BMJ Open

Validation of a brief scale to assess ambulatory patients' perceptions of reading visit notes: a scale development study

Julie A Wright ,[1] Suzanne G Leveille ,[2,3,4] Hannah Chimowitz,[2] Alan Fossa,[2,5] Rebecca Stametz,[6] Deserae Clarke,[7] Jan Walker [2,3]

To cite: Wright JA, Leveille SG, Chimowitz H, *et al*. Validation of a brief scale to assess ambulatory patients' perceptions of reading visit notes: a scale development study. *BMJ Open* 2020;**10**:e034517. doi:10.1136/bmjopen-2019-034517

► Prepublication history for this paper is available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2019-034517).

Received 30 September 2019
Revised 06 August 2020
Accepted 10 August 2020

Check for updates

© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to
Dr Julie A Wright;
Julie.Wright@umb.edu

ABSTRACT

Objectives To develop and evaluate the validity of a scale to assess patients' perceived benefits and risks of reading ambulatory visit notes online (open notes).

Design Four studies were used to evaluate the construct validity of a benefits and risks scale. Study 1 refined the items; study 2 evaluated underlying factor structure and identified the items; study 3 evaluated study 2 results in a separate sample; and study 4 examined factorial invariance of the developed scale across educational subsamples.

Setting Ambulatory care in three large health systems in the USA.

Participants Participants in three US health systems who responded to one of two online surveys asking about benefits and risks of reading visit notes: a psychometrics survey of primary care patients, and a large general survey of patients across all ambulatory specialties. Sample sizes: n=439 (study 1); n=439 (study 2); n=500 (study 3); and n=250 (study 4).

Primary and secondary outcome measures Questionnaire items about patients' perceived benefits and risks of reading online visit notes.

Results Study 1 resulted in the selection of a 10-point importance response option format over a 4-point agreement scale. Exploratory factor analysis (EFA) in study 2 resulted in two-factor solution: a four-item benefits factor with good reliability (alpha=0.83) and a three-item risks factor with poor reliability (alpha=0.52). The factor structure was confirmed in study 3, and confirmatory factor analysis of benefit items resulted in an excellent fitting model, $X^2(2)=2.949$; confirmatory factor index=0.998; root mean square error of approximation=0.04 (0.00, 0.142); loadings 0.68–0.86; alpha=0.88. Study 4 supported configural, measurement and structural invariance for the benefits scale across high and low-education patient groups.

Conclusions The findings suggest that the four-item benefits scale has excellent construct validity and preliminary evidence of generalising across different patient populations. Further scale development is needed to understand perceived risks of reading open notes.

Strengths and limitations of this study

- ► The OpenNotes patient benefit and risk scales were developed and evaluated in order to assess patient experiences and perceptions of reading their ambulatory visit notes.
- ► Three studies demonstrated improved construct validity and factorial invariance across patient subgroups in a modified version of the benefits questionnaire compared with the original survey version.
- ► The study was conducted using patient surveys in three large health systems across the USA.
- ► The original intent was to develop a single scale to assess both the benefits and risks of patients reading visit notes; however, the risk items were too few to develop a robust construct.

INTRODUCTION

Secure online patient portals connected to electronic medical records are proliferating, and researchers and practitioners need valid measures to understand how these new tools influence patient care and health outcomes. Portals can provide patients with an easier way to access their clinicians' visit notes, and patient access to these 'open notes' is advocated by the international OpenNotes movement.[1] In the original study with primary care doctors and their patients in three sites in the USA, patients were enthusiastic about reading their visit notes online, and doctors reported only modest effects on their work lives.[2]

Since the original study, more than 200 organisations have adopted open notes, not only in primary care but across their ambulatory practices.[1] By 2014, the three original sites had implemented open notes throughout their ambulatory practices, providing access to clinical notes written by virtually all outpatient clinicians (eg, nurse practitioners, therapists and physician assistants). Given this expansion, the investigators

BMJ Open: first published as 10.1136/bmjopen-2019-034517 on 20 October 2020. Downloaded from http://bmjopen.bmj.com/ on September 17, 2024 by guest. Protected by copyright.

planned a broader evaluation of open notes in the three sites to examine patient experiences in specialty settings, and with all types of health professionals in ambulatory care. The validity of future investigations depends largely on the availability of valid instruments. Moreover, researchers and healthcare systems who seek to evaluate implementation of open notes can benefit from employing valid measures to gain a more accurate understanding of patients' attitudes towards reading their notes.

The original study surveyed doctors and patients using questionnaires that evaluated the acceptability and feasibility of open notes. These surveys were initially developed by a team of investigators who used information from patient focus groups and provider interviews, and subsequent piloting of the instruments.[3] The patient questionnaires included validated scales such as the *Ambulatory Care Experiences Survey*[4] and *Perceived Efficacy in Patient-Physician Interactions*,[5] along with items created by the OpenNotes research team to understand patients' attitudes and experiences, particularly the benefits and risks of reading their notes. While the original study provided a survey with good content validity of the benefits and risks, the construct validity was not assessed and requires further study, and no validated scales were available to assess patients' perceptions of reading visit notes.

