BMJ Open

Development and validation of the health belief model questionnaire to promote smoking cessation for nasopharyngeal cancer prevention: a cross-sectional study

Martin Tze Wah Kueh , Fairuz Fadzilah Rahim , Abdul Rashid

To cite: Kueh MTW, Rahim FF, Rashid A. Development and validation of the health belief model questionnaire to promote smoking cessation for nasopharyngeal cancer prevention: a cross-sectional study. *BMJ Open* 2022;**12**:e057552. doi:10.1136/bmjopen-2021-057552

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2021-057552).

Received 29 September 2021
Accepted 08 July 2022

Department of Public Health Medicine, RCSI & UCD Malaysia Campus, Georgetown, Malaysia

Correspondence to
Martin Tze Wah Kueh;
martin1kueh@hotmail.com

ABSTRACT

Objective Lifestyle-induced nasopharyngeal carcinoma is a serious but preventable risk factor. This study serves to develop and validate a questionnaire that aims to predict the health behavioural intention on smoking cessation in Sarawak, Malaysia using the Health Belief Model (HBM).

Design A cross-sectional study.

Setting Urban and suburban areas in Sarawak, Malaysia.

Participants The preliminary items of the instrument were developed after extensive literature review. The instrument was translated into the Malay language using the forward-backwards method before commencing with the content validity by a panel of 10 experts. Face validity was done both quantitatively and qualitatively by 10 smokers. The construct validity of the instrument was evaluated through exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). A total of 100 smokers participated in phase 1 for EFA, while 171 smokers participated in phase 2 for CFA. Internal consistency was measured using Cronbach's alpha coefficients to evaluate the reliability.

Results In the exploratory stage, the factor loading of each item remained within the acceptable threshold. The final revised CFA yielded appropriate fit of the seven-factor model with the following model fit indices: χ^2=641.705; df=500; p<0.001; comparative fit index=0.953; Tucker-Lewis Index=0.948; root mean square error of approximation=0.041. Satisfactory convergent validity and divergent validity were shown, with the exception of one pairwise construct. The internal reliability of these scales was above the desirable threshold, with Cronbach's alpha coefficients ranging from 0.705 to 0.864 and 0.838 to 0.889 in phases 1 and 2, respectively.

Conclusions The study substantiated the instrument to be valid and reliable for predicting smokers' health behavioural intention to reduce cancer risk. The instrument is made up of 34 items, categorised into two sections, six HBM constructs and health behavioural intention. The instrument can be utilised for other smoking cessation-related cancers in different at-risk populations.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ This study established a novel instrument to assess smokers' health behavioural intention to reduce nasopharyngeal carcinoma risks, based on a well-known framework in a series of systematic validation stages. This instrument can potentially be applied to other smoking-related cancers as well.
⇒ Face validity was undertaken both qualitatively and quantitatively to sufficiently reflect the demographic during the assessments of psychological constructs. The validation was markedly aided by experts' evaluations.
⇒ This study was conducted in two phases, involving both urban and suburban smokers, to examine concept validity, convergent validity and divergent validity.
⇒ The study's generalisability may be limited because this was a cross-sectional study of a convenience sample.

INTRODUCTION

According to the WHO, tobacco smoking is a public health concern that accounts for over 8 million deaths per year and is the leading avoidable cause of illness, disability and death globally.[1] Annually, exposure to smoking is associated with 2.4 million deaths from cancer throughout the world.[2] The report from the Surgeon General of the USA associating smoking and cancer was a watershed moment in public health towards tobacco's adverse effects on human health. This was followed by the subsequent discovery that tobacco smoke comprises approximately 7000 compounds, 72 of which are carcinogenic.[3 4] Tobacco use is now causally associated with at least 20 cancer types. There are wide-ranging immediate and long-term health benefits accompanying smoking cessation.[2] However, the harmful consequences of tobacco smoking are widely neglected or underestimated, despite the fact that it remains a significant public health hazard among the impoverished, and marginalised, as well as those in

developing nations, which bear a disproportionate share of the burden.[5]

Cancer is a leading cause of death as well as a major obstacle to improving lifespan in every country.[6] There is an approximately 1% reduction in the overall cancer mortality rate across both sexes in both high-income and low-income countries.[7] However, there are variations in the frequency and distribution of aetiological aspects such as socioeconomic, geographical, genetic, biological, ethnic, social and physical factors, as well as disparities across cancer types.[7 8] For instance, nasopharyngeal cancer (NPC) is uncommon but unique among head and neck cancers with its own distinct epidemiological and risk factors. Global data from the WHO illustrate poorer outcomes of NPC in endemic areas like Southeast Asia which has an unbalanced global burden of 67%.[9]

In Malaysia, NPC is a nationwide public health concern and the fifth leading form of cancer, amounting to 4597 new diagnoses of NPC for the 2012–2016 period. A recent report from the Malaysian National Cancer Registry reported that the lifetime risk of developing NPC among men and women is 1 in 175 and 1 in 482, respectively.[10] There is a substantial geographical variance within the country, with Sarawak exhibiting a higher prevalence rate of NPC. A previous study has shown a significant high age-standardised rate in males (13.5/100 000, 95% CI=12.2 to 15.0) and females (6.2/100 000, 95% CI=5.7 to 6.7). The high-risk ethnic groups in Sarawak which include Bidayuh, Chinese, Iban, Malays and Melanau collectively rank top globally. In particular, the risk among the Bidayuh ethnic population, which is a native indigenous group, exceeds the risk for the male and female general population in Sarawak by 2.3 times and 1.9 times, respectively.[11]

