been used in various studies to characterize rates of internalizing and externalizing problems among youth and inform policy decisions. Yet this study, and others, have been criticized for not measuring the needs of young people with tools that were designed or tested on the target population (Barbic et al., 2018). In order to measure outcomes and experiences of young people, we must be confident that these measures are fit for purpose for the context of use and the target population. Specifically, it is important that screening tools provide reliable and valid estimates of their outcome measures.

Mental Health Screening in Youth

Proponents of screening for mental illness have cited its importance in early detection of mental illness, decreasing youth suicidality, and promoting overall mental health (Dowdy, Ritchey, & Kamphaus, 2010). Where and when to screen youth is a topic of recent debate (Gill et al., 2017). Increasingly, integrated youth health services are emerging across Canada, including Foundry (British Columbia), Access Open Minds (pan-Canadian), and Youth Wellness Hubs Ontario (Ontario). Services include primary care, substance use and mental health support, peer support, and social services (Hetrick et al., 2017). Due to their accessibility and rapid access to subsequent care if necessary, integrated youth health services may represent an ideal setting for screening for diverse health needs including mental illness. They also present an excellent opportunity to work closely with young people across Canada to test existing measures and develop solutions for meaningful measurement in this population. One area of emerging interest is clinical staging.

A recent paradigm shift in psychiatry has focused on better characterizing the progression of mental disorders from youth into adulthood through the creation of a clinical staging model (McGorry et al., 2007). Youth often present symptoms in "transdiagnostic clusters" which heavily overlap across the diagnostic criteria of various mood, emotional, and behavioural disorders (Weersing, Rozenman, Maher-Bridge, & Campo, 2012). In order for this model of

clinical staging to be useful, screening tools which measure the risk of mental illness must be shown to be valid and reliable in youth populations with sufficient evidence to support their use. There is an additional need to determine if cut-points developed for adults populations are appropriate for screening youth and identify youth-specific clinical cut-points that might be more suitable for screening in this population.

The GAIN-SS

The Global Appraisal of Individual Needs is a family of measures that consists of tools that are clinical assessments as well as brief screening tools. The Global Appraisal of Individual Needs – Initial (GAIN-I) is a semi-structured assessment tool that has been used to provide a formal psychiatric diagnosis based on the Diagnostic and Statistical Manual for Mental Disorders (DSM-IV) (Dennis, White, Titus, & Unsicker, 2008). One of the most widely used variants of the GAIN-I, is the GAIN Short Screener (GAIN-SS) (Dennis, Feeney, & Titus, 2013). In the context of an integrated youth health service, there are various advantages associated with the use of a tool such as the GAIN-SS. It provides a rapid screen of two major domains of psychopathology (internalizing and externalizing problems), as well as substance disorders and crime and violence problems.

The GAIN-SS serves three purposes, to: (1) screen in diverse clinical populations, (2) estimate the need for referral across different health systems, and (3) serve as a measure of change in behavioural health over time. The GAIN-SS contains 23 items which screen for significant internalizing (6 items), externalizing (7 items), substance use (5 items), and crime and violence problems (5 items). These items are reported to be used as sub-screeners that can function as brief screening tools when used independently of the rest of the tool, each with their own unique identifiers: Internalizing Disorder Screener (IDScr), Externalizing Disorder Screener (EDScr), Substance Disorder Screener (SDScr), and Crime/Violence Screener (CVScr). When scoring the tool, the total scores from each individual screener are summed to form the Total Disorder Screener (TDScr) score (Dennis, Feeney, & Titus, 2013). The tool is reported to be easy to self-administer or staff-administer on paper or by computer as part of an intake assessment or primary care visit (Dennis, Feeney, & Titus, 2013). When a staff member is administering the GAIN-SS, it can be performed as a semi-structured assessment where the interviewer is able to clarify and record patient responses exactly as they are said. As noted above, the GAIN-SS is derived from the GAIN-I which is a diagnostic tool, but it cannot provide a formal diagnosis. Rather, if an individual screens positive on the GAIN-SS, he/she/they would proceed to a clinical assessment which may include administration of the full version of the GAIN-I.