Patients' perceptions of the benefits and risks of reading notes can be conceptualised as theoretical constructs. Multiple theories of behaviour include constructs that are similar in meaning to benefits and risks. For example, the constructs of the pros and cons from decision-making theory[6] assess the potential gains or losses to oneself and one's significant others if one were to take action. The health belief model[7] similarly includes the perceived benefits of taking a health action weighed against the perceived barriers of taking that action, typically with the context of susceptibility to or severity of a health condition. Regardless of the theory, these constructs are often used to understand one's decision whether to engage in a behaviour. Developing valid measurement scales to assess these constructs can provide researchers with tools to better understand why patients read visit notes and how their perceptions of the benefits and risks of doing so change across time.

Objectives

The objectives of this study were (1) to compare and report on the psychometric properties of two versions of questions (from the questionnaire used in the original study, and a modified version) to be used in benefit and risk scales; (2) to optimise the scales' length to reduce respondent burden, (3) to assess the scales' construct validity, and (4) to assess the factorial invariance of the scales across subgroups.

METHODS

Overall study design and procedures

This study includes four substudies, with each analysing data from a unique sample of patients from Beth Israel Deaconess Medical Center (BIDMC) in Boston, Geisinger (formerly Geisinger Health System) in Pennsylvania, or University of Washington Medicine (UW) in Seattle. The three institutions participated in the original study[2] in 2010 and have since implemented open notes across ambulatory care. To assess recent experience, we completed two patient surveys: a 'psychometrics survey' testing two different question formats, conducted at BIDMC and Geisinger from October 2016 to February 2017; and a large 'general survey'[8] asking about multiple aspects of patients' experiences with open notes, conducted at all three institutions from July to October 2017. Each participant in this scale development and evaluation study had completed one of the two online surveys. Requirements for written informed patient consent for the surveys were waived.

The four substudies were performed in sequential order to develop and evaluate survey items that were hypothesised to reflect the theoretical constructs of benefits and risks of reading one's open notes. In brief, the purpose of each substudy was: study 1 to evaluate and refine the questions (items); study 2 to determine how many constructs underlie the set of items and identify items that perform better or worse to develop a parsimonious scale; study 3 to confirm the structure based on the analytical results from study 2; and study 4 to test factorial invariance of the developed scale across educational subsamples.

Patient and public involvement

The original patient survey was designed with patient input.[3]

Study 1

While the original study included items that were created to assess advantages (benefits) and disadvantages (risks) of reading visit notes online, the aim of study 1 was to refine and evaluate the questions for use in future evaluations. We developed the psychometrics survey for this purpose and compared two similar versions of the questions. In general, the versions differed by the item stems and response option formats, and item content was held constant. A detailed description of the differences is provided in the Measurement section.

Participants

For the psychometrics survey, we randomly selected 1000 primary care patients at both BIDMC and Geisinger who (1) were at least 18 years old, (2) were registered on the patient portal, and (3) had opened at least one primary care visit note in the previous 12 months according to portal tracking data. We sent them invitations and up to two subsequent reminders with links to the online survey via portal messaging to BIDMC patients in October and November 2016, and via personal email to Geisinger patients in January and February 2017; the field period was 7 weeks at both sites.

Invitations and reminders included a description of the survey and its intended purpose of psychometric

BMJ Open: first published as 10.1136/bmjopen-2019-034517 on 20 October 2020. Downloaded from http://bmjopen.bmj.com/ on September 17, 2024 by guest. Protected by copyright.

evaluation. A statement in the invitation informed patients that the survey 'intentionally included repetitive questions to help us create the best survey possible'. Invitations and reminders also informed patients that their care would not be affected by their participation in the survey, and that 15 respondents would be randomly selected at each site to receive a $100 check or gift card.

Measurement

The psychometrics survey included 73 closed-ended and free-text questions evaluating attitudes and experiences related to open notes, and items about the patient's education and general health status. Patient age and gender were extracted from administrative data and were available only at BIDMC.

The survey included two versions of the benefit and risk questions; version 1 items were those used in the original questionnaire.[2] The stem was, 'As a result of reading your visit notes', followed by 10 statements rated on a 4-point Likert scale (disagree, somewhat disagree, somewhat agree, agree). These questions were developed by a team of investigators using information from patient focus groups, and the instruments were subsequently pilot tested.[3] These early steps suggest good content validity of this scale. However, version 1 resulted in bimodal distributions (unpublished data) that prompted the investigators to modify and evaluate an alternative stem and response format to improve on the bimodal distribution.

The investigators reviewed the items' wording and content and developed an alternative, version 2, based on investigator consensus, participant interviews and prior studies; for example, a question was added about sharing notes because a significant number of patients reported sharing notes with others.[2] Version 2 of the scale tested a modified stem, 'How important is reading your visit notes for', followed by 10 statements answered on a 10-point importance scale, from 1=not at all important to 10=extremely important. The wording of the original and modified benefit and risk items is presented in online supplementary table 1.

Both versions of the benefit and risk questions were included in the survey. In the sequence of the questionnaire, the two versions were separated, with version 2 items near the beginning and version 1 items near the end of the survey.

Data analysis

Descriptive statistics were used to present characteristics of the respondents at the two sites according to education and self-rated health. Education was collapsed into high school/some college or 4-year degree or higher, and self-rated health was collapsed into poor-fair or good-excellent; differences were evaluated using Pearson χ^2 tests. Means, medians, modes, ranges, skew and kurtosis were calculated for the two versions of the benefit and risk questions to assess the normality of the items and provide data for visual comparison. IBM SPSS Statistics (V.25; IBM) was used for analyses in study 1.

