Appendix 18 – THS-PBA-01-US

Summary Document

PBA01 – Scales Development and Validation Project

PMI RESEARCH & DEVELOPMENT

Summary Document
PBA01 – Scales Development and Validation Project

Sponsor: Philip Morris Products S.A.
PMI Research & Development
Quai Jeanrenaud 5
2000 Neuchâtel, Switzerland

Authors: Stefan Cano, Methodological Consultant, ScaleReport, UK
Thomas Salzberger, Methodological Consultant, University of Vienna, Austria
Christelle Chrea, Senior Behavioral Scientist, Philip Morris Products S.A.

TABLE OF CONTENTS

TABLE OF TABLES

TABLE OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	Definition
ASCQ	Adolescent Smoking Consequences Questionnaire
CC(s)	conventional cigarette(s)
CDI(s)	cognitive debriefing interviews
(C)ITC	(Corrected) Item-Total Correlation
CI	confidence interval
CTT	Classical Test Theory
df	degrees of freedom
DIF	differential Item Functioning
E-CIG	electronic cigarette
EC	Ethics Committee
FDA	Food and Drug Administration
HPHCs	harmful and potentially harmful constituents
ICF	informed consent form
IIC	Inter-Item Correlation
IOM	Institute of Medicine
IRB	Institutional Review Board
ITUQ	Intent to Use Questionnaire
KOL(s)	key opinion leader(s)
MRTP(s)	Modified Risk Tobacco Product(s)
n	number of observation
N	sample size
NLST	National Lung Screening Trial

Abbreviation	Definition
NRT(s)	nicotine replacement therapy product(s)
p	p value (associated with test statistic)
PBA	Perception and Behavior Assessment
PMI	Philip Morris International
PRI	Perceived Risk Instrument, see PRI-G and PRI-P
PRI-G	Perceived Risk Instrument applied to risk to users in general
PRI-P	Perceived Risk Instrument applied to risk to the individual participant
PRO	patient-reported outcome
PSI	person separation index
QSI	Quantitative Study I
QSII	Quantitative Study II
RMT	Rasch Measurement Theory
RRP(s)	Reduced Risk Product(s) (see definition p.10)
SCQ	The Smoking Consequences Questionnaire
SCQ-A	The Smoking Consequences Questionnaire Adult version
SD	standard deviation
SRD(s)	smoking-related disease(s)
t	t value (Student's t-test)
TCQ	Tobacco Craving Questionnaire
THS	Tobacco Heating System
UK	United Kingdom
US	United States

Abbreviation	Definition
VAS	visual analog scale
WHO	World Health Organization
WLE	weighted likelihood estimation

DEFINITION OF TERMS

The following explains how special terms are used in this report.

Term	Definition
Cessation	Having smoked conventional cigarettes in the past but having successfully stopped smoking and not using any tobacco or nicotine-containing products. It is referred to as an Object.
Classical Test Theory	A body of related psychometric theory whose aim is to understand and improve the reliability and validity of psychological tests based on the fundamental assumption that each person has a true score that would be obtained if there were no errors in measurement.
Conceptual framework	Organization of ideas to achieve the purpose of the instrument development, comprising a definition of the construct to be measured and its limits of applicability.
Construct	Not directly observable, hypothetical property or attribute measured by the instrument.
Corrected item-total correlations	Correlation between an item and the total score (excluding that item). As the item of interest has been omitted, this is corrected for overlap.
Cronbach's coefficient alpha	A measure of internal consistency reliability computed from the intercorrelations between the items of a scale. Theoretically, and under specific circumstances, alpha is an estimate of the mean of all possible split-half reliabilities.
Data quality	Completeness of item- and scale-level data.
Differential item functioning	The presence of bias in an item examined through the differences observed between the different levels of a person factor (e.g., gender).
Discrimination, item discrimination	In the Rasch model, item discrimination indicates the extent to which higher scores on an item correspond to higher scores on the whole instrument. Technically, discrimination is the slope of the empirical item characteristic curve, which describes the relationship of average scores and estimated person measures.
Domain	Represented by a set of items of an instrument, a domain measures a part of a larger construct comprised of multiple domains.

Term	Definition
Fit residual, item fit residual	In the Rasch model, item fit statistic that indicates the actual discrimination of the item.
Floor/ceiling effects	Floor effects refer to the proportion of the sample at minimum possible scale or item score; ceiling effects refer to proportion of the sample at the maximum possible scale or item score.
Inter-item correlations	Correlations between individual items in a scale; typically reported as the average and the range of correlation coefficients.
Internal consistency	The extent to which items in a scale are positively related to each other; typically assessed by Cronbach's coefficient alpha.
Intent to Use Instrument/ Questionnaire	Intent to Use Instrument, the name for the scale under development up to the conclusion of Quantitative Study I, when it was renamed Intent to Use Questionnaire
Item fit	In the Rasch model, the degree to which observed item responses are consistent with the expected item responses predicted by a mathematical model – here the Rasch model.
Item locations	The position of items along a line (continuum) representing the construct of interest.
Item characteristic curve	A graph describing the relationship between the ability (trait, attitude, perception, etc.) of examinees and the probability of answering a given item correctly (ability) or endorsing it (attitude, perception).
Person measure, Participant measure	Represents, for a participant, the level of perceived risk scaled onto a latent dimension (e.g., perceived health risk) and associated to the tobacco product the scale has been applied to. This measure is derived from the sum raw score based on the Rasch Measurement Theory. In the original metric of the Rasch model, the point of origin is defined by the mean item location. The person measures can be transformed to a more convenient metric by applying a linear transformation rule.
Person fit	In the Rasch model, the degree to which the response pattern of a participant matches the expected response pattern given the estimated item parameters.