This trend has been ascribed to potential risk factors, which include Epstein-Barr virus, genetic susceptibility, consumption of food with nitrous compounds and volatile nitrosamines, and complex interaction with environmental factors.[12] Among the many risk factors that are associated with NPC, tobacco smoking is the most important modifiable cause of NPC.[13 14] A meta-analysis 32 epidemiological studies (28 case–control studies and four cohort studies) on the association of tobacco smoking and NPC from 1979 to 2011 reported that tobacco-correlated NPC cases were 60% higher compared with non-smokers.[15] The Malaysian National Health and Morbidity Survey 2015 showed that the prevalence of tobacco smoking among the population in Sarawak was 25.4%. The native indigenous male (61.2%) and female (10.7%) smokers in Sarawak were among the highest nationwide.[16] Rahman *et al*[17] indicated that the average number of cigarettes smoked per day in Sarawak was 13.6.[17] This lifestyle-induced NPC is a serious concern, and necessitates a preventative strategy centred on modifying health risk behaviours.

Despite the robust establishment of cumulative impact of tobacco smoking on the risk of cancer, in Malaysia, there is still a paucity of published studies and research to evaluate the effectiveness of comprehensive strategies to promote cancer prevention among smokers. It is imperative to create a questionnaire that focuses on behavioural factors and is customised to the interests of local smokers to risk of cancer. The Health Belief Model (HBM) is an approachable theoretical model that aids in the understanding of the individual's or a smoker's belief on the health-related behavioural intention.[18 19] This HBM can predict the smokers' effort to improve health or their health-seeking behaviours for the prevention of NPC. This study serves to develop and validate a HBM-based questionnaire that aims to predict the health behavioural intention of smokers and their perspective and motivation towards smoking cessation.

Conceptual framework

The HBM is the underpinning conceptual framework for this study to focus on psychological variables to predict health behavioural intention. Developed in the 1950s by social psychologists at the US Public Health Service, the HBM is currently one of the most extensively used cognitive model and theoretical framework to help researchers understand and predict health behaviours in the population and ultimately guide health promotion and intervention activities.[18 19] A large volume of studies conducted in numerous countries, both in developed and developing countries, have utilised HBM to examine health-promoting behaviours for the prevention of different cancers.[20–22] The HBM is a value-expectancy theory, based on the hypothesis of Lewin *et al* that highlights the influence of two variables on behaviour: (1) the value that a person places on the outcome of the behaviour and (2) the person's perception of how likely the behaviour will lead to that outcome, in the event of an illness.[19] Having evolved over the past decades, HBM currently consists of six elements: (i) perceived susceptibility, (ii) perceived severity, (iii) perceived benefits, (iv) perceived barriers, (v) cues to action and (vi) self-efficacy.[23]

The first four elements refer to a person's subjective perceptions regarding (1) his/her risk of getting the disease; (2) how severe the consequences are of getting the disease; (3) the benefits from performing a health behaviour in preventing, curing or managing the disease and (4) obstacles to that health behaviour, for example, financial and time costs, side effects and so on.[23] 'Cue to action' is the stimulus, which may be internal (eg, physical sensations) or external (eg, friends with the disease and social media), that is required for that health behaviour to occur, and 'self-efficacy' refers to the person's confidence on how capable he/she is to successfully undertake that health behaviour.[19 23]

METHODS

Study design and setting

A cross-sectional study was conducted in Sarawak, Malaysia in two phases: phase 1 from October 2020 to January 2021 and phase 2 from January to April 2021. Sarawak is the largest Malaysian state situated on the island of Borneo

with a population of more than 2.6 million, made up of 26 different ethnic groups. Sarawak is divided into 12 divisions, each of which is further divided into districts and subdistricts. The two divisions selected for the study were Miri and Bintulu, which had the respective populations of 433 800 and 266 200.

The sample population in phase 1 were residents residing in urban and suburban areas in Miri, Sarawak. Phase 2 mainly involved residents of Bintulu, Kuala Baram (a federal constituency in Miri Division), and remote rural areas in Miri. They were mainly local employees working in the agricultural industry. Data were collected by eight trained research assistants. All research assistants were given a crash course in the research aims, methodology and data collection, as well as a trial run to simulate real-world situations. Prior to the distribution of the questionnaires, participants were briefed regarding the objective and methodology of the study, as well as the benefits and risks involved. Involvement in this study was entirely voluntary. On clarifying the study details, informed consent was obtained from each participant. The anonymity of the respondents' details is assured. The research assistants provided clarification to smokers who requested assistance. The methodology of the study and data collection were recorded precisely and accurately throughout the process of the research.