Study 2

The purpose of study 2 was to determine how many constructs underlie the set of items that emerged from study 1, and to identify items that performed better or worse. Responses to the version 2 questions from study 1 were used in the analysis.

Data analysis

EFA was performed to explore whether the items represent one or more underlying constructs.[9 10] Maximum likelihood (ML) extraction method and oblimin rotation with Kaiser normalisation were used. The number of factors was determined using eigenvalues >1.0 and scree plots. Decisions to remove items to develop a shorter version of the questionnaire were based on a number of factors: items' statistics (item means, variances and correlations); factor loadings below 0.40; model fit; and Cronbach's alpha coefficient, interitem correlations and conceptual analysis of each item's contribution to the depth and breadth of the construct.[9–11] All analyses for study 2 were performed using IBM SPSS Statistics (V.25; IBM).

Study 3

The purpose of study 3 was to confirm the resulting structure of the scales from study 2. Our approach was to evaluate the construct validity of the revised scale in a new sample of patients. We used split-half, cross-validation methods to explore the structure of the scales from study 2 with EFA in the first half of the sample, and subsequent confirmatory factor analysis (CFA) using structural equation modelling methods in the second half for hypothesis testing.

Participants

We drew a random sample of 500 UW patients who responded to the general survey.[8] The survey methods and eligibility criteria were the same as in the psychometrics survey used in studies 1 and 2, except that eligibility was not limited to primary care patients.

Measurement

Study 3 data included responses to the version 2 questions used in study 2, and age, gender, education and general health status. We split the 500 responses randomly into two equal exploratory and confirmatory samples.

Data analysis

Common factor analysis using ML extraction and oblimin rotation was used to evaluate the factor structure (underlying dimensionality of the scale) in the first half of the sample (n=250). For the second half of the sample, CFAs were performed to test the hypothesis that the model, optimised in the EFA, fit the data. ML estimation and full information maximum likelihood (FIML) were used to handle missing data. Criteria for a good fitting model were assessed using the confirmatory factor index (CFI) >0.95, root mean square error of approximation (RMSEA) <0.05, standardised root mean square residual (RMR)

BMJ Open: first published as 10.1136/bmjopen-2019-034517 on 20 October 2020. Downloaded from http://bmjopen.bmj.com/ on September 17, 2024 by guest. Protected by copyright.

<0.05 and root means square residual near 0. Normality of the data was assessed using kurtosis of <7 and multivariate normality was assessed with a value <5.[12 13] Non-normal data were analysed using ML estimations with bootstrapping using 200 samples to assess model fit. A Bollen-Stine bootstrap p value >0.05 indicates good model fit. CFA was completed in IBM Amos V.24 and SPSS was used for the EFA.[12 13]

Study 4

The purpose of study 4 was to evaluate the factorial invariance of the scale refined in study 3. Factorial invariance tests whether the items provide the same results across different populations. Invariance testing assesses whether the model generalises across groups. Three levels of invariance should be tested to establish invariance: (1) configural, which assesses whether the factor structure is similar across groups, a necessary first step, although the most minimal test; (2) measurement, also known as weak, which assesses the equivalence of factor loadings across groups; and (3) structural, also known as strong, which assesses the equivalence of the item intercepts.[13]

Participants

We chose to examine the factorial invariance in educational subgroups (high school/some college vs 4 years college or higher) because education level is a key social determinant of health.[14] The model in study 3 was developed using UW patients who had high levels of education (65% reported 4 years of college or more). The same general survey provided an opportunity to test the model on Geisinger respondents, 69% of whom reported having less than a college degree. Therefore, we drew a random sample of 250 participants from the Geisinger respondents to the general survey.[8]

Data analysis

Data were assessed for normality as described in study 3. Factorial invariance was assessed using multigroup analysis procedures in Amos V.24. Invariance was assessed by examining whether the fit of the multigroup analysis was consistent across groups and computationally with a X^2 difference test and a CFI difference test. Invariance is confirmed when the difference in the two X^2 is not significantly different. Analysis provides additional evidence for invariance when the difference in the CFIs is not more than 0.01 when comparing the unconstrained model to the measurement model or the structural model.[13]

RESULTS

Table 1 displays the demographic characteristics of the respondents in the four studies. The patients sampled from the three separate healthcare systems were significantly different from one another in terms of education (p<0.001) and perceived health (p<0.009). The distribution of these sample characteristics was similar to that in the original OpenNotes survey.[2]

Study 1

Of the 2000 patients invited to participate, 439 (22%) completed the survey. The majority of BIDMC (83.6%) and Geisinger (77.3%) respondents reported being in good to excellent health. The mean age of BIDMC respondents was 56.5 years (SD 14.2) and 65% were female (table 1).