Term	Definition
Person separation index (PSI)	A reliability statistic, comparable to Cronbach's alpha but computed from linear person measurements rather than raw summed scores. It quantifies the error associated with the measurements of people in this sample. Note: the PSI for any scale is sample dependent.
Psychometrics	Methods for constructing and evaluating measurement scales and measurement properties.
Rasch Measurement Theory	A sophisticated psychometric technique for constructing and evaluating rating scales, and for analyzing rating scale data. It tests the extent to which a scale is working as a measurement instrument, and if performing as such, enables linear measurements and fit statistics) to be constructed from the ordered category responses of rating scale items.
Reduced Risk Products	Reduced Risk Products is the term PMI uses to refer to products with the potential to reduce individual risk and population harm in comparison to smoking cigarettes.
Reliability	The degree to which a measure is free from random error.
Reproducibility (test-retest reliability)	Assesses whether a scale yields the same results on repeated applications (in respondents who have not changed on the construct being measured).
Scaling assumptions	Extent to which it is legitimate to sum a set of item scores, without weighting or standardisation, to produce a single total score. Justified by evidence of data meeting the model assumptions.
Score, person score, raw score	The unweighted sum score of a participant’s responses to items forming a scale.
Targeting	Extent to which the theoretical range of the variable measured by the scale matches the actual range of that variable in the study sample.
Thresholds for item response options	Points of crossover on the latent trait measured between two adjacent response categories. The point on the latent continuum at which the probability of a person responding to two adjacent categories (e.g., 0 and 1) is equal (50%).

1 EXECUTIVE SUMMARY

The Scales Development and Validation Project (PBA01) aimed at developing scientifically validated measurement instruments for two concepts: (1) the *Perceived Risks* associated with the use of tobacco and nicotine-containing products and (2) the *Intent to Use* a tobacco product in the future (including intention to try and intention to use). This project led to the development of the *Perceived Risk Instrument* (PRI), a self-report psychometrically validated instrument designed to assess the perceived risks associated with the use of tobacco and nicotine-containing products including conventional cigarettes (CCs), nicotine replacement therapy products (NRTs), and reduced risk tobacco products (RRPs, see definition p.10), also refered by the US Food Drug and Administration as modified risk tobacco products (MRTPs). The PRI can also be used to assess the perceived risks associated with the past use of CC (e.g., "*Cessation*", or having successfully quitting smoking CC and not using any tobacco or nicotine-containing product). In addition, this project led to the development of the *Intent to Use Questionnaire* (ITUQ), comprised of two sets of single items assessing for CC, NTPs and NRTs: (i) the intention to try (at least once) and (ii) intention to use (on a regular basis). The PRI and ITUQ were designed to be administered to different consumer groups based their smoking status (i.e., adult smokers, adult former smokers and adult never smokers).

In this summary document, an overview of the parallel development conducted for the PRI and ITUQ is provided, describing more specifically (1) the target population; (2) the target concepts of interest measured by the instruments; and (3) the validation of these instruments. The PRI and ITUQ are based on an underlying conceptual framework developed from a range of extensive qualitative studies including:

- focus groups (12 focus groups in US population, total n=93); and 17 focus groups in Italian, UK and Japanese populations [total n=136] with subpopulations of adult current smokers, adult former smokers, and adult never smokers),
- a literature review, and
- input from an advisory board panel of experts.
- The preliminary instruments went then through cognitive debriefing interviews (CDIs) with UK population (n=40) and US-population (n=48).

The UK-US English version of the draft instruments – PRI and ITUQ - were subsequently field-tested through 2 quantitative cross-sectional US-based web-surveys (Survey 1, 2020 completers and Survey 2, 1640 completers) to evaluate their psychometric properties. Psychometric evaluation was based on both Classical Test Theory (CTT) as well as Rasch Measurement Theory (RMT).

Survey 1 data analyses aimed at identifying the best performing items to form a scale for each of the draft instruments. With regard to the concept of Perceived Risks, this led to the formation of an instrument comprised of an 18-item *Perceived Health Risk* scale and a 7-item *Perceived Addiction Risk* scale, plus 2 standalone items for *Perceived Harm to Others*. With regard to the concept of Intent to Use, the items constituting the draft Instrument did not show satisfactory psychometric properties to justify these being treated as a comformable set to form a total score. Thus, these items are to be considered

as standalone single categorical items, which could be selected on a study-by-study basis within the *Intent to Use Questionnaire* (ITUQ). Findings from Survey 2 supported the summing of items to form a total score for each of the two scales for the PRI (i.e. the 18-item *Perceived Health Risk* scale and the 7-item *Perceived Addiction Risk* scale). Importantly, the stability of the scales across demographics (age, sex, and education), different tobacco products, and smoking status groups was supported by the absence of differential item functioning (DIF). All data surpassed psychometric criteria supporting the conclusions that the PRI is a psychometrically robust instrument applicable for general and personal risk perception measurement for different types of products (including CCs, NRTs, RRPs and Cessation) and smoking status groups (i.e., adult current smokers with and without intention to quit, adult former smokers, adult never smokers).

2 BACKGROUND

2.1 RRP as a Tool to Help Reduce Tobacco-Related Mortality and Morbidity

Although smoking prevalence has declined in many countries over the last decades, millions of adults continue to smoke cigarettes and due to population growth, the number of smokers is likely to further increase (Ng 2014). Recognizing this, the policy of tobacco harm reduction is being put forward by a multitude of stakeholders – including public health organizations, healthcare professionals and regulators – to complement the other major strategies often based on tobacco control, for reducing smoking-related harm (Royal College of Physicians 2007, WHO 2009, Zeller 2009). With the advent development of a range of products employing unconventional technologies and with potential to reduce health risks compared to CCs (Pederson 2007), the question arises as to what their potential public health impact would be.