The inclusion criteria were as follows: (1) 18 years old and above, (2) smoked for at least a year and (3) is a Sarawakian. Participants who did not consent to participate, were pregnant or smoked e-cigarettes only were excluded from this study.

Patient and public involvement

No patient was involved in this study.

Instrument development and validation

The questionnaire was self-administered based on prior validated studies. A search was conducted using the National Library of Medicine (PubMed), Google Scholar and Cochrane Library databases by exploring various keywords: health belief model, HBM, nasopharyngeal cancer, NPC, cancer, smokers, smoking behaviours, smoking habit, and questionnaire. Questionnaires and prior literature were examined primarily, and explicitly on smokers around NPC prevention using HBM. The initial questionnaire consisted of 41 items.

The questionnaire was developed in English and translated to Malay by two local fluent bilingual translators. An experienced researcher, whose mother tongue is Malay, compared the Malay version of the questionnaire to the English version. A 'back-translation' approach to English was taken independently by another two bilingual translators based on the Beaton-recommended guidelines.[24]

To determine the content validity, 10 healthcare professionals including public health experts, hospital directors and health officers were invited to evaluate the survey instrument. Using Lawshe's model, a questionnaire was designed and organised to assist and allow the panellists to express clearly their views on the importance of including different components in a model. Experts received attachments of the questionnaire via email, which was graded on a three-point scale: essential, useful but not essential and not essential. Based on Lawshe's table, items with a content validity ratio value greater than 0.62 were retained.[25] To minimise ambiguity, the experts evaluated each item's accuracy, phrasing and grammar as well as their relevance to the construct. Modifications were made for subsequent analysis based on the experts' comments and suggestions. Overall, eight items were deleted, two items were rephrased and one item was allocated into a different construct.

Face validity was evaluated through a pilot study by 10 local smokers of different ethnic backgrounds, both qualitatively and quantitatively. Smokers who participated in the pilot study were exempted from the main study. The pilot study is conducted on a small-scale basis to ascertain the feasibility of the proposed larger study.[26] In the qualitative stage, cognitive interviews were conducted face-to-face individually to obtain participants' feedback on their comprehension and answers. Items which were not well understood were identified from the cognitive interview.[27] Minor revisions were made to better suit a linguistically and culturally diverse context. Based on the smokers' perspective, an item-oriented person's faith or spirituality to own health was also included. With the consensus of the researchers, an item was added, 'I think getting nose and throat cancer is my destiny and quitting smoking will not change it'. In the quantitative stage, a survey was disseminated based on a Likert scale of 1 (least importance) to 5 (extremely important) to determine the clinical impact of each item. The importance score was calculated based on the 'clinical impact method', in which the clinical impact of each item was determined from the proportion of participants who identified it as important. This technique was chosen for better clarity where the items were ranked according to their impact score. The mean importance score of each item was computed using the following formula: Impact Score=Frequency (Proportion)×Importance. Factors were kept if the Impact Score equal or more than 1.5. These factors were defined as deemed suitable and kept for further evaluation.[28] In the current study, the impact score for each item ranged from 1.7 to 4.6, therefore, no item was eliminated.

Subsequently, in the main study, to determine the construct validity, exploratory and confirmatory factor analyses were performed in two periods. In the first period, EFA (n=100) was used to determine the number of latent factors or the relationships between the common factors. The model was later adjusted in the second period with CFA (n=171) via structural equations using AMOS. CFA confirmed the overall fit of the model and indicated that the measures were in the acceptable range.[29] Convergent validity and discriminant validity were also carried out in phase 2. The flow diagram for the questionnaire development and validation is shown in figure 1.

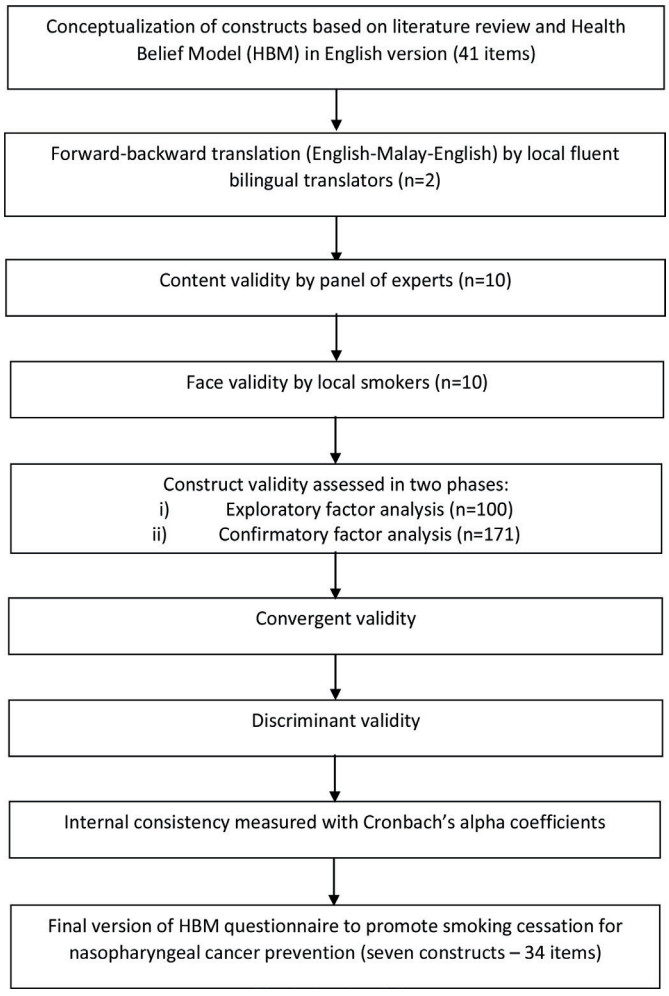

Figure 1 Flow diagram of the development and validation of the Health Belief Model questionnaire to promote smoking cessation for nasopharyngeal cancer prevention.