We first examined the characteristics of responses to the two versions of the benefit and risk questions. The means for the version 1 questions using the 4-point

Table 1 Characteristics of patient survey respondents

	Studies 1 and 2			Study 3	Study 4
	October 2016 to February 2017			August to October 2017	July to September 2017
Patient characteristics	**Total (n=439)**	**BIDMC (n=285)**	**Geisinger (n=154)**	**UW (n=500)**	**Geisinger (n=250)**
Age (years), mean (SD)	n/a	56.5 (14.2)	n/a	53.6 (17.2)	59.45 (14.7)
Female, n (%)	n/a	184 (64.6)	n/a	332 (66.4)	163 (65.2)
Male, n (%)	n/a	101 (35.4)	n/a	168 (33.6)	87 (34.8)
Education, n (%)	435 (99.0)	284 (99.6)	151 (98.1)	481 (96.2)	226 (90.4)
High school/some college	161 (36.7)	58 (20.3)	103 (68.9)	156 (31.2)	148 (59.2)
4-year college or higher	274 (62.4)	226 (79.3)	48 (31.2)	325 (65.0)	78 (31.2)
Missing	4 (0.009)	1 (0.004)	3 (0.02)	19 (3.8)	24 (9.6)
General health, n (%)	437 (99.5)	285 (65.2)	152 (98.7)	478 (95.6)	227 (90.8)
Good-excellent	365 (83.1)	246 (86.3)	119 (77.3)	384 (76.8)	173 (76.2)
Fair-poor	72 (16.4)	39 (13.7)	33 (21.4)	94 (18.8)	54 (21.6)
Missing	2 (0.005)	0	2 (0.01)	22 (0.04)	23 (9.2)

n/a denotes data not available for Geisinger sample (n=154).
BIDMC, Beth Israel Deaconess Medical Center; UW, University of Washington Medicine.

BMJ Open: first published as 10.1136/bmjopen-2019-034517 on 20 October 2020. Downloaded from http://bmjopen.bmj.com/ on September 17, 2024 by guest. Protected by copyright.

Table 2 Study 1: descriptive statistics for version 1 (original) benefit and risk items using a 4-point Likert agreement scale* (n=439)

Item No	Items	n	% DK	Mean	Med	Mode	SD	SK	Kur
Benefit items									
1	I understand my health and medical conditions better.	427	3	3.69	4	4	0.62	−2.31	5.64
2	I take better care of myself.	419	5	3.32	4	4	0.84	−1.22	0.94
3	I remember the plan for my care better.	428	2.5	3.65	4	4	0.68	−2.2	4.79
4	I feel more in control of my healthcare.	430	2	3.64	4	4	0.64	−2.16	5.37
5	I am better prepared for visits.	420	4	3.41	4	4	0.87	−1.49	1.42
6	I do better with taking my medications as prescribed†.	364	7	3.21	4	4	1.05	−1.09	−0.17
Risk items									
1	I worry more.	427	3	1.47	1	1	0.79	1.49	1.06
2	I am concerned about my privacy.	428	2.5	2.06	2	1	1.14	0.49	−1.28
3	The notes are more confusing than helpful.	435	1	1.16	1	1	0.49	3.43	12.13
4	I felt offended.	432	1.6	1.12	1	1	0.43	3.87	14.97

*Item stem: As a result of reading my notes. Response options: 1=strongly disagree, 2=disagree, 3=agree, 4=strongly agree, DK=don't know.
†Asked of 392 patients who reported taking medications.
Kur, kurtosis; Med, median; SK, skewness.

agreement format were at either end of the distribution (either 1 or 4), had small SDs and six of the 10 items were highly skewed or kurtotic, with values above 2 and 3, respectively (table 2).[10] The means for the version 2 questions using the 10-point format were relatively closer to the midpoint of the scale, had larger SDs and had normal skew and kurtosis, less than ±2.0 for all items except one (table 3). Based on this analysis, we narrowed our focus in studies 2–4 to the version 2 items.

Table 3 Study 1: descriptive statistics for version 2 (modified) benefit and risk items using 10-point response scales* (n=439)

Item No	Items	Mean	Med	Mode	SD	SK	Kur
Benefit items							
Imp1	(Importance of) sharing your visit notes with others.	5.45	5	1	3.16	0.03	−1.32
	How important is reading your visit notes for:						
Imp2	Taking care of your health?	8.3	9	10	2	−1.34	1.4
Imp3	Remembering the plan for your care?	7.8	8	10	2.47	−1.09	0.35
Imp4	Helping you feel in control of your care?	8.14	9	10	2.11	−1.31	1.3
Imp5	Preparing for your office visits?	7.02	8	10	2.69	−0.64	−0.65
Imp6	Helping you with your medications†?	6	6	10	3	−0.21	−1.11
Imp7	Feeling like you have an active role in your medical care?	7.82	9	10	2.48	−1.16	0.48
Risk items							
Risk1	How much did you worry about your health after reading your visit notes?	7.46	8	10	2.45	−0.79	−0.105
Risk2	Now that the office visit notes are available online, how concerned are you about privacy?	6.69	7	10	3.04	−0.53	−1.03
Risk3	How confusing were the office visit notes?	8.67	10	10	2	−1.82	3.02

*Importance item 1–7 (Imp1 to Imp7) response options: 1=not at all important to 10=extremely important. Risk items 1–3 (Risk1 to Risk3) used item-specific anchors for the 1–10 scale (1=not at all worried/concerned/confusing to 10=extremely worried/concerned/confusing) and were reverse coded.
†Asked of 392 patients who reported taking medications.
Kur, kurtosis; Med, median; SK, skewness.