An important factor in evaluating the full potential of any RRP on public health, is to understand the perceptions and future behavior towards the product among both current tobacco users and non-users. This is an important consideration for regulatory authorities such as the Food Drug Administration (FDA) who highlighted the issue in their draft guidance for Modified Risk Tobacco Products (MRTP).

The challenge is to develop tobacco products that reduce health risk and are acceptable to adult smokers as substitutes for CC. In addition, it is important that these products generate low intention to use among adult former smokers and adult never smokers, and do not impact the intention to quit of those adult smokers who are motivated to quit.

Communicating the substantiated risk profile of a RRP compared to CC is an important element in gaining acceptance of the product by adult smokers. At the same time, it is essential that such communications are correctly understood and do not encourage initiation or discourage cessation.

Philip Morris International (PMI) aimed to meet these objectives can be met by a potential RRP known as the Tobacco Heating System (THS). This product contains tobacco and replicates much of the ritual of smoking, but without combustion of tobacco. THS heats tobacco to temperatures lower than CC (below 400°C), thereby significantly

reducing or eliminating harmful and potentially harmful constituents (HPHCs) compared to CC smoke.

2.2 RRP Research on Consumer Perception and Behavior

To measure the potential benefit of marketing an RRP to the public, the FDA Draft Guidance recommends investigating several areas, including (1) the effect on tobacco use behavior among current tobacco users, (2) the effect on tobacco use initiation among non-users, and (3) the effect of marketing on consumer understanding and perceptions. In response to the areas of investigation, PMI has undertaken a comprehensive Perception and Behavior Assessment (PBA) program to study consumer perception and behavior toward potential RRPs.

2.3 Developing Self-Report Instruments for RRP Research

To address consumer perception and behavior prior to market launch, self-report instruments are needed. As recognized by Rees et al. (Rees 2009), current measures to assess consumer responses to RRPs lack evidence of validity and reliability which are needed to support future regulation of RRPs. To fill this gap and adequately address the FDA MRTP Draft Guidance recommendations, the Scales Development and Validation Project (PBA01) was undertaken to develop scientifically validated measurement instruments that assess two concepts: (1) the *Perceived Risks* that are associated with the use of tobacco and nicotine-containing products and (2) the *Intent to Use* the tobacco product in the future (including *Intention to Try* and *Intention to Use*). The project consisted of the typical qualitative and quantitative phases associated with the development of self-reported measurement instruments following the current best practice guidelines (Aaronson 2002); (Mokkink 2010), including the US FDA scientific requirements for Patient-Reported Outcomes (FDA 2009).

3 METHODOLOGICAL APPROACH

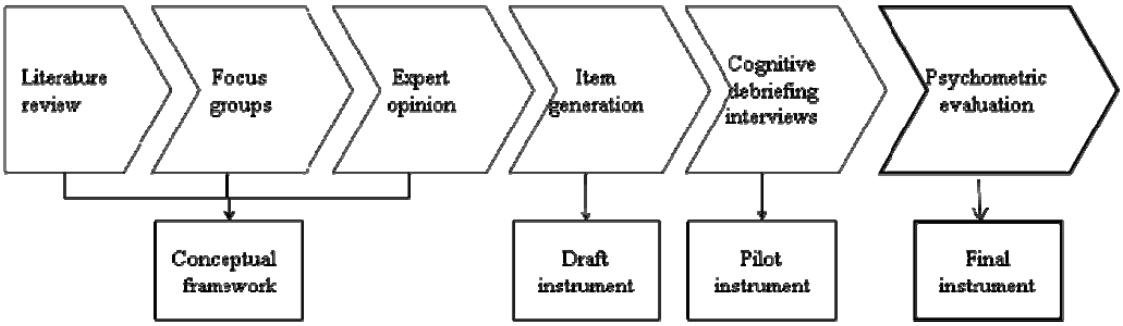
As summarized in Abbreviations: PRI=Perception Risk Instrument; ITUQ=Intent to Use Questionnaire

Note: All but the last step in this figure defines the qualitative development of the PRI and ITUQ. The last step refers to the quantitative evaluation and final validation of these two instruments, consisting of a small pilot study followed by two quantitative cross-sectional studies.

Figure 1, The FDA guidelines for the development of self-reported measurement instruments (FDA 2009) particularly stress the fundamental importance of conceptual frameworks and their definitions. These are best achieved using a mix of qualitative investigations to identify the dimensionality of the construct; generate potential items representing the construct; establish the most appropriate item phrasing, questionnaire structure, and context (i.e., target population and tobacco-related products assessed by the questionnaire); and conduct cognitive debriefing interviews to ensure clarity and consistency in the meaning of each element of the new instrument. To this end, and in the context of PBA01, the qualitative research included a literature review, consumer focus groups, discussion with experts, and cognitive debriefing interviews. An important additional objective of PBA01 lay in its international scope. Thus, studies revealing the

conceptual foundations of the instruments were carried out in multiple countries, namely Italy, Japan, the United Kingdom (UK), and the United States (US).

Following this qualitative research, quantitative research was undertaken. The draft instruments were first administered in a small pilot study to assess the feasibility of developing multiple scales in parallel. The scales were subsequently field-tested in large sample of participants with different smoking status in two subsequent quantitative cross-sectional studies.



Abbreviations: PRI=Perception Risk Instrument; ITUQ=Intent to Use Questionnaire

Note: All but the last step in this figure defines the qualitative development of the PRI and ITUQ. The last step refers to the quantitative evaluation and final validation of these two instruments, consisting of a small pilot study followed by two quantitative cross-sectional studies.