Using the results from GAIN-I clinical assessments as a gold standard, the initial version of the GAIN-SS was determined to have excellent sensitivity (90%) and specificity (92%) for its total disorder screener (Dennis et al., 2006). Along with their initial validation of the GAIN-SS, Dennis et al. (2006) also reported that the total disorder scale of the GAIN-SS had good internal consistency (alpha = 0.96), and that it was strongly correlated with the GAIN-I (Pearson r = 0.94). This study was conducted with a sample of nearly 6,000 adolescents and 2,000 adults drawn from a diverse range of sites, with some participants being in residential treatment or involved in the juvenile or criminal justice system. The work of Dennis and colleagues was important in establishing the validity of the GAIN-SS, and provided the foundation for the current study to provide additional evidence on the tool. Smith et al. (2017) were among the first to examine the sensitivity and specificity of a GAIN-SS sub-screener, the Substance Disorder Screener (SDScr). Based on a large sample (n=9,808), the scores of the SDScr were examined against the GAIN-SS total disorder score. The SDScr was found to have a sensitivity of 83% and specificity of 95% when a cut-point of 2 was used. This study suggested that the SDScr may be acceptable for use in emerging adults (youth between the ages of 18 and 25 years). It was also among the first to validate a GAIN-SS sub-screener to determine if it can be used to triage individuals to appropriate substance use treatment programs. To date, we are unaware any existing literature studying the psychometric properties of the Internalizing Disorder Screener (to be called the IDScr for the remainder of this paper).

Objectives

This study aims to provide additional evidence on the concurrent validity of the GAIN-SS and IDScr to inform their use as tools for screening youth (aged 17 to 24 years) for mental health disorders in an integrated health service. Concurrent validity is a form of criterion validity, the effectiveness of test to estimate a participants performance on another outcome measure. Specifically, concurrent validity estimates performance on difference tests of similar nature administered at approximately the same time (Lin & Yao, 2014). Using other validated screening tools as reference measures, we will determine Cronbach's alpha to examine internal consistency, as well as sensitivity and specificity to comment on the tool's concurrent validity. We will generate Receiver Operator Characteristic (ROC) curves to determine optimal cut-points for the tool when used in

youth populations. Finally, using a similar method, we will examine the concurrent validity and determine optimal cut-points for the IDScr, a sub-section of the GAIN-SS known as the internalizing disorder screener.

Methods

Study Setting and Recruitment

Located in British Columbia, Foundry Vancouver-Granville, is one of eleven operating locations that currently make up the Foundry network and was the primary setting for this study. Foundry Vancouver-Granville provides a wide range of health and social services to youth centralized in one building, including primary care, mental health support, substance use support, peer support, and social services.

This study used a convenience sample of youth visiting Foundry Vancouver-Granville to recruit its participants. Youth were recruited via flyers placed in high traffic areas of the clinic; additionally, on select days a member of the research team set up a table in the clinic to answer questions and enrol participants. All primary and reference measures along with the GAIN-SS were self-reported, completed, and automatically scored on a computer tablet provided by a member of the research team on site at Foundry Vancouver-Granville. Permission was obtained from all authors to use these measures in our study.

Data Source

The data used for this study came from a sample of 204 youth participants who visited Foundry Vancouver-Granville between July and August 2018, who were between the ages of 17 and 24. Eligible youth, aged 19 and over, followed consenting protocols outlined by the University of British Columbia Ethics Board. For youth aged 18 or less, assent protocols, followed by a meeting with a trained research team member, were used. There was no apriori sample size or power calculation completed for the current study.

Demographic information collected in the questionnaire included age, gender, ethnicity, self-reported mental health diagnoses. In addition to the GAIN-SS, participants also completed the Kessler Psychological Distress Scale (K10), Patient Health Questionnaire (PHQ-9), and Generalized Anxiety Disorder 7-item scale (GAD-7).

Primary Measures

The GAIN-SS and sub-section IDScr (Version 3) were the primary tools of interest for this study. Scoring of the GAIN-SS is based on the timeframe when someone experienced symptoms in the past year, with symptoms experienced more recently scoring higher. For example, an item may ask someone when was the last time they experienced significant problems with "feeling very trapped, lonely, sad, blue, depressed, or hopeless about the future". The participant is able to choose from the following responses: "past month", "2 to 3 months ago", "4 to 12 months ago", "1+ years ago" or "never". Someone who has experienced significant problems with symptoms in the past month (maximum score of 3) will score higher on this particular item than someone who has not experienced significant problems with symptoms at all in the past year (minimum score of 0). When scoring the GAIN-SS, clinic staff will flag clients in need of further assessment when clients score a minimum of 1 on the total disorder screener. A score of 1 or 2 in the total disorder screener (GAIN-SS) is considered "moderate" likelihood of receiving a diagnosis after in both adolescent and adult populations. For adolescents in outpatient settings, a median score of 6 has been observed, with 78% of participants scoring high between 3 and 23. The IDScr is the internalizing disorder screener, a sub-section within the total disorder screener and is made up of six items that originate from the GAIN-SS (Dennis et al., 2013).