BMJ Open: first published as 10.1136/bmjopen-2019-034517 on 20 October 2020. Downloaded from http://bmjopen.bmj.com/ on September 17, 2024 by guest. Protected by copyright.

Table 4 Study 2: first run of EFA (maximum likelihood extraction with oblique rotation) with version 2 items (10-point scale)*

Item No	Items	Factor 1	Factor 2
Benefit items			
Imp1	(Importance of) sharing your visit notes with others.	0.56	−0.26
Imp2	How important is reading your visit notes for: Taking care of your health?	0.76	−0.09
Imp3	Remembering the plan for your care?	0.7	−0.15
Imp4	Helping you feel in control of your care?	0.78	−0.1
Imp5	Preparing for your office visits?	0.74	−0.2
Imp6	Helping you with your medications?	0.7	−0.26
Imp7	Feeling like you have an active role in your medical care?	0.77	−0.16
Risk items			
Risk1	How much did you worry about your health after reading your visit notes?	−0.29	0.54
Risk2	Now that the office visit notes are available online, how concerned are you about privacy?	−0.06	0.37
Risk3	How confusing were the office visit notes?	−0.01	0.76

n=392, includes only patients who answered all items in the analysis; Cronbach's alpha=0.87 for the benefit component and 0.51 for the risk component.
*Importance item 1–7 (Imp1 to Imp7) response options: 1=not at all important to 10=extremely important. Risk items 1–3 (Risk1 to Risk3) used item-specific anchors for the 1–10 scale (1=not at all worried/concerned/confusing to 10=extremely worried/concerned/confusing) and were reverse coded.
EFA, exploratory factor analysis.

Study 2

EFA of the 10 items in version 2 resulted in a two-factor solution: a benefits scale (seven items) and a risks scale (three items). Table 4 presents the correlations between the variables and their factors from the initial run, and the scree plot is shown in online supplementary figure 1. The two factors were correlated, r=−0.24. Table 5 displays the final four-item benefits scale and the three-item risk scale; the two factors were correlated, r=−0.19.

The four-item benefits scale had good internal consistency reliability with a Cronbach's alpha of 0.83, whereas the three items representing risks of reading open notes had an unacceptable Cronbach's alpha of 0.52.[9] A Cronbach's alpha >0.7 is needed to reflect an adequately internally consistent set of items, that is, the items should reliability represent the same construct.[9 15]

Study 3

The UW sample from the general survey (n=500) was 66.4% female, and the average age was 53.6 years (SD 17.2, range 18–92) (table 1). Less than 5% of the data were missing in the exploratory sample of 250 (7 cases or 2.8%); listwise deletion resulted in an analytical sample of n=243. The confirmatory sample of 250 had four missing cases (1.6%); FIML was used to handle missing data.

Table 5 Study 2: the results of the final run of EFA, including item means, SDs and factor loadings* (n=392)

Item No	How important is reading your visit notes for:	Mean	SD	Factor loading
Imp2	Taking care of your health?	8.3	2	0.78
Imp3	Remembering the plan for your care?	7.8	2.5	0.72
Imp4	Helping you feel in control of your care?	8.1	2.1	0.75
Imp5	Preparing for your office visits?	7	2.7	0.74
Risk1	How much did you worry about your health after reading your visit notes?	3.6	2.5	0.56
Risk2	Now that the office visit notes are available online, how concerned are you about privacy?	4.3	3.1	0.37
Risk3	How confusing were the office visit notes?	2.3	2	0.74

n=392, includes only patients who answered all items in the analysis.
*Importance item 1–7 (Imp2 to Imp5) response options: 1=not at all important to 10=extremely important. Risk items 1–3 (Risk1 to Risk3) used item-specific anchors for the 1–10 scale (1=not at all worried/concerned/confusing to 10=extremely worried/concerned/confusing) and were reverse coded.
EFA, exploratory factor analysis.

BMJ Open: first published as 10.1136/bmjopen-2019-034517 on 20 October 2020. Downloaded from http://bmjopen.bmj.com/ on September 17, 2024 by guest. Protected by copyright.