Figure 1. Schematic of the Process of Development of the PRI and ITUQ

4 CONTEXT OF USE OF THE SELF-REPORT INSTRUMENTS

The PRI and ITUQ were developed to be administrated to different populations of interest classified in four main smoking status groups: adult smokers with no intention to quit, adult smokers with the intention to quit, adult former smokers, and adult never smokers. Below are the operational definitions used in the context of the instrument development for each smoking status group in accordance with guidelines established by the World Health Organization (WHO 1998):

- **Adult Smokers with no Intention to Quit CC**:

  Current adult smokers having smoked at least 100 CC, are currently smoking at least one CC/day, and having no intention of quitting as defined by Prochaska and DiClemente's Stages of Change Model ("pre-contemplation" stage) (Prochaska 1982).

- **Adult Smokers with the Intention to Quit CC**:

  Current adult smokers having smoked at least 100 CC , are currently smoking at least one CC/day, and having the intention to quit CC as defined by Prochaska and DiClemente's Stages of Change Model ("contemplation" and "preparation" stages) (Prochaska 1982).

- **Adult Never Smokers:**

Adults having either never smoked at all, or never having been daily smokers and having smoked less than 100 CC.

- **Adult Former Smokers:**

Adults having formerly been daily smokers (i.e., having smoked at least 100 CC) and, at the time of study, having quit smoking CC more than 30 days ago. Applicability of the Self-Report Instruments to Different Tobacco and Nicotine-Containing Products

For the development of the instruments, the PRI and the ITUQ were applied to different products products that contain tobacco (e.g. conventional cigarettes [CCs], Tobacco Heating System [THS], a product that heats instead of burning tobacco), as well as products that contain nicotine but no tobacco (NRTs, electronic cigarettes [E-CIG]).). In addition, the PRI was also administered to "Cessation", which is defined as having successfully stopped smoking CC and not using any nicotine-containing product.

5 CONCEPTS OF INTEREST

5.1 Perceived Risks

The Institute of Medicine (IOM) in its report, Scientific Standards for studies on MRTPs (IOM 2012), recommends the measure of perception of tobacco-related outcomes to include the perception of short- and long-term risks and addiction. Perceptions of general harm, such as overall risk of harm or addiction, as well as perceptions of specific harm, such as risk of lung cancer or heart disease should also be included. The IOM also stipulates that the measure of risk perception should allow for comparison of an MRTP with existing tobacco products and comparison among different smoking status groups. In response to the IOM recommendations, the concept of interest measured by the PRI is focused around health-related outcomes that are expected to be experienced with different tobacco and nicotine-containing products.

In addition, the PRI was developed to assess two sets of perceived risks: 1) to the ***individual*** respondent (personal risk; PRI-P); and 2) to users of the product in ***general*** (general risk; PRI-G). The PRI-P and the PRI-G share the same items and response options; only the opening stems differ. Specifically, for the PRI-P, the opening stem is adapted for each smoking status in order to take into account the personal smoking history of the respondent (e.g., instructions for a never smoker assessing the risks of smoking cigarettes "If you were to start smoking, what do you think would be the risk, if any, to you personally of getting the following (sometime during your lifetime) because you smoke cigarettes"). For the PRI-G, the opening stem is identical to all smoking status groups (e.g., "In general, what do you think is the risk, if any, to smokers of getting the following (sometime during their lifetime) because of smoking cigarettes").

5.2 Intent to Use

As recommended by the IOM (IOM 2012), several types of intentions to engage in MRTP should be studied, including intention to try the MRTP, intention to use the MRTP to quit smoking, intention to use the MRTP while continuing conventional tobacco products, and how soon one would start using the MRTP. Based on these considerations, the concept of interest measured by the ITUQ is focused around two main domains:

- Intention to try (at least once)
- Intention to use (on a regular basis, including use in conjunction or replacement of other products and time to start using).

6 QUALITATIVE PHASE OF SCALE DEVELOPMENT AND VALIDATION

6.1 Development of the Conceptual Framework

The conceptual frameworks underlying the proposed instruments measuring perceived risks and intent to use were defined on the basis of a literature review, input from participants in focus groups, and expert opinions.

6.1.1 Literature Review

The literature review was conducted in December 2012. The literature search covered publication dates from January 2000 through September 2012 in Embase® and Medline® (see Appendix A for the details on the search strategy). In addition, three experts in public health and quality of life, consumer risk perception, and scale development were consulted to identify additional literature not covered by our search (e.g., articles published before 2000).

Perceived Risk. For Perceived Risks, a total of 172 abstracts were screened; 42 relevant papers (39 papers focusing on risk perception) were selected for further review (see Appendix B for the literature search Workflow). The first part of this review focused on existing self-report instruments; the second part on key themes extracted from observational studies. All instruments were reviewed with considerations for appropriateness to use in target populations (e.g., smokers and nonsmokers, users of specific tobacco-related products), content coverage (e.g., domains, type of items, response options), and psychometric properties (e.g., validity and reliability, Table 1). This review did not identify any of the existing instruments as fit for purpose, given their content and psychometric characteristics. For perceived risks, the literature review identified four broad domains: health risk to the individual user of tobacco and nicotine-containing products, social risks, financial risks, and time-related aspects of perceived risk (Table 2).

The most widely addressed aspect of risk was health risk to the individual user of tobacco-products (all 21 papers referenced in Table 2), typically cigarette smokers. In this context, risks relating to cancer and heart disease (16 and 8 papers, respectively), as well as addiction (7 papers) were most often referenced. In addition, but much less frequently, the literature review disclosed social, financial, and time-related aspects of perceived risk (8, 2, and 2 papers, respectively).