Reference Measures

The K10, developed by Kessler in 2003, is a 10-item screening tool used to measure global distress or non-specific mood and anxiety disorders (Furukawa, Kessler, Slade, & Andrews, 2003). The development of the tool was informed by the DSM-IV which was the latest version of the DSM at the time of development (Furukawa et al., 2003). As the K10 is a measure of global distress related to non-specific mood and anxiety disorders, its use can be likened to that of the GAIN-SS as a trans-diagnostic screening tool. Uniquely, the K10 is a much briefer scale and is primarily a measure of global distress that does not directly assess issues related to externalizing disorders, substance use, and crime and violence. Although the K10 focuses on assessing symptom severity, it can be used as a screening tool to identify people with suspected clinical conditions. For example, the K10 has been validated multiple times among adults, across various cut-points, with values ranging between 77% and 86% for sensitivity, and 74% and 83% for specificity (Anderson et al., 2013; Sampasa-Kanyinga, Zamorski, & Colman, 2018). There is less evidence of the validity and reliability of the tool in youth populations. When using the optimal cut-point (27) for a sample of Hong Kong youth from their analysis, the sensitivity was 85% and the specificity was 83% (Chan & Fung, 2014). The study also reported the internal consistency of the K10 to be 0.86 using Cronbach's Alpha (Chan & Fung, 2014). In adult populations,

subjects with a score of 20 or greater are described as having a mild mental disorder, 25 or greater are described as having a moderate mental disorder, and 30 or greater are described as having a severe mental disorder. Given the limited evidence on youth-specific cut-points a cut-point of 20 or greater was used in this study.

The Patient Health Questionnaire (PHQ) is a family of measurements containing the PHQ-9, a screening tool used for the screening and measurement of depressive symptom severity (Kroenke, Spitzer, & Williams, 2001). The PHQ-9 consists of 9 items, each scored on a 4-point rating scale, with a higher score indicating greater severity of depression symptomology. At the time of its development, the PHQ-9 was considered unique because it was half the length of traditional tools used for the measurement of depression and was based on the diagnostic criteria for depression described in the DSM-IV (Kroenke et al., 2001). The PHQ-9 has demonstrated acceptable validity and reliability through testing in multiple adult clinical populations, with sensitivity and specificity both being reported as 88% and internal consistency reliability measured with Cronbach's alpha ranging from 0.86 to 0.89 (Kroenke et al., 2001). A "threshold" cut-point of 10 identified in a systematic review and meta-analysis of the PHQ-9 has been used in screening for depression (Kroenke, Spitzer, Williams, & Löwe, 2010). One report found that using the PHQ-9 to screen adolescents for depressive symptoms required a higher cut-point (11) than adults. Despite this, the findings were similar to adults with sensitivity of 89.5% and specificity of 77.5% (Richardson et al., 2010). For screening purposes, a minimal cut-point of 5 represents the presence of mild depressive symptoms severity. Individuals scoring between 10 and 14 are flagged as having moderate depressive symptoms severity, with scores between 15 and 19 as moderately severe, and scores between 20 and 27 as severe. The current study used the minimum cut-point of 5 to be consistent with the choice of using the minimal cut-point score of all other reference measures.

The GAD-7 (Generalized Anxiety Disorder) is a screening tool that was developed around the diagnostic criteria for anxiety as per the DSM-IV, similar to the PHQ-9 (Spitzer, Kroenke, Williams, & Löwe, 2006). The original tool consisted of 13 items, nine of which were the DSM-IV criteria for diagnosing GAD in patients and the remaining four items were chosen based on their use in pre-existing screening tools (Spitzer et al., 2006). The seven items that were chosen for the final version of the tool were those that had the greatest correlation coefficient with the full version containing 13 items. A cut-point of 10 was found to maximize sensitivity and specificity of the GAD-7 to 89% and 82% respectively (Spitzer et al., 2006). Less evidence exists on the reliability of the GAD-7 in youth; however, Löwe et al. reported good internal consistency reliability (Cronbach's Alpha = 0.89) in the general population (2008). For screening purposes, a minimal cut-point of 5 represents the presence of mild anxiety severity. When scoring the GAD-7, the cut-point of 10 represents moderate anxiety severity and 15 represents severe anxiety severity.