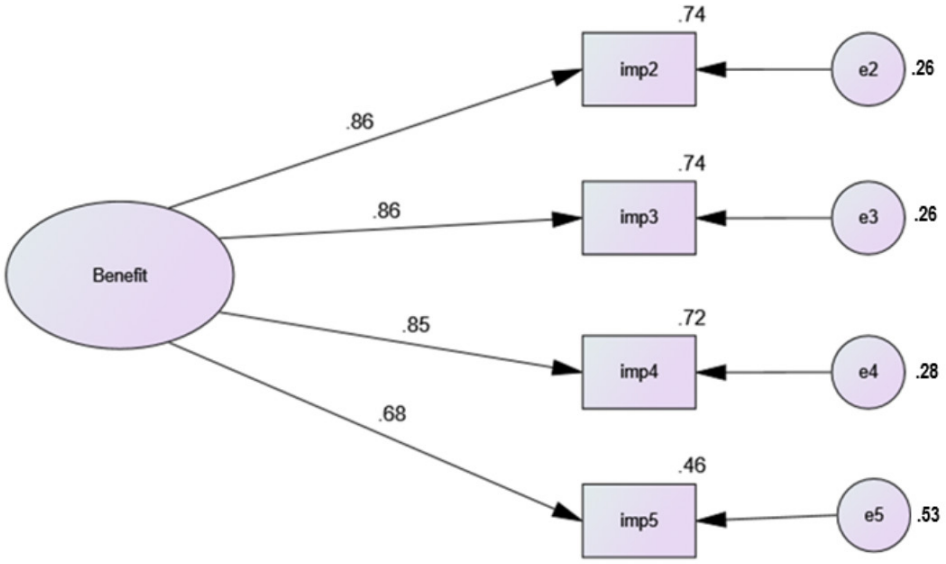

Figure 1 Study 3 results of the confirmatory factor analysis (CFA) four-item model with factor loadings with confirmatory sample of University of Washington Medicine (UW) patients (n=246). Factor loadings (standardised regression weights) are displayed on the arrow from the latent variable Benefit to observe variables (questionnaire items) abbreviated as Imp2, 3, 4, 5, which correspond to the items shown in table 5. Squared multiple correlations are displayed on top right of each variable (an indicator of stability); e2=error term. Standardised error variances displayed to right of error term. X^2 df(2)=2.95, p=0.23, confirmatory factor index (CFI)=0.998, root mean square error of approximation (RMSEA)=0.04 (0.00, 0.14) and standardised root mean square residual (SRMR)=0.012.

The factor structure of the four benefit and three risk items that were retained from study 2 was examined in the exploratory sample of study 3. The EFA using common factor analysis resulted in a two-factor, good fitting model, $X^2(8)=9.48$, p=0.30. The factors were correlated, r=−0.29. Item loadings ranged from 0.71 to 0.90 for the benefits factor, and 0.26–0.49 for the risk factor (online supplementary table 2). The Cronbach's alpha for the benefits was excellent, 0.88, whereas the alpha for the risk factor was poor, 0.30. Given that only two of the three items on the risk factor loaded >0.40 and that the internal consistency was poor, the risk items were not pursued further.

The four-item solution was further tested in the confirmatory sample using CFA to test the hypothesis that the model fit the data. All items entered into the model were normal with a kurtosis <1.0, although multivariate normality was >5.0 criterion. To account for the non-normal multivariate, the model was tested using ML estimations with bootstrapping with the sample that had no missing cases (n=246) resulting in a Bollen-Stine bootstrap p value >0.05 (p=0.37) indicating a good model fit. The CFA (n=246) resulted in an excellent fitting model as indicated by the following indices: $X^2(2)=2.95$, p=0.23, CFI=0.998, RMSEA=0.04 (0.00, 0.14), standardised RMR=0.012, and root means square residual=0.086. The factor loadings ranged from 0.68 to 0.86 (figure 1).

Study 4

The Geisinger sample from the general survey (n=250) was 65.8% female, similar to the sample in studies 1 and 2 (table 1). The average age was slightly older, 59.76 (SD 14.78), range 19–89 years. We removed 22 incomplete responses, leaving an analytical sample of 228. Prior to running the multigroup analysis, we assessed the normality of the data and RMSEA. All items assessed had a kurtosis <2.5; however, the criterion for multivariate normality is >5.0. Therefore, the model was tested using ML estimations with bootstrapping resulting in a Bollen-Stine bootstrap p value >0.05 (p=0.63) indicating a good model fit. Additionally, RMSEA=0.000, 90% CI 0.00 to 0.13. Item correlations are shown in online supplementary table 3.

Factorial invariance testing using multigroup analysis supports configural, measurement and structural invariance (see online supplementary table 4), indicating that the benefit items are assessing the same construct in both low-education and higher education groups. There were no significant differences between the unconstrained and measurement models and the structural model for the X^2 difference test. The CFI difference tests were less than the 0.01 criteria for both the measurement (0.001) and the structural model (0.006) needed for invariance.

DISCUSSION

This study compared the psychometric properties of the original benefit and risk items using Likert scale agreement responses, to a modified set of benefit and risk items using 10-point importance scale responses. While the original items had established content validity, the findings of this study suggest that the latter modified benefit items have improved psychometric properties with good construct validity. The four-item benefits scale has a depth and breadth of the benefits construct with excellent reliability. The scale reflects a sound distribution

BMJ Open: first published as 10.1136/bmjopen-2019-034517 on 20 October 2020. Downloaded from http://bmjopen.bmj.com/ on September 17, 2024 by guest. Protected by copyright.

and captures variability in patients' responses, providing researchers and clinicians with a shorter set of benefit items with which to evaluate patient experiences with open notes.

This study did not result in the development of a risk scale as the analysis did not provide evidence for the reliability of the risk construct. Scales with poor internal consistency will not adequately represent the construct. It is likely that the limited number of risk items on the patient questionnaire was not sufficient to represent this construct. Ideally, an item pool of 10–20 items would be needed for future scale development in this domain. As open notes are more widely implemented, additional risks may be observed and the item pool could be expanded. Perceived risks may contribute to why some patients choose not to read their notes, though in previous work we found the main reasons for not reading notes had to do with forgetting or not knowing they were available, or having difficulty finding notes on the portal.[2 8] We have not found evidence in surveys of substantial risks to patients who read their clinicians' visit notes.[2 8 16–18]