Beside the identification of domains, the literature review revealed two important insights with regard to health risks. First, short-term risks, i.e., consequences that are likely to set in early, can be considered distinct from long-term risks that become manifest only at a later stage (e.g., (Slovic 2000)). Second, the perceived risk for the individual user of the product (i.e., "what does it mean for me personally") was distinct from the perceived risk for users in general (i.e., "what does it mean for users in general") (Weinstein 2005).

Intent to Use. The literature review showed that questions related to intent to use mainly focused on single items looking at smokers' and former smokers' perceptions of tobacco products; these items were geared toward understanding whether participants would use a particular tobacco product. Thus, for intent to use, the findings from the literature review supported that the concept of 'intent to use' may be captured using single item(s) rather than a psychometric scale. Given this finding, it was important to measure Intent to Use concepts aligned with the IOM (IOM 2012), which refers to several types of intentions to engage in MRTP, including intention to try the MRTP, intention to use the MRTP to quit smoking, intention to use the MRTP while continuing conventional tobacco products, and how soon one would start using the MRTP.

Table 1. Characteristics of Existing Scales to Measure *Risk Perception* of Tobacco Products

Scale	Domains (no. of items)	Response Scale	Population used for the Psychometric Validation	Development and Validation Criteria
The Attitudes and Beliefs about the Consequences of Smoking Scale (ABS Smoking Scale) (Budd 2001)	• Emotional benefits (7) • Health hazards (9) • Self-confidence (8) • Body image (2)	Five point Likert Format (1=strongly agree, 2=agree, 3=neither agree nor disagree, 4=disagree, and 5=strongly disagree)	178 US undergraduate students (79.8% females; 19.1% current smokers)	Item Generation • Literature • Expert opinion Item Reduction • Expert opinion • Factor analysis Psychometric Analyses • Internal reliability • Criterion validity
The Multidimensional Smoking Behavior Questionnaire (Gilliard 2001)	• Dependence (7) • Social integration (7) • Regulation of negative affect (7) • Hedonism (7)	Not indicated	Sample of 150 adult French smokers (75 men, 75 women) aged 18 to 70 years	Item Generation • Literature • Subject input Item Reduction • Factor analysis Psychometric Analyses • Internal reliability • Criterion validity

(table continues)

Scale	Domains (no. of items)	Response Scale	Population used for the Psychometric Validation	Development and Validation Criteria
The Tobacco Craving Questionnaire (TCQ) (Heishman 2003)	• Emotionality (5) • Expectancy (3) • Compulsivity (5) • Purposefulness (4)	Seven-point Likert format (ranging from strongly disagree to strongly agree)	Sample of 213 American current cigarette smokers not attempting to reduce or quit smoking (99 women, 114 men)	Item Generation • Literature • Subject input Item Reduction • Factor analysis Psychometric Analyses • Content validity • Convergent validity • Internal reliability
The Smoking Consequences Questionnaire Adult version (SCQ-A) (Jeffries 2004)	**SCQ-A** • Negative affect reduction (4) • Social facilitation (4) • Taste-sensorimotor manipulation (3) • Negative physical feelings (3) • Weight control (3) • Health risk (4) • Stimulation-state enhancement (3) • Negative social impression (3) • Boredom reduction (3)	Ten-point Likert format (1=not likely at all; 10=extremely likely)	484 Adult African American smokers	Item Generation • Items from the SCQ Item Reduction • Factor analysis Psychometric Analyses • Internal reliability • Predictive validity

(table continues)

Scale	Domains (no. of items)	Response Scale	Population used for the Psychometric Validation	Development and Validation Criteria
Adolescent Smoking Consequences Questionnaire (ASCQ) (Lewis-Esquerre 2005)	**ASCQ** • Negative affect reduction (8 items) • Taste-sensorimotor manipulation (2 items) • Social facilitation (8 items) • Weight control (5 items) • Negative physical feelings (5 items) • Boredom reduction (2 items) • Negative social impression (3 items)	Five-point Likert format (never, rarely, sometimes, often, always)	437 inexperienced and experienced adolescents (11-19 years old)	Item Generation • Adaptation of SCQ items • Literature • Subject input • Expert opinion Item Reduction • Factor analysis Psychometric Analyses • Test-retest reliability • Internal reliability • Concurrent validity
The Risk Perception Questionnaire (Park 2009)	Perceived personal risk (4 questions on the likelihood and danger of developing lung cancer and other smoking-related diseases (SRDs) Perceived comparative risk (6 questions)	Five-point Likert scales about the likelihood (very unlikely to very likely) and danger (strongly disagree to strongly agree) of developing lung cancer and a SRD	630 US adult participants enrolled in the National Lung Screening Trial NLST (345 current and 285 former smokers)	Item Generation • Literature • CDIs Psychometric Analyses • Internal reliability • Know-group validity

Abbreviations : SRD=smoking-related disease ; CDI=Cognitive debriefing interviews

Table 2. Preliminary Conceptual Model of Risk Perception based on Literature Review