Analyses

Descriptive statistics were used to characterize the total sample and groups defined by gender. Internal consistency reliability of the primary measures was assessed using Cronbach's alpha. Concurrent validity of the GAIN-SS with the K10 as a reference measure was estimated by calculating the sensitivity and specificity with the cut-points of each tool applied to create binary measures. The rationale for this choice was that although the K10 is a measure of global distress, research indicates that it can also be used as a screening tool to identify people with suspected mental health disorders. To determine the cut-points at which sensitivity and specificity of the GAIN-SS are optimized, Receiver Operator Characteristic (ROC) curves were created to help visualize the relationship between the true positive rate (sensitivity) and the false positive rate (specificity – 1). To identify the optimal cut point, Youden's J Statistic [sensitivity + (specificity – 1)] was calculated and the highest value between 0 and 1 was noted. The purpose of the ROC curve is to visually depict the relationship between the true positive rate and the false positive rate of the tool across various cut-points (Fawcett, 2006). Area Under the Curve (AUC) values, which provide an estimate of the predictive power of the tool, were calculated by inputting data points from the ROC curve into the RStudio package "MESS" (Ekstrøm, 2019). Prior to determining the AUC values, it was noted that a minimum AUC value of 0.7 would be acceptable, with values greater than 0.8 being considered excellent (Hosmer, 2013). Analysis of concurrent validity for the IDScr followed a similar process to the GAIN-SS, where the K10, PHQ-9, and GAD-7 were used as reference measures to determine sensitivity, specificity, and generate ROC curves as well as optimal cut-points.

Missing Data Analysis

Of the original 204 participants, three (<1%) did not provide valid responses to the majority of the survey questions and were removed from the dataset. Among the remaining 201 participants, the amount of missing data was below 3% for all items and the missing values appeared to be missing completely at random (Little's MCAR test: Chi-Square = 223, DF = 604, $p > .05$). To maintain sample size, missing

Table 1. Descriptive statistics and demographic characteristics of the study sample by gender (N=201).

	Women (n=88)	Men (n=89)	Non-Binary (n=24)	Total (n=201)
Variables	Mean (SD)			
Age (years)	21.4 (1.8)	21.2 (2.0)	18.4 (2.1)	21.3 (2.0)
	% (n)			
Ethnicity				
White	53.4 (47)	47.2 (42)	54.2 (13)	51.7 (104)
First Nations, Métis, Inuit	12.5 (11)	19.1 (17)	4.2 (1)	14.4 (29)
Multiple Ethnicities	12.5 (11)	13.5 (12)	25.0 (6)	14.4 (29)
Other	21.6 (19)	18.0 (16)	8.3 (2)	18.4 (37)
Highest Level of Education Attained				
Secondary	60.2 (53)	77.5 (69)	70.8 (17)	69.2 (139)
Post-Secondary	39.8 (35)	22.5 (20)	20.8 (5)	29.9 (60)
Self-Reported Diagnoses**				
Anxiety	78.4 (69)	52.8 (47)	83.3(20)	67.7 (136)
Bipolar Disorder	26.1 (23)	13.5 (12)	16.7 (4)	19.4 (39)
Depression	70.5 (62)	50.6 (45)	79.2 (19)	62.7 (126)
Post-Traumatic Stress Disorder	38.6 (34)	24.7 (22)	87.5 (21)	35.3 (71)
Schizophrenia or Schizoaffective Disorder	2.3 (2)	18.0 (16)	8.3 (2)	10.0 (20)
Other	36.4 (32)	39.3 (35)	45.8 (11)	38.8 (78)

**Participants were allowed to select more than 1 diagnosis, thus it should be noted that total column reflects the total number of participants who selected that diagnosis relative to the total sample size.

responses for the 201 participants were imputed using single imputation with the EM algorithm in SPSS Version 25 ("IBM SPSS Statistics for Windows," 2017). Given the low amount of missing data and the results of Little's MCAR test we chose to impute missing data with single imputation.

Results

Basic demographics for the total sample and gender-specific groups are reported in Table 1. The average age of participants was 21.3 years (SD=2.0), the median age was 21 and ranged from 17 to 24 years. Of study participants, 43.8% of participants identified as women, 44.3% identified as men, and 11.9% identified as non-binary or a gender other than woman or man. The majority of this sample identified as "White" (51.7%). Participants identifying their ethnicity as "First Nations, Métis, and Inuit" represented 14.4% of this sample. Participants who identified more than one ethnicity in the survey represented 14.4% of the sample. When asked to self-report any mental health diagnoses, the majority of participants reported having Anxiety (78.4% of women, 52.8% of men, 83.3% of non-binary participants) and/or Depression (70.5% of women, 50.6% of men, 79.2% of non-binary participants).

Analysis of Cronbach's alpha indicated the GAIN-SS and IDScr have generally acceptable (greater than 0.80) internal consistency (Nunnally & Bernstein, 1994) with the GAIN-SS having an alpha of 0.91 and the IDScr having an alpha of 0.83. The Receiver Operator Characteristic (ROC) curve plotting the GAIN-SS against the K10 is summarized in Figure 1. When using the recommended cut-point of 1 for the GAIN-SS, the sensitivity and specificity are 100.0% and 9.6% respectively. Using Youden's J Statistic, an optimal cut-point of 7 was identified. At this point the observed sensitivity was 96.0% and the specificity was 40.4%.