Although multiple validation studies are ideal,[19] these results suggest that the modified scale that assesses importance of reading notes may be useful for researchers and practitioners. Moreover, asking about the importance of a potential benefit may get at the underlying value that a patient places on reading the visit notes, better than asking a patient if he or she agrees with a statement. Longitudinal research is needed to examine how the importance scale may predict changes in reading visit notes across time. As importance increases, the commitment to reading notes after every visit may increase. Using longitudinal data, one could examine the correspondence of perceived importance to note reading behaviour. Changes in health may be associated with changes in importance of note reading. Alternatively, the importance may change over time if patients do not find reading notes beneficial, or find them repetitive. In that case, the importance of reading notes will decrease and patients will not continue accessing their notes. Tracking the perceptions of patient populations could help practitioners or healthcare systems understand open notes' value to patients or possibly determine if improvements in the note's content are needed. As features and content areas of open notes expand, the scale will facilitate tracking the impact of changes to the open notes platforms over time. The benefits scale offers a first step towards a standardised approach to assessing overall patient benefits from note reading. It will allow for comparisons between institutions and implementation practices for opening visit notes to patients. Assessment of risks and development of the risk scale will be needed to determine a fuller picture of patients' experiences with open notes.

This research defines a selected set of benefits having to do with the impact of notes on patients' personal experience of their healthcare. Others have used items from instruments or developed new items regarding benefits and risks of accessible electronic health records and visit notes. In a national patient survey in Sweden, Moll and colleagues reported benefits such as improving communication between providers and patients and making patients feel safer.[20] In a mixed methods study of patients' experiences with in Virginia, Mishra and colleagues reported that patients generally liked having access to the notes and better understood their care after reading them.[21] In another study from Sweden, Wass and Vimarlund studied a smaller sample using mixed methods and described patients' attitudes about having online access to electronic health records in general, such as making it easier to talk with clinicians and be more involved in their treatment.[22] Clinicians and researchers will likely identify other benefit domains to explore, such as the impact of note reading on health self-management behaviours and on patient safety. Further scale development will be essential for understanding the scope of patient benefits as well as the hazards of note reading.

This study has important limitations. The authors originally intended to develop a single scale that could assess both the benefits and risks of reading visit notes; however, the risk items were too few to develop a robust construct. Including both of these constructs would allow researchers to fully assess the decision-making process that influences a health behaviour, specifically the lifestyle behaviours that are important to managing one's health (eg, medication adherence, home monitoring, diet and exercise). Health behaviour change research has shown that weighing the pros (advantages) and cons (disadvantages) of engaging in a health behaviour corresponds to stages of change.[23] For example, high pro scores and low con scores correspond to the *maintenance stage of change*, described as engaging in a behaviour for 6 months or more.[23] The patterns of pros and cons could provide a fuller picture of whether the note will be read by the patient. Further exploration of both the advantages and disadvantages of constructs is recommended for future iterations of these assessments.

Another limitation of this study is that it does not examine the scale's association with reading visit notes, an objective criterion that could provide some evidence for criterion validity. Only patients who had read at least one visit note in the past 12 months were included in the study. Moreover, to examine criterion validity, a longitudinal study would be necessary to assess the consistency with which patients read their notes after each visit. It would be expected that those who read their note after each visit would have higher importance scores. A longitudinal study would also allow for predictive validity to be examined.

A strength of this study is that sample characteristics varied among the sites. Generally, BIDMC and UW respondents were highly educated with a higher perceived health status, while Geisinger patients were less educated with a lower perceived health status. Studying the three groups for the analyses provided a more diverse sample that is more likely to be generalisable to other US patient populations. Nonetheless, future studies should examine the generalisability of the scale using invariance testing

BMJ Open: first published as 10.1136/bmjopen-2019-034517 on 20 October 2020. Downloaded from http://bmjopen.bmj.com/ on September 17, 2024 by guest. Protected by copyright.

in additional demographic subgroups,[24] and could also include assessing residual invariance or invariant uniqueness, the strictest test of invariance,[25] which examines the equivalence of residuals. Future studies should examine additional aspects of validity to continue to build evidence for the scales' validity. Invariance testing can provide some evidence for the scales' generalisability to low literacy and ethnically diverse groups, while examining the predictive validity may have utility for healthcare systems interested in engaging more patients with the notes.

CONCLUSION

This study provides evidence for the construct validity of the new benefits scale. The scale may be a useful tool for researchers and healthcare systems wishing to evaluate patients' experiences with note reading and to understand the impact of note reading on various patient groups. Further research is necessary to develop the risks construct for a risk scale. As with all scale development, future research should continue to examine aspects of the scale's validity, and evaluate other domains that may be influenced by patients' note reading. Such instruments are key to building the evidence about the impact on patients of emerging technologies to engage patients in their care and improve health overall.

Author affiliations
[1]Exercise and Health Sciences, University of Massachusetts, Boston, Massachusetts, USA
[2]General Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA
[3]Harvard Medical School, Boston, Massachusetts, USA
[4]College of Nursing and Health Sciences, University of Massachusetts, Boston, Massachusetts, USA
[5]Epidemiology, University of Michigan, Ann Arbor, Michigan, USA
[6]Steele Institute for Health Innovation, Geisinger, Danville, Pennsylvania, USA
[7]Clinical Data Analytics and Decision Support, University of Arizona College of Medicine, Tucson, Arizona, USA

Contributors Conception and design of the study: JAW, SGL. Instrument development: SGL, JW, HC. Data acquisition: AF, HC, RS, DC, JW, SGL. Analysis and interpretation of data: JAW, SGL. Drafting and revision for important intellectual content: JAW, SGL, JW.