Domain	Theme	Selected papers referencing this theme
HEALTH		
	Accelerates aging	(Budd 2001)
	Addiction	(Budd 2001, Kropp 2004, Tilleczek 2006), (Halpern-Felsher 2004, Lyna 2002, Rindfleisch 1999)
	Bad colds	Halpern-Felsher 2004, (Kropp 2004, Morrell 2010)
	Cancer	(Budd 2001, Halpern-Felsher 2004, Hamilton 2004, Jeffries 2004, Kropp 2004, Lyna 2002, Morrell 2010, Oncken 2005, Parascandola 2009, Park 2009, Peretti-Watel 2007, Rindfleisch 1999, Song 2009, Weinstein 2005)
	Causes cough	(Gilliard 2001, Kropp 2004, Lewis-Esquerre 2005, Morrell 2010, Song 2009)
	Dependence	(Gilliard 2001)
	Emphysema	(Oncken 2005, Rindfleisch 1999)
	Heart disease, heart attack	(Budd 2001, Kropp 2004, Morrell 2010, Oncken 2005, Rindfleisch 1999, Tilleczek 2006)
	Irritates mouth / throat	(Jeffries 2004)
	Make lungs hurt	(Lewis-Esquerre 2005)

(table continues)

Domain	Theme	Selected papers referencing this theme
	Mortality / Premature mortality, die from a smoking-related disease	(Budd 2001, Kropp 2004)
	Physical side effects (e.g., weight gain, breathlessness)	Budd 2001
	Quality of life	(Borland 2004, Wilson 2009)
	Reduce physical fitness	(Rindfleisch 1999)
	Shortness of breath, trouble catching breath	(Budd 2001, Halpern-Felsher 2004, Kropp 2004, Morrell 2010)
	Smell like an ashtray	(Halpern-Felsher 2004, Kropp 2004, Morrell 2010, Song 2009)
	Smokers are sick more often	Budd 2001
	Smoking / other drug use	(Budd 2001, Rindfleisch 1999)
	Stroke	(Oncken 2005, Rindfleisch 1999)
	Wrinkles on face	(Halpern-Felsher 2004, Kropp 2004, Morrell 2010, Oncken 2005, Song 2009)
SOCIAL		
	Bad impression	Rindfleisch 1999, Jeffries 2004
	Damage reputation	Tilleczek 2006

(table continues)

Domain	Theme	Selected papers referencing this theme
	Friends don't like it / get upset	(Halpern-Felsher 2004, Lewis-Esquerre 2005)
	Get into trouble	(Halpern-Felsher 2004, Morrell 2010, Song 2009)
	Irritating / Annoying others	(Rindfleisch 1999)
	Less attractive	(Jeffries 2004, Lewis-Esquerre 2005)
	Look ridiculous	Jeffries 2004, Lewis-Esquerre 2005
	Negative family view	Tilleczek 2006
	Negative social impression, negative view from others	Jeffries 2004
	People think less of me if they see me smoking	(Jeffries 2004, Rindfleisch 1999)
FINANCIAL		
	Borrow money to buy cigarettes	(Rindfleisch 1999)
	Never having any extra money	(Rindfleisch 1999)
	Spending a lot of money on cigarettes	Tilleczek 2006
	Starting accidental fires	(Rindfleisch 1999)

(table continues)

Domain	Theme	Selected papers referencing this theme
	Wasting money that could used for something else	(Rindfleisch 1999)
TIME		
	Being late for class because of smoking	(Rindfleisch 1999)
	Losing studying time by stopping to have a cigarette	(Rindfleisch 1999, Tilleczek 2006)
	Wasting a large portion of the day smoking	(Rindfleisch 1999)
	Wasting a lot of time by having to go outside to smoke	(Rindfleisch 1999)

6.1.2 Focus Groups

Seventeen focus group discussions were conducted in the UK (N=9), Italy (N=4), and Japan (N=4) in December 2012 to better understand how individuals describe perceptions of the risks and intent to use tobacco products. Twelve additional focus groups were conducted in the US in August 2013 following ethical approval by the New England institutional review board (IRB). All focus group participants, recruited by market research agencies (Schmiedl Marktforschung [Italy and the UK], Market Research Asia [Japan], Delve Marketing Research [Philadelphia and Atlanta], and Plaza Research [Los Angeles]), were of legal age of smoking and provided informed consent to participate. A maximum variation sample was chosen to ensure that a broad spectrum of age, ethnicity, and smoking status were represented. Focus group participants were classified into adult smokers (with and without intention to quit), adult former smokers, and adult never smokers as described in Section 4. In line with the objective of developing instruments that would be applicable to a broad range of products, CC, e-cigarettes and an NRT product were presented to focus group participants in order to stimulate discussions accordingly. Focus groups were tape-recorded and transcribed. In the analysis, the principle of thematic saturation (Kerr 2010) was applied during the transcription and codification process.

Perceived Risk. Analysis of the focus group results identified 88 concepts for Perceived Risks that were grouped into three thematic clusters: health/addiction risks; societal/social risks; and material/financial risks (Table 3). Health risks, including addiction, were most often cited by participants across all products (40%). However, at a country level, participants from Japan did not emphasize health risks (33%) over societal/social (31%) and material/financial risks (36%). The findings from the Focus Group phase supported and elaborated upon the literature review findings in terms of the domains underlying the concept of perceived risks.

***Intent to Use*.** Insights on intent to use were very poor as the focus group participants provided mainly information on the situational contexts they would use tobacco products (e.g., participants reported on factors influencing the use of each type of product, such as stress, peer-group values, curiosity, wanting to quit other forms of tobacco use, having it with food and alcohol, where these factors varied by type of product), rather than on the concept of intent to use per se.