Sensitivity and specificity values observed when the IDScr was tested against the K10, PHQ-9, and GAD-7 are summarized in Table 2, and corresponding ROC curves are provided in Figure 1. An optimal cut-point of 4 was determined for the IDScr consistently throughout analysis via Youden's J Statistic. When tested against the K10, the sensitivity and specificity of the IDScr were 99.3% and 19.2% for the recommended cut-point, and then 91.3% and 48.1% at the

Table 2. Values for sensitivity, specificity, and Area Under the Curve (AUC) for the GAIN-SS and IDScr across optimal cut-points determined through Receiver Operator Characteristic (ROC) curves.

	Cut-point	Sensitivity (%)	Specificity (%)	Youden's J	AUC
GAIN-SS	1	100	9.6	0.1	**0.67**
Reference: K10	**7**	**96**	**40.4**	**0.4**	
	8	90.6	42.3	0.3	
	9	87.9	44.2	0.3	
IDScr	1	99.3	19.2	0.2	**0.71**
Reference: K10	2	99.3	21.2	0.2	
	3	97.3	30.8	0.3	
	4	**91.3**	**48.1**	0.4	
IDScr	1	97.8	33.3	0.3	**0.78**
Reference: PHQ-9	2	97.8	38.1	0.4	
	3	94.4	47.6	0.4	
	4	**86.7**	**66.7**	**0.5**	
IDScr	1	98.3	30.8	0.3	**0.72**
Reference: GAD-7	2	98.3	34.6	0.3	
	3	94.3	38.5	0.3	
	4	**86.3**	**53.8**	**0.4**	

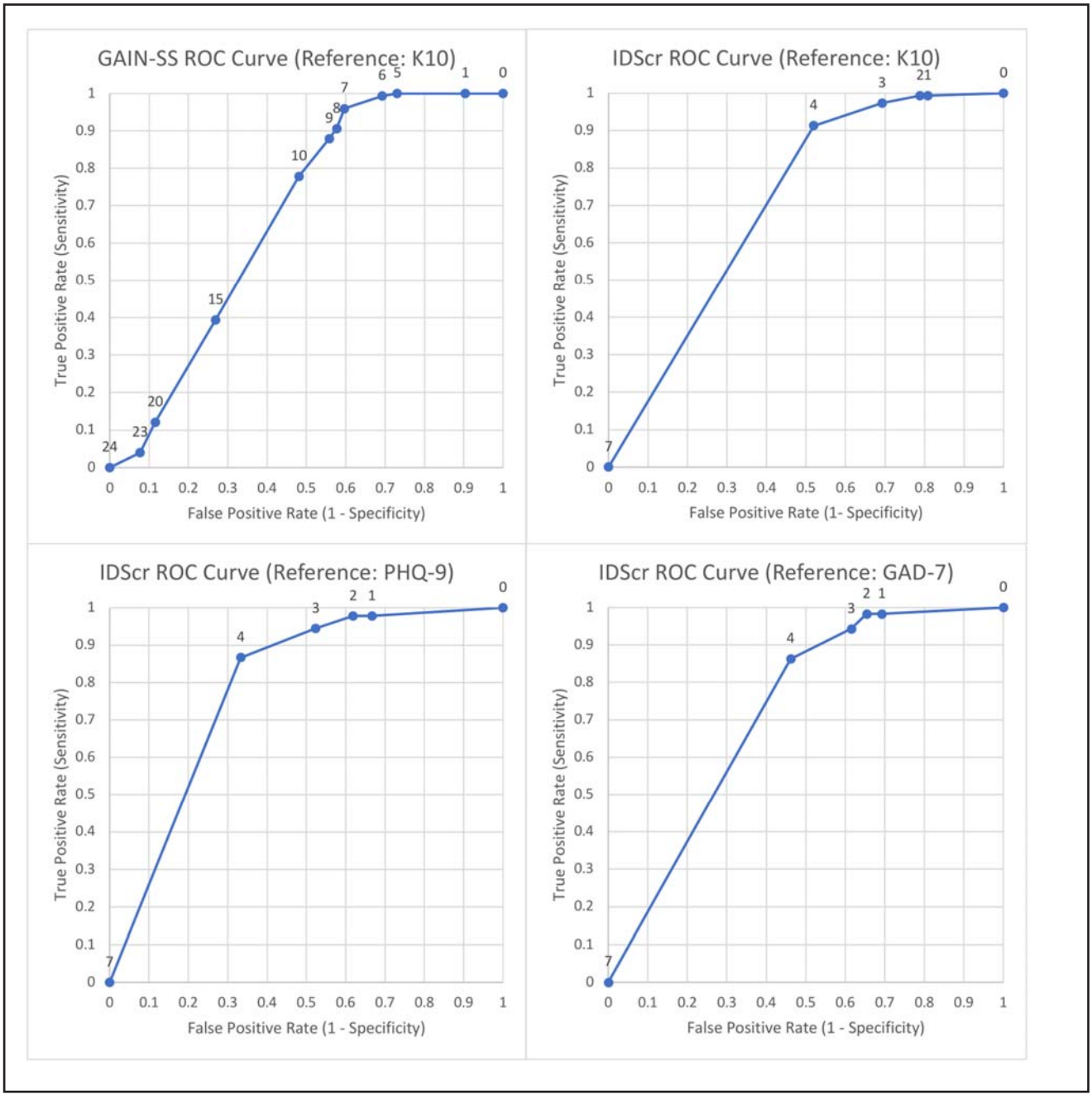
optimal cut-point. When tested against the PHQ-9, the sensitivity and specificity of the IDScr were 97.8% and 33.3% at the initial recommended cut-point, and then 86.7% and 66.7% at the optimal cut-point. When tested against the GAD-7, the initial sensitivity and specificity of the IDScr for the recommended cut-point were 98.3% and 30.8%, and then 86.3% and 53.8% at the optimal cut-point.