Funding This work was supported by Robert Wood Johnson Foundation (grant number 73038), Gordon and Betty Moore Foundation (grant number 4926), Peterson Center on Healthcare (grant number 16019) and Cambia Health Foundation (grant number 28584).

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not required.

Ethics approval The study was reviewed and approved by the Institutional Review Board of each participating organisation: BIDMC IRB No 2016P000226, Geisinger IRB No 2016-0193 and UW IRB No 00002358.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iDs
Julie A Wright http://orcid.org/0000-0002-5115-7011
Suzanne G Leveille http://orcid.org/0000-0003-0427-1580
Jan Walker http://orcid.org/0000-0001-9366-1200

REFERENCES

1 OpenNotes homepage, 2020. Available: https://www.opennotes.org/ [Accessed 21 Sept 2020].
2 Delbanco T, Walker J, Bell SK, *et al*. Inviting patients to read their doctors' notes: a quasi-experimental study and a look ahead. *Ann Intern Med* 2012;157:461–70.
3 Leveille SG, Walker J, Ralston JD, *et al*. Evaluating the impact of patients' online access to doctors' visit notes: designing and executing the OpenNotes project. *BMC Med Inform Decis Mak* 2012;12:32.
4 Safran DG, Karp M, Coltin K, *et al*. Measuring patients' experiences with individual primary care physicians. Results of a statewide demonstration project. *J Gen Intern Med* 2006;21:13–21.
5 Maly RC, Frank JC, Marshall GN, *et al*. Perceived efficacy in patient-physician interactions (PEPPI): validation of an instrument in older persons. *J Am Geriatr Soc* 1998;46:889–94.
6 Jervis RL, Janis I, Mann L. Decision making: a psychological analysis of conflict, choice, and commitment. *Polit Sci Q* 1978;93:134.
7 Janz NK, Becker MH. The health belief model: a decade later. *Health Educ Q* 1984;11:1–47.
8 Walker J, Leveille S, Bell S, *et al*. OpenNotes after 7 years: patient experiences with ongoing access to their clinicians' outpatient visit notes. *J Med Internet Res* 2019;21:e13876.
9 DeVellis R. *Scale development: theory and applications*. 2nd edn. Thousand Oaks, CA: SAGE Publications, 2003.
10 Noar SM. The role of structural equation modeling in scale development. *Struct Equ Model A Multidiscip J* 2003;10:622–47.
11 Goldberg LR, Velicer WF. Principles of exploratory factor analysis. In: Strack S, ed. *Differentiating normal and abnormal personality*. 2nd edn. New York, NY: Springer, 2006: 209–37.
12 Schreiber JB. Update to core reporting practices in structural equation modeling. *Res Social Adm Pharm* 2017;13:634–43.
13 Byrne B. *Structural equation modeling with amos: basic concepts, applications, and programming*. 3rd edn. New York, NY: Routledge, 2016.
14 Office of Disease Prevention and Health Promotion. Social determinants of health. Healthy people 2020, 2020. Available: https://www.healthypeople.gov/2020/topics-objectives/topic/social-determinants-of-health
15 Nunnally JC. *Psychometric theory*. 2nd edn. New York, NY: McGraw-Hill, 1978.
16 DesRoches CM, Bell SK, Dong Z, *et al*. Patients managing medications and reading their visit notes: a survey of OpenNotes participants. *Ann Intern Med* 2019;171:69
17 O'Neill S, Chimowitz H, Leveille S, *et al*. Embracing the new age of transparency: mental health patients reading their psychotherapy notes online. *J Ment Health* 2019;28:1–9.
18 Oster NV, Jackson SL, Dhanireddy S, *et al*. Patient access to online visit notes: perceptions of doctors and patients at an urban HIV/AIDS clinic. *J Int Assoc Provid AIDS Care* 2015;14:306–12.
19 Holmbeck GN, Devine KA. Editorial: an author's checklist for measure development and validation manuscripts. *J Pediatr Psychol* 2009;34:691–6.
20 Moll J, Rexhepi H, Cajander Åsa, *et al*. Patients' experiences of accessing their electronic health records: national patient survey in Sweden. *J Med Internet Res* 2018;20:e278.
21 Mishra VK, Hoyt RE, Wolver SE, *et al*. Qualitative and quantitative analysis of patients' perceptions of the patient portal experience with OpenNotes. *Appl Clin Inform* 2019;10:10–18.
22 Wass S, Vimarlund V. Same, same but different: Perceptions of patients' online access to electronic health records among healthcare professionals. *Health Informatics J* 2018;1.
23 Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. *Am J Health Promot* 1997;12:38–48.
24 Meredith W, Horn J. The role of factorial invariance in modeling growth and change. In: *New methods for the analysis of change*. Washington: American Psychological Association, 2001: 203–40.
25 Putnick DL, Bornstein MH. Measurement invariance conventions and reporting: the state of the art and future directions for psychological research. *Dev Rev* 2016;41:71–90.

BMJ Open: first published as 10.1136/bmjopen-2019-034517 on 20 October 2020. Downloaded from http://bmjopen.bmj.com/ on September 17, 2024 by guest. Protected by copyright.