Final scoring of the instrument

In the main study, the total mean score for each HBM component was formulated on the 5-point Likert scale options from strongly disagree to strongly agree, with scores ranging from 1 to 5 points. With the exception of 'Perceived Barriers', which is inversely proportional, a greater score reflects a firmer desire to quit smoking.

Data management

The data were processed in Microsoft Excel before being analysed in SPSS V.26.0 and AMOS V.23.0.[30] Listwise deletion approach was done for missing data of less than 2%. Sociodemographic characteristics are presented as number and percentage distribution. Cronbach's alpha coefficients were used to determine internal consistency. Exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) were used to test the construct validity of each construct. Specifically, EFA was evaluated in phase 1 and CFA in phase 2. A p-value below 0.05 was deemed to be statistically significant.

EFA was performed in phase 1 to reveal the fundamental structure of a large set of variables.[29] The factors were extracted using principal component analysis with a varimax rotation. The Kaiser-Meyer-Olkin (KMO)>0.6 and Bartlett's test for sphericity ($p<0.05$) were used for adequacy and item checking. All loading factors below 0.3 were excluded from the constructs.[31]

CFA was performed in phase 2 to assess the data integrity and the structural model.[29] The acceptable level of standardised factor loading was set at 0.5 and above to ensure a satisfactory association between items and corresponding factors.[32] Different fit indices were utilised to estimate the model fit. These include a comparative fit index (CFI) of>0.90, Tucker-Lewis Index (TLI)>0.90 and root mean square error of approximation (RMSEA)<0.05.[33] Additionally, convergent validity and discriminant validity were evaluated based on the composite reliability (CR) and average variance extracted (AVE). The HBM components were analysed with the minimum and maximum scores, total mean score and SD.

RESULTS

A total of 100 and 171 smokers participated in phases 1 and 2 of the study, respectively. The response rate for phase 1 was 100%, whereas in phase 2, the response rate was 98.3% (171/174). The majority of the participants were males (phase 1, 86.0%; phase 2, 89.5%) and were in the 30–39 age group (phase 1, 40.0%; phase 2, 40.9%). A little over one-third of the participants smoked for more than 10 years for both phase 1 (37%) and phase 2 (40.4%). In phase 1, most of the participants were Iban (34%), followed by Malay (19%), Chinese (18%), Others (13%), Melanau (9%) and Bidayuh (7%). Most of the participants in phase 2 were Chinese (33.9%), followed by Iban (24.6%), Malay (12.3%), Bidayuh (9.9%), Melanau (9.9%) and Others (9.4%). Details of the smokers' demographic information are presented in table 1.

In phase 1, KMO Measure of Sampling Adequacy was 0.697 and Barlett's test of the sphericity was significant (χ^2=1746, p-value<0.001). EFA was conducted to analyse the factor structure with principal component analysis with a varimax rotation. A decision was made to go for seven-factor structures since there is clarity of seven constructs. The EFA found that seven variables had eigenvalues larger than Kaiser's threshold of 1 and explained 63.0% of the variance when combined. Factor loadings higher than 0.3 were found in all the items. Four items had cross-loading with values greater than 0.3, which are PBar5, HBI2, HBI3 and HBI4. All items remained because the contents of the items were regarded as relevant based on the decision and judgement of the researchers. Table 2 shows the EFA with total items and the factor loading of each construct for the seven-factor model.

CFA was performed in phase 2 to assess whether the seven-factor model indicated by the EFA could sufficiently represent the data. The items in their respective constructs were loaded between 0.586 and 0.898 (table 3). For the model's fitness to increase, items with less than 0.6 and a MI of more than 10 should have been eliminated. However, they were kept because they were

Table 1 Demographic characteristics of participants

	Phase 1 (n=100)		Phase 2 (n=171)	
	Number	Percentage (%)	Number	Percentage (%)
Age				
18–29	34	34.0	46	26.9
30–39	40	40.0	70	40.9
40–49	14	14.0	43	25.1
50–64	12	12.0	10	5.8
65 and above	0	0.0	2	1.2
Gender				
Male	86	86.0	153	89.5
Female	14	14.0	18	10.5
Ethnic groups				
Malay	19	19.0	21	12.3
Chinese	18	18.0	58	33.9
Bidayuh	7	7.0	17	9.9
Iban	34	34.0	42	24.6
Melanau	9	9.0	17	9.9
Others	13	13.0	16	9.4
Years of smoking				
1–5 years	32	32.0	44	25.7
6–10 years	31	31.0	58	33.9
More than 10 years	37	37.0	69	40.4

essential for the conceptual framework. Before arriving at the final model, six pairs of correlated errors were added to improve robustness. The resulting model was suitable for testing, as evidenced by the following model fit indices: χ^2=641.705; df=500; p<0.001; CFI=0.953; TLI=0.948; RMSEA=0.041 (90% CI=0.031 to 0.050).