Discussion

The purpose of this study was to evaluate the concurrent validity and internal consistency reliability of the GAIN-SS and the IDScr for use in youth (aged 17 to 24 years) utilizing integrated health services using other validated screening tools (the K10, PHQ-9, and GAD-7). Sensitivity and specificity of the GAIN-SS were analyzed using the K10 as a reference. The results indicated that the GAIN-SS has acceptable sensitivity; however, this appears to be accompanied by a steep trade off in specificity, even when an optimized cut-point is used. To better understand this compromise it is important to first consider how the GAIN-SS is scored. In order to score the minimum of 1 point on a GAIN-SS item, a participant must have indicated they have been experiencing significant problems with an item in the past year (Dennis, 2008). When interpreting scores, the GAIN manual recommends that anyone whose total score was between 1 to 2 may benefit from a full assessment, and estimated that 50% of participants who scored in this range would receive a diagnosis following a clinical consultation (Dennis, 2008). Given this, it seems reasonable to conclude that the GAIN-SS is designed to flag as many participants as possible who present any significant mental health problems in the past year and then provide a secondary assessment.

One of the controversial aspects of screening tests with high sensitivity and low specificity is a phenomena known as "over-screening" (Jenniskens et al., 2017). In oncology, over-screening is seen as an issue for detecting breast cancer, where the second stage may involve an invasive procedure such as a mammogram or biopsy (Seigneurin et al., 2016). Conversely, the second stage of screening for mental illnesses typically involves a clinical assessment and the potential result of a formal diagnosis. While the use of the GAIN-SS may lead to a high number of false positives initially, the consequences associated with a false positive i.e. "over-screening" may be reduced in integrated youth health services where secondary assessment may be readily available. The data from this study suggest that if integrated youth health centres administer the GAIN-SS, they should be resourced and prepared to conduct notable follow ups with a high proportion of youth that complete the screen. In the follow up, particular attention should be given to how young people screen on individual items, rather than

Figure 1. Receiver Operator Characteristic (ROC) curves for the GAIN-SS and IDScr with their respective gold standard and selected cut-points.



the total score. As well, based on the results of this study, we strongly recommend that all secondary assessment and follow-up care are provided in the context of the needs of the young person.

With regards to the IDScr in this context of use, this study provided new evidence that contributes to the ongoing validation of the GAIN-SS. This appears to be one of the first studies to provide evidence on the validity of the IDScr independent of the rest of the tool. The findings for the IDScr were similar to those of the GAIN-SS, in that both tools will have acceptable sensitivity and low specificity when using the recommended cut-point of 1. As with the GAIN-SS, the low sensitivity of the IDScr can be addressed by ensuring that valid second stage assessments are easily accessed by youth who initially screen positive on the IDScr. At this time, clinicians are faced with the challenge of using multiples tools to screen people "in to services" rather than "out". This study has shown that many legacy measures provide information that may be valid and reliable, but not necessarily fit for purpose. In order to drive meaningful engagement of young people in care, it is important that the tools we use to screen youth are meaningful and produce metrics that can inform the allocation of appropriate services where and when youth need them.