Table 3. Risk Perception Thematic Clusters across Focus Groups

Risk Perception Thematic Clusters	Broad Definition and Examples	Total Number of Occurrences				
		UK (n = 72)	Italy (n = 32)	Japan (n = 32)	US (n = 93)	Total (n = 229)
Health/Addiction Risks	Any reference to general or specific risks to healthy function or illness, including any physical harm to the body (e.g., cancer, cardiac and respiratory diseases, skin health), and addiction	263	99	69	329	760
Societal/Social Risks	Any adverse experience in social settings or reaction from people associated with smoking, including harm from or to others (e.g., secondhand smoke), smell and unfavorable social appearance/reaction	232	72	65	292	661
Material/Financial Risks	Including personal material risks, financial risks or other personal risks of product use, including property damage, expenses as well as uncertainty, unknown risks or that the product would not meet users expectations, product unavailability	133	55	77	188	453

6.1.3 Expert Opinions

On 19 February, 2013, an expert panel consisting of four opinion leaders (KOLs) and two academic consultants in scale development was convened in Zurich, Switzerland. The KOLs (see Appendix C for KOL curriculum vitaes and academic consultants) were subject matter experts in fields of nicotine addiction, motivational aspects of consumer perception, and other relevant area on approaches to measurement (e.g., public health, behavioral epidemiology, regulatory submissions). The purpose of this meeting was twofold. First, in an open elicitation phase, experts were invited to suggest relevant themes related to risk perception and intent to use. Second, the panel was asked to review and respond to themes and concepts uncovered in the literature review and the focus groups. The conclusions of the panel meeting widely supported the literature review and the focus groups for both Risk Perception and Intent to Use.

6.2 Resulting Conceptual Framework

6.2.1 Perceived Risks

Given the broad agreement of the findings from the literature review, the focus groups, and expert opinion, the following conceptual framework was developed for the *Perceived Risk Instrument* (PRI):

- *Perceived Health Risk:* The perceived negative risk (or impact) of product use to the user's physical health, going from minor immediate concrete manifestations of health risk (e.g., having poor gum health) to more serious long terms ones (e.g., having lung cancer);
- *Perceived Addiction Risk*: The perceived negative risk (or impact) that product use may have on the user's sense of being addicted to using the product;
- *Perceived Health Risk to Others:* The perceived negative risk (or impact) to the physical health of nonsmokers when being around during the product use;
- *Perceived Social Risk:* The perceived negative risk (or impact) that product use will affect interpersonal interactions adversely or how the user is perceived by others;
- *Perceived Practical risk:* The perceived negative risk (or impact) that product use may have on the user's time and finances.

6.2.2 Intent to Use

Based mainly on the interpretation of the FDA draft guidance on MRTP and the IOM report, the hypothesized domains for the concept of Intent to Use were derived as the followings:

- *Intention to Try:* Intention of using a tobacco product a few times without commitment to keep on using it on a regular basis.
- *Intention to Use:* Intention to engage in tobacco usage on a regular basis (including use in conjunction or replacement of other products and time to start using).

6.3 Draft Instruments

6.3.1 Item Generation

Item generation involved developing an exhaustive list of potential items for each domain within the conceptual framework for *Perceived Risks* and *Intent to Use*, respectively. This was done through several workshops involving the project team and expert panels between February 20th and April 12th 2013.

Items were developed using information generated primarilary from focus groups with consumers, who represented current adult smokers and adult non-smokers of CCs. We also examined existing published measures and added items not discussed by our focus groups. Finally, we had experts nominate items that were missing from their perspective. Items in the pool were initially grouped into the domains defined in the conceptual framework (i.e., health risk to self, health risk to others, addiction risk, practical risk, social risk for *Perceived Risks* and intention to try and intention to use for *Intent to Use*) based on their conceptual meaning to represent coherent constructs.

Risk Perception. For the new instrument of perceived risk, each item denotes a symptom, a disease, a condition, or, generally speaking, an adverse consequence of using a tobacco or nicotine-containing product. For item formulation, two versions of a sentence stem were generated and shared by all items within a domain; one referring to the personal risk to the individual respondent (e.g., with regard to cigarette smoking: "What do you think is the risk, if any, to you personally of getting the following (sometime during your lifetime) because you smoke cigarettes … "), and one referring to the risk to a user of a product in general ("In general, what do you think is the risk, if any, to smokers of getting the following (sometime during their lifetime) because of smoking cigarettes … "). A five-point fully verbalized rating scale ranging from "no risk" to "very high risk" was considered in order to offer the opportunity to express a medium level of perceived risk ("moderate risk"). Furthermore, the option "don't know" was added in order to not to enforce a response if respondents do not relate to some items and therefore lack a perception.

Intent to Use. The generation of items for ITUQ was based on the interpretation of the FDA draft guidance on MRTPs (FDA 2012) and the IOM report (IOM 2012) complemented with the input from experts. A four-point fully Likert rating scale was drafted ("very unlikely", "somewhat unlikely", "somewhat likely", "very likely") to be response options for a total of 18 draft items, including "Based on what you know about THS, how likely or unlikely are you to try THS" and "If you try THS and like it, how likely or unlikely are you to use THS regularly.".

6.3.2 Cognitive Debriefing Interviews (CDIs)

The CDIs were conducted to review the understanding of all aspects of the PRI and ITUQ instruments, including instructions, sentence stems, items, as well as the appropriateness of the response options before proceeding to field test the instruments. To ensure conceptual equivalence of the instruments in UK- and US-English, CDIs were conducted in both countries (April 2013 for the UK and November 2013 for the US). Forty UK participants and 48 US participants varying in smoking status (see Target

Population in Section 4) were asked to review the draft version of the instruments to determine their understanding of each item, to point out any unclear or ambiguous items, and to comment on the response options and recall periods. Overall, participants found the content to be comprehensive, sentence stems to be clear, and items and response formats straightforward to complete.

Risk Perception. Importantly, participants were able to discriminate between the two versions of the PRI and assess personal versus general risks accordingly. A few changes were made to the draft version of the draft PRI, including (1) the adjustment of stems for participants with different smoking status and for different product types; (2) the removal of two items in the health risk to others domain due to ambiguity and lack of relevance; and (3) improvements of the wording of some items. In addition, feedback from adult never smokers suggested that it was difficult to estimate their personal risk associated with products (specifically NRT) that they would never consider using. This led to the decision to not administer adult never smokers the PRI-P where the object assessed was either NRT or cessation.