The AVE and CR values, which are listed in table 3, were obtained after the structural model's fit was investigated to check if the items were loaded appropriately. The AVE readings were all over the cut-off value of 0.5, ranging from 0.512 to 0.617. The CR values were all over the cut-off value of 0.7, ranging from 0.839 to 0.888. All seven constructs featured sufficient convergent validity.

Discriminant validity was evaluated using Fornell-Larcker criteria by comparing the squared correlations and AVE scores for each of the pairwise constructs.[34] With the exception of perceived benefit < — > cue to action, all paired constructs have shown established discriminant validity (table 4).

Internal consistency was deemed to be acceptable if Cronbach's alpha coefficients were more than 0.7. According to reliability analyses, Cronbach α of the perceived susceptibility, perceived severity, perceived benefits, perceived barriers, cues to action, self-efficacy and health behavioural intention were 0.83, 0.81, 0.86, 0.80, 0.81, 0.85 and 0.71, respectively, in phase 1 and 0.87, 0.88, 0.86, 0.89, 0.84, 0.87 and 0.86, respectively, in phase 2.

The detailed mean and SD of each HBM component and health behavioural intention to the sociodemographic characteristics are displayed in table 5. The total mean score for each domain ranged from 14.03 to 21.32.

Discussion

At the time when the study was conducted, there were no published papers based on HBM to evaluate cancer health perception among at-risk smokers. Geography, ethnicity, national, social and genetic-related factors contribute to the disproportionate burden of cancer. Sarawak's population vulnerability to NPC is among the highest in the world. Although genetic predisposition may be the most important risk factor leading to higher incidence of NPC in Sarawak, individual behavioural variables are a key driver of community health that must not be underestimated.[35]

A total of 34 items in the questionnaire were formulated consistent with the HBM and divided into two sections: HBM scale for smokers' perception of NPC and health behavioural intention to smoking cessation. Both sections are constructed with a 5-point Likert scale ranging from '1=strongly disagree' to '5=strongly agree'. The first section consists of 30 items and is arranged into six subcategories, each representing the six constructs of the health belief model—perceived susceptibility (five items), perceived seriousness (five items), perceived benefits (four items), perceived barriers (six items), cues

Table 2 Result of exploratory factor analysis in phase 1 (n=100)

Constructs	Items	Component						
		1	2	3	4	5	6	7
Perceived susceptibility	PSus1			0.665				
	PSus2			0.750				
	PSus3			0.810				
	PSus4			0.548				
	PSus5			0.736				
Perceived severity	PSev1						0.558	
	PSev2						0.720	
	PSev3						0.744	
	PSev4						0.843	
	PSev5						0.730	
Perceived benefit	PBen1	0.767						
	PBen2	0.765						
	PBen3	0.763						
	PBen4	0.792						
Perceived barrier	PBar1					0.594		
	PBar2					0.725		
	PBar3					0.658		
	PBar4					0.709		
	PBar5					0.590		
	PBar6					0.617		
Cue to action	CUE1				0.721			
	CUE2				0.584			
	CUE3				0.736			
	CUE4				0.723			
	CUE5				0.745			
Self-efficacy	EFF1		0.721					
	EFF2		0.730					
	EFF3		0.770					
	EFF4		0.868					
	EFF5		0.625					
Health behavioural intention	HBI1							0.681
	HBI2							0.307
	HBI3							0.733
	HBI4							0.596
Rotation sums of squared loading								
Total		3.469	3.371	3.246	3.072	3.071	3.012	2.185
Percentage of variance		10.203	9.913	9.548	9.034	9.032	8.859	6.425
Cumulative percentage		10.203	20.116	29.664	38.698	47.730	56.588	63.014

to action (five items) and self-efficacy (five items). The second section includes four items that predict health behavioural intention (see online supplemental file).

Based on the validity and reliability tests, including face and content validity, construct validity and internal consistency, the findings of the current study indicate that the questionnaire has promising psychometric properties. Ten experts advised on content validity, and eight items that did not reach the threshold of CVR based on Lawshe's table and were judged superfluous were removed.[25] Face validity was examined in a pilot study with 10 smokers who fulfilled the eligibility requirements to ensure cultural acceptance and assess relevance and readability within the local community. Given a satisfactory impact score of each item, no item was eliminated in the face validity stage. In the main study, at-risk

Table 3 Result of confirmatory factor analysis in phase 2 (n=171)