The current study has several limitations that may have impacted the results. First, the sample size for this study is smaller relative to other GAIN-SS validation studies. It is difficult to comment on whether this study's sample was large enough as no a priori power calculation was performed. Second, previous studies which assessed the concurrent validity of the GAIN-SS used the complete GAIN-I as their reference measure, which is effectively a gold standard and confirms a diagnosis (Dennis et al., 2006; Smith et al., 2017). The current study used the K10, the PHQ-9, and the GAD-7 as reference measures, all of which are screening tools that do not determine a diagnosis. For the purpose of this study this choice was determined appropriate because these instruments have all been validated in the past in similar settings among adolescents and adults with acceptable sensitivity and specificity. As noted in the description of these measures, we used the current minimum cut-points for adults. Future research could examine the impact of using alternative cut-points for these reference measures on the results presented in this study. When validated with the full version of the GAIN-I, the sensitivity and specificity of the initial version of the GAIN-SS were much higher than the values found in this study (Dennis et al., 2006). The inclusion of substance use and crime and violence items not measured by the K10 may have also impacted the correlation of the tools. In addition, the AUC values associated with most of the ROC curves, with the exception of the IDScr with the PHQ-9, indicate low or poor predictive abilities of the GAIN-SS and the IDScr. Although there are clear limitations in this study, the decision to validate the GAIN-SS and IDScr with the K10, PHQ-9 and GAD-7 is supported by existing evidence on the validity of these tools for use in primary care settings.

In conclusion, the results of this study indicated that the GAIN-SS and IDScr are tools with acceptable sensitivity and will likely flag the majority of individuals who present with transdiagnostic symptoms characteristic of a range of psychiatric disorders. Provided that the GAIN-SS is used in a screening process where a follow-up clinical assessment is readily performed, the low specificity of the tool may not pose major consequences. The ideal context for using the GAIN-SS would thus be within an integrated youth health service, where resources exist for secondary assessment and immediate intervention if needed, a situation that may not exist in all programs. The clinical significance of these results can be framed as balancing between the efficiency of using the GAIN-SS for screening large populations while maintaining some level of confidence in the ability of the screening tool to effectively rule positive cases in and negative cases out. Ideally, a screening tool could effectively rule out negative cases and reduce the costs of secondary assessments which may be more time consuming and expensive. Finally, the GAIN-SS is a brief tool that can easily be self-administered before seeing a clinician, but these results suggest that clinicians and researchers exercise some caution using the GAIN-SS for screening purposes without a follow-up clinical assessment. Additional evidence of the concurrent validity (with other validated assessments) of the GAIN-SS and the IDScr as well as the remaining subscales in youth populations would further inform their use as screening tools. Future studies should also focus on the inter-rater and test-retest reliability of the GAIN-SS as this is a psychometric property of the tool that has not been well studied.

Acknowledgements

The authors of this paper would like to acknowledge that grant funding from the Canadian Institute for Health Research (CIHR) was received in order to complete this study.

Conflicts of Interest

The authors have no financial relationships to disclose.

References

Anderson, T. M., Sunderland, M., Andrews, G., Titov, N., Dear, B. F., & Sachdev, P. S. (2013). The 10-item Kessler psychological distress scale (K10) as a screening instrument in older

individuals. *The American Journal of Geriatric Psychiatry*, *21*(7), 596-606.

Barbic, S., Mathias, S., Tee, K., Manion, I., & Zivanovic, R. (2018). Precisely measuring and

targeting what matters to young adults with mental illness. *Early Intervention in Psychiatry, 12*, 229-229.

Chan, S. M., & Fung, T. C. T. (2014). Reliability and validity of K10 and K6 in screening depressive symptoms in Hong Kong adolescents. *Vulnerable Children and Youth Studies*, *9*(1), 75-85.

Dennis M. L., White M., Titus J., & Unsicker J. (2008). *Global Appraisal of Individual Needs: Administration guide for the GAIN and related measures, version 5.* Normal, IL: Chestnut Health Systems.

Dennis, M. L., Feeney, T., & Titus, J. C. (2013). *Global Appraisal of Individual Needs-Short Screener (GAIN-SS): Administration and scoring manual, version 3.* Normal, IL: Chestnut Health Systems.

Dennis, M. L., Chan, Y. F., & Funk, R. R. (2006). Development and validation of the GAIN Short Screener (GSS) for internalizing, externalizing and substance use disorders and crime/violence problems among adolescents and adults. *American Journal on Addictions*, *15*(s1), 80-91.

Dowdy, E., Ritchey, K., & Kamphaus, R. W. (2010). School-based screening: A population-based approach to inform and monitor children's mental health needs. *School Mental Health*, *2*(4), 166-176.