Constructs	Items	Factor loadings	AVE	CR
Perceived susceptibility	PSus1	0.898	0.577	0.871
	PSus2	0.819		
	PSus3	0.739		
	PSus4	0.685		
	PSus5	0.683		
Perceived severity	PSev1	0.680	0.597	0.881
	PSev2	0.675		
	PSev3	0.806		
	PSev4	0.867		
	PSev5	0.766		
Perceived benefit	PBen1	0.752	0.603	0.858
	PBen2	0.812		
	PBen3	0.740		
	PBen4	0.733		
Perceived barrier	PBar1	0.586	0.572	0.888
	PBar2	0.670		
	PBar3	0.626		
	PBar4	0.886		
	PBar5	0.878		
	PBar6	0.770		
Cue to action	CUE1	0.646	0.512	0.839
	CUE2	0.721		
	CUE3	0.727		
	CUE4	0.767		
	CUE5	0.723		
Self-efficacy	EFF1	0.645	0.574	0.869
	EFF2	0.694		
	EFF3	0.827		
	EFF4	0.859		
	EFF5	0.744		
Health behavioural intention	HBI1	0.654	0.617	0.864
	HBI2	0.836		
	HBI3	0.867		
	HBI4	0.773		

AVE, average variance extracted; CR, composite reliability.

smokers from various ethnic groups participated, which was conducted in urban, suburban and rural regions of Miri (the northern region of Sarawak). The KMO test yielded a result of 0.697 (phase 1) and 0.830 (phase 2), while Bartlett's test of sphericity obtained 1746.76 (phase 1) and 3362.86 (phase 2), both with p-value<0.001, indicating that the sample size was adequate and the correlation between the items was sufficient for factor analysis.

Construct validity primarily concerns with the degree to which a concept measures what it claims to measure.[36] Similar to a previous study,[37] a number of analyses were conducted to assess the construct validity, including EFA, CFA, as well as convergent and discriminant validity. The EFA demonstrated in phase 1 that the seven-factor structure accounted for 63.01% of the overall variance. The cut-off point for factor loading was fixed at 0.3. According to EFA suggestion, HBI2 item (I am trying to quit smoking to prevent nose and throat cancer at this time) should be grouped together with perceived benefit since the factor loading (0.537) is higher than when it is grouped with health behavioural intention (0.307). This item requires immediate smoking cessation, which may

Table 4 Result of discriminant validity in phase 2 (n=171)

	Factor correlation	Correlation squared	Discriminant validity
Perceived susceptibility<-->perceived severity	0.312	0.097	Established
Perceived susceptibility <--> perceived benefit	0.260	0.068	Established
Perceived susceptibility <--> perceived barrier	0.204	0.042	Established
Perceived susceptibility <--> cue to action	0.172	0.030	Established
Perceived susceptibility <--> self-efficacy	0.236	0.056	Established
Perceived severity <--> perceived benefit	0.575	0.331	Established
Perceived severity <--> perceived barrier	–0.042	0.176	Established
Perceived severity <--> cue to action	0.238	0.057	Established
Perceived severity <--> self-efficacy	0.257	0.066	Established
Perceived benefit <--> perceived barrier	–0.085	0.007	Established
Perceived benefit <--> cue to action	0.826	0.682	Not established
Perceived benefit <--> self-efficacy	0.349	0.122	Established
Perceived barrier <--> cue to action	–0.001	0.000	Established
Perceived barrier <--> self-efficacy	–0.157	0.025	Established
Cue to action <--> self-efficacy	0.600	0.360	Established

spark inconsistent answers as some respondents may not be prepared to quit. Cronbach's α coefficient value of 'health behavioural intention' construct appeared to be satisfactory (phase 1: 0.705; phase 2: 0.861) and this item could potentially be essential on a larger scale. Thus, the researchers agreed not to delete this item.

In phase 2, CFA was performed to see if the seven-factor model derived by EFA could validate the association among the items based on the chosen framework. The loadings of all factors were more than 0.5. The goodness of fit was demonstrated to have an acceptable fit with the data, χ^2=641.705; df=500; p<0.001; CFI=0.953; TLI=0.948; RMSEA=0.041 (90% CI=0.031 to 0.050). Following CFA, convergent and discriminant validity were tested in phase 2. The CR and AVE values for each component have to be higher than 0.7 and 0.5; respectively, which are fulfilled in the current study, suggesting an acceptable convergent validity.[38] Discriminant validity was evaluated between the HBM constructs (excluding health behavioural intention construct). To establish acceptable discriminant validity, the factors' correlation coefficients with other factors must not be greater than each factor's AVE square root.[34] The current findings demonstrated established discriminant validity for all except for perceived benefit <–> cue to action. The explanation for this might be that the greater the perceived benefits of smoking cessation, the more likely smokers will look for cues to participate in such health-promoting behaviour, or vice versa. Future studies could delve deeper into the strength of the correlation, particularly between perceived benefit and cue to action. In terms of reliability, each construct for both phases 1 and 2 showed rationally acceptable Cronbach's α coefficient values as all of which were higher than 0.7, which demonstrates a high internal consistency.[39]

This study has several limitations. First, it is a cross-sectional study using convenience sampling, and thus, it is susceptible to recall and selection bias. However, to mitigate recall bias, participants were given ample time to consider their responses before answering the questions. This study was carried out with the cooperation of smokers from accessible areas in Sarawak. However, due to the widespread locations, the population sizes were difficult to determine, which contributed to the second limitation. The third limitation was the relatively small sample size for both EFA (n=100) and CFA (n=171). However, Kline[40] indicated that for EFA, a sample size of 100 is sufficient,