Ekstrøm, C. T. (2019). MESS: Miscellaneous Esoteric Statistical Scripts (Version 0.5.7) [Software]. Available from https://cran.rproject.org/web/packages/MESS/index.html

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861-874.

Furukawa, T. A., Kessler, R. C., Slade, T., & Andrews, G. (2003). The performance of the K6 and K10 screening scales for psychological distress in the Australian National Survey of Mental Health and Well-Being. *Psychological Medicine*, *33*(2), 357-362.

Gill, P. J., Saunders, N., Gandhi, S., Gonzalez, A., Kurdyak, P., Vigod, S., & Guttmann, A. (2017). Emergency department as a first contact for mental health problems in children and youth. *Journal of the American Academy of Child & Adolescent Psychiatry*, *56*(6), 475-482.

Gore, F. M., Bloem, P. J., Patton, G. C., Ferguson, J., Joseph, V., Coffey, C., ... & Mathers, C. D. (2011). Global burden of disease in young people aged 10–24 years: a systematic analysis. *The Lancet*, *377*(9783), 2093-2102.

Hetrick, S. E., Bailey, A. P., Smith, K. E., Malla, A., Mathias, S., Singh, S. P., ... & Moro, M. R. (2017). Integrated (one-stop shop) youth health care: Best available evidence and future directions. *Medical Journal of Australia*, *207*(s10), S5-S18.

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression*. Hoboken, NJ: John Wiley & Sons Inc.

IBM SPSS Statistics for Windows. (2017). [Computer software]. Armonk, NY: IBM Corp.

Jenniskens, K., De Groot, J. A., Reitsma, J. B., Moons, K. G., Hooft, L., & Naaktgeboren, C. A. (2017). Overdiagnosis across medical disciplines: a scoping review. *BMJ Open*, *7*(12), e018448.

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*(9), 606-613.

Kroenke, K., Spitzer, R. L., Williams, J. B., & Löwe, B. (2010). The patient health questionnaire somatic, anxiety, and depressive symptom scales: a systematic review. *General Hospital Psychiatry*, *32*(4), 345-359.

Lin, W. L., & Yao, G. (2014). Concurrent validity. In A.C. Michalos (Ed.), *Encyclopedia of quality of life and well-being research* (pp. 1184-1185). Dordrecht: Springer Netherlands.

Löwe, B., Decker, O., Müller, S., Brähler, E., Schellberg, D., Herzog, W., & Herzberg, P. Y. (2008). Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. *Medical Care*, *46*(3), 266-274.

McGorry, P. D., Purcell, R., Hickie, I. B., Yung, A. R., Pantelis, C., & Jackson, H. J. (2007). Clinical staging: A heuristic model for psychiatry and youth mental health. *Medical Journal of Australia*, *187*(s7), S40-S42.

Mental Health Commission of Canada. (2013). *Making the case for investing in mental health in Canada.* https://www.mentalhealthcommission.ca/sites/default/files/Investing_in_Mental_Health_FINAL_Version_ENG_0.pdf

National Longitudinal Survey of Children and Youth (NLSCY). (2010). Statistics Canada. Retrieved May 15, 2020, from: https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=4450

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory (3rd ed.)*. New York, NY: McGraw-Hill.

Richardson, L. P., McCauley, E., Grossman, D. C., McCarty, C. A., Richards, J., Russo, J. E., ... & Katon, W. (2010). Evaluation of the Patient Health Questionnaire-9 Item for detecting major depression among adolescents. *Pediatrics*, *126*(6), 1117-1123.

Sampasa-Kanyinga, H., Zamorski, M. A., & Colman, I. (2018). The psychometric properties of the 10-item Kessler Psychological Distress Scale (K10) in Canadian military personnel. *PLOS ONE*, *13*(4), e0196562.

Seigneurin, A., Labarère, J., François, O., Exbrayat, C., Dupouy, M., Filippi, M., & Colonna, M. (2016). Overdiagnosis and overtreatment associated with breast cancer mammography screening: a simulation study with calibration to population-based data. *The Breast*, *28*, 60-66.

Smith, D. C., Bennett, K. M., Dennis, M. L., & Funk, R. R. (2017). Sensitivity and specificity of the gain short-screener for predicting substance use disorders in a large national sample of emerging adults. *Addictive Behaviors*, *68*, 14-17.

Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine*, *166*(10), 1092-1097.

Weersing, V. R., Rozenman, M. S., Maher-Bridge, M., & Campo, J. V. (2012). Anxiety, depression, and somatic distress: Developing a transdiagnostic internalizing toolbox for pediatric practice. *Cognitive and Behavioral Practice*, *19*(1), 68-82.