Table 5 Result of the minimum and maximum scores, total mean score and SD in phase 2 (n=171)

Constructs	N	Minimum	Maximum	Mean	SD
Perceived susceptibility	171	5	25	14.03	4.04
Perceived severity	171	5	25	16.81	4.44
Perceived benefits	171	7	20	16.16	2.86
Perceived barriers	171	6	30	21.32	4.99
Cue to action	171	9	25	19.97	3.14
Self-efficacy	171	9	25	18.04	3.87
Health behavioural intention	171	5	20	15.64	3.26

while Anderson and Gerbing[41] suggested that CFA/SEM may be reliably examined with a minimum sample size of 100–150.[40 41]

This study provides practical implications. With it being valid and reliable, the public health officials and researchers now have a reason to launch a larger population-based study on health behavioural intention to minimise NPC. Given that the current smoking rates in Malaysia remain high, this questionnaire can help in understanding and determining the factors that influence smokers' health decisions. Subsequently, cancer risk can be reduced by better prediction, a comprehensive tobacco control programme, policy creation and health interventions.

Conclusions

In order to examine the variables affecting smoking cessation for cancer prevention, this study made an effort to develop a comprehensive HBM-based questionnaire. The results depict consistently satisfactory psychometric properties, confirming the validity and reliability. Considering that smoking is a major contributor to cancer, it is critical to address the health behavioural intention to uncover obstacles and implement improvements for a more successful intervention. The authors propose further studies to use the instruments for application in other smoking-related cancers in different susceptible populations and geographical locations.

Acknowledgements We express our gratitude to the hospital director in Miri and local estate managers for their helpful assistance and the participants of this study.

Contributors Conceptualisation: MTWK, FFR and AR. Formal analysis, investigation, methodology, project administration, resources: MTWK and FFR. Software: MTWK. Supervision: FFR and AR. Validation: MTWK. Visualisation: MTWK. Writing—original draft: MTWK. Writing—review and editing: MTWK, FFR and AR. Guarantor: MTWK.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval This study involves human participants. The study was approved by Joint Penang Ethics Committee (approval no. JPEC 20 0027). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository.

ORCID iDs
Martin Tze Wah Kueh http://orcid.org/0000-0002-7473-343X
Fairuz Fadzilah Rahim http://orcid.org/0000-0002-0248-6046
Abdul Rashid http://orcid.org/0000-0003-1520-8480

REFERENCES

1 Institute for Health Metrics and Evaluation (IHME). Findings from the Global Burden of Disease Study 2017. In: *IHME*. Seattle, WA, 2018. http://www.healthdata.org/policy-report/findings-global-burden-disease-study-2017
2 World Health Organization. Who report on cancer: setting priorities, investing wisely and providing care for all, 2020. Available: https://apps.who.int/iris/handle/10665/330745
3 Wellmann KF. [SMOKING AND HEALTH. ON THE REPORT OF THE ADVISORY COMMITTEE TO THE SURGEON GENERAL OF THE PUBLIC HEALTH SERVICE]. *Dtsch Med Wochenschr* 1964;89:1085–6.
4 Hecht SS. Research opportunities related to establishing standards for tobacco products under the family smoking prevention and tobacco control act. *Nicotine Tob Res* 2012;14:18–28.
5 Viswanath K, Herbst RS, Land SR, *et al*. Tobacco and cancer: an American association for cancer research policy statement. *Cancer Res* 2010;70:3419–30.
6 Sung H, Ferlay J, Siegel RL, *et al*. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2021;71:209–49.
7 Hashim D, Boffetta P, La Vecchia C, *et al*. The global decrease in cancer mortality: trends and disparities. *Ann Oncol* 2016;27:926–33.
8 Zavala VA, Bracci PM, Carethers JM, *et al*. Cancer health disparities in racial/ethnic minorities in the United States. *Br J Cancer* 2021;124:315–32.
9 Lam K-O, Lee AWM, Choi C-W, *et al*. Global pattern of nasopharyngeal cancer: correlation of outcome with access to radiation therapy. *Int J Radiat Oncol Biol Phys* 2016;94:1106–12.
10 Azizah A, Hashimah B, Nirmal K, Siti Zubaidah A, Puteri N, Nabihah A. Malaysian National Cancer Registry Report 2012-2016 [Internet]. Malaysia National Cancer Institute; 2019. https://www.moh.gov.my/moh/resources/Penerbitan/Laporan/Umum/2012-2016%20(MNCRR)/MNCR_2012-2016_FINAL_(PUBLISHED_2019).pdf
11 Devi BCR, Pisani P, Tang TS, *et al*. High incidence of nasopharyngeal carcinoma in native people of Sarawak, Borneo island. *Cancer Epidemiol Biomarkers Prev* 2004;13:482–6.
12 Tsao SW, Yip YL, Tsang CM, *et al*. Etiological factors of nasopharyngeal carcinoma. *Oral Oncol* 2014;50:330–8.
13 Linton RE, Daker M, Khoo AS-B, *et al*. Nasopharyngeal carcinoma among the Bidayuh of Sarawak, Malaysia: history and risk factors. *Oncol Lett* 2021;22:514.
14 Guo S-S, Huang P-Y, Chen Q-Y, *et al*. The impact of smoking on the clinical outcome of locoregionally advanced nasopharyngeal carcinoma after chemoradiotherapy. *Radiat Oncol* 2014;9:246.
15 Xue W-Q, Qin H-D, Ruan H-L, *et al*. Quantitative association of tobacco smoking with the risk of nasopharyngeal carcinoma: a comprehensive meta-analysis of studies conducted between 1979 and 2011. *Am J Epidemiol* 2013;178:325–38.
16 Institute for Public Health. National health and morbidity survey 2015 (NHMS 2015) Ministry of Health Malaysia Kuala Lumpur; 2015. https://www.moh.gov.my/moh/resources/nhmsreport2015vol2.pdf
17 Rahman M, Arif M, Razak MFA. Factors associated with tobacco use among the adult population in Sarawak, Malaysia: a cross sectional study. *Epidemiol Biostat Public Health* 2015;12:1–9.
18 Champion VL, Skinner CS. The health belief model. In: *Health behavior and health education: theory, research, and practice*. 4th ed. San Francisco, CA, US: Jossey-Bass, 2008: 45–65.
19 Maiman LA, Becker MH. The health belief model: origins and correlates in psychological theory. *Health Educ Monogr* 1974;2:336–53.
20 Odedina FT, Dagne G, Pressey S, *et al*. Prostate cancer health and cultural beliefs of black men: the Florida prostate cancer disparity project. *Infect Agent Cancer* 2011;6:S10.
21 Solomon K, Tamire M, Kaba M. Predictors of cervical cancer screening practice among HIV positive women attending adult anti-retroviral treatment clinics in Bishoftu town, Ethiopia: the application of a health belief model. *BMC Cancer* 2019;19:989.
22 Tavafian SS, Hasani L, Aghamolaei T, *et al*. Prediction of breast self-examination in a sample of Iranian women: an application of the health belief model. *BMC Womens Health* 2009;9:37.
23 LaMorte. WW. The health belief model: behavioral change models, 2019. Available: http://sphweb.bumc.bu.edu/otlt/MPH-Modules/SB/BehavioralChangeTheories/BehavioralChangeTheories2.html

24 Beaton DE, Bombardier C, Guillemin F, *et al*. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 2000;25:3186–91.
25 Lawshe CH. A quantitative approach to content validity. *Pers Psychol* 1975;28:563–75.
26 Ferreira Antunes JL, Samet JM, Wipfli H, Platz EA. A dictionary of epidemiology. *J Epidemiol Community Health* 2009;63:1449–51.
27 Boateng GO, Neilands TB, Frongillo EA, *et al*. Best practices for developing and validating scales for health, social, and behavioral research: a primer. *Front Public Health* 2018;6:149.
28 Lacasse Y, Godbout C, Sériès F. Health-related quality of life in obstructive sleep apnoea. *Eur Respir J* 2002;19:499–503.
29 Suhr DD. Exploratory or confirmatory factor analysis? 2006. Available: https://support.sas.com/resources/papers/proceedings/proceedings/sugi31/200-31.pdf
30 Kueh M, Rahim FF, Rashid A. Data from: development and validation of the health behavioural intention on smoking cessation to prevent nasopharyngeal cancer in Sarawak, Malaysia based on the health belief model: a cross-sectional study. *Dryad Digital Repository* 2021.
31 Munro BH. *Statistical methods for health care research*. lippincott williams & wilkins, 2005.
32 Xiang B, Wong HM, Cao W, *et al*. Development and validation of the oral health behavior questionnaire for adolescents based on the health belief model (OHBQAHBM). *BMC Public Health* 2020;20:701.
33 Brown TA. *Confirmatory factor analysis for applied research*. New York, NY, US: The Guilford Press, 2006.
34 Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. *J Mark Res* 1981;18:39–50.
35 McGinnis JM, Williams-Russo P, Knickman JR. The case for more active policy attention to health promotion. *Health Aff* 2002;21:78–93.
36 Cohen RJ, Swerdlik ME, Phillips SM. *Psychological testing and assessment: an introduction to tests and measurement*. Mayfield Publishing Co, 1996.
37 Shamsalinia A, Moradi M, Farahani MA, *et al*. Designing and psychometric evaluation of disease-related fear scale (D-RFS) in adults with epilepsy: a sequential exploratory mixed methods design. *Epilepsy Behav* 2020;110:107169.
38 Schreiber JB, Nora A, Stage FK, *et al*. Reporting structural equation modeling and confirmatory factor analysis results: a review. *J Educ Res* 2006;99:323–38.
39 Kline P. *The Handbook of psychological testing*. Psychology Press, 2000.
40 Kline P. *An easy guide to factor analysis*. 1st ed. London, 1994: 208.
41 Anderson JC, Gerbing DW. Structural equation modeling in practice: a review and recommended two-step approach. *Psychol Bull* 1988;103:411–23.