Intent to Use. Adjustments to the ITUQ included (1) the addition of two extreme response options (i.e., "definitely" and "definitely not"); (2) the removal of 9 items evaluating respondents' intent to use an MRTP to focus on the key behaviors of concerns identified by the FDA and IOM as well as to reduce complexity; and (3) improvements of the wording of some items.

6.3.3 Final Draft Instruments Resulting from Entire Qualitative Phase

Upon completion of the CDIs, the final draft versions of PRI-P and PRI-G comprised a total of 67 items each, related to five domains: Perceived Health Risk to Self (31 items); Perceived Health Risk to Others (3); Perceived Addiction Risk (11); Perceived Social Risk (13); and Perceived Practical risk (9). The draft version of the ITUQ comprised a total of 9 items related to two domains: Intention to Try (2); Intention to Use (7).

7 QUANTITATIVE PHASE OF SCALE DEVELOPMENT AND VALIDATION

Three quantitative validation studies were conducted: a pilot study and two full-size US-based online cross-sectional studies. This work evaluated quantitatively the final draft instruments based on the qualitative research and development of the PRI and ITUQ.

7.1 Pilot Quantitative Study

Subsequent to the qualitative phase, the draft PRI and ITUQ were administered in a pilot quantitative study (N=233) in December 2013 with US participants to assess the feasibility of developing the multiple scales in parallel. A web-based data capture tool (i.e., Confirmit Horizons version 16) was used to gather responses from US participants from a proprietary database maintained by the web-survey provider Toluna Group Ltd. (Wilton, Connecticut USA). The preliminary psychometric assessment conducted on the pilot data led to the following conclusions:

- ***Risk Perception:*** Preliminary results indicated that the PRI's focus is to be restricted to the three health-related proposed scales (i.e., perceived health risk to

self, perceived addiction risk, perceived health risk to others), due to considerable floor effects (between 12 and 41%) that occurred for perceived social and practical risks when applied to products other than conventional cigarettes. As suggested by the literature review, perceived health and addiction risks are the most predominant risks reported by consumers across different tobacco and nicotine-containing products. This is also in line with IOM requirements on MRTP assessment (IOM 2012).

- ***Intent to Use.*** Pilot data analyses showed that a majority of the item thresholds were disordered, indicating that the response categories for these items are not working as intended, and therefore, the scoring function for that item is not valid. These pilot results supported further the outcomes of the qualitative phase that the ITUQ's items may be considered as standalone single categorical items.

7.2 Quantitative Studies I and II

7.2.1 Design and Procedure

Quantitative Study I and II were designed as cross-sectional, US population-based internet surveys with stratified sampling of four subpopulations defined according to self-reported smoking status at the time of data collection (see Target Population in Section 4). The studies utilized quota sampling based on the demographic characteristics age, gender, and education within each smoking status group. A web-based data capture tool (i.e., Confirmit Horizons version 16) was used to gather responses from study participants from a proprietary database maintained by Toluna Group Ltd. (Wilton, Connecticut USA), consisting of individuals with expressed interest to participate in online survey research. Quantitative Study I (administered between February and March 2014) and Quantitative Study II (administered between May and June 2014) were both approved by the New England Institutional Review Board and the participants received complete information about the study before agreeing to participate by signing an informed consent form (ICF). The total participation time for each survey was between 30 and 45 minutes and participants were rewarded with 3,500 points to exchange for vouchers or gifts at the reward partner network of the company hosting the survey (Toluna Group Ltd).

In Quantitative Study I, respondents completed the PRI for four "objects": CC, THS, nicotine patch as an example of NRTs, and Cessation. They also completed the ITUQ. In Quantitative Study II, e-cigarettes were included as a fifth object and nicotine patch was replaced by NRT as a general category in order to assess further the applicability of the instruments to various product categories. For each survey, participants were quota-randomized to pre-determine sequences so that an equal number of participants of each demographic stratum would be exposed to a specific sequence of objects. A minimum of 1600 completers, with an equal representation of each of the four subpopulations defined on smoking status (adult smokers with and without intention to quit, adult former smokers and adult never smokers), was estimated as an appropriate sample size for psychometric evaluation for each survey (Hobart 2012).

Only participants of legal smoking age agreeing with the ICF were enrolled in the survey. In total, 2411 and 2400 eligible participants were enrolled in Survey 1 and Survey 2, respectively.

7.2.2 Measurements

Each measure, the PRI and ITUQ was expected to function as a generic instrument that could be applied to tobacco and other nicotine-containing products.

Risk Perception. Three draft scales, in both versions of the PRI (i.e., PRI-P and PRI-G), were evaluated by different groups of participants: perceived health risk to self (31 items), perceived addiction risk (11 items) and perceived health risk to others (3 items).

Intent to Use. The ITUQ was designed as having two domains: 1) Intention to Try (sample at least once, 2 items), and 2) Intention to Use (continued usage, 7 items). The instrument was administrated to assess intent to use of conventional cigarettes (CC), a candidate MRTP (THS 2.2), and Nicotine patch (NRT).

Supportive Measures. The following supportive data were collected: (1) Tobacco use history was captured by the Smoking Questionnaire (Weitkunat 2013), addressing current and past use of tobacco-related products, (2) A summary measure of the participant's perceived short and long-term risk of smoking as assessed by the short-term and long-term consequences of Smoking Questionnaire (Slovic 2000), (3) a measure of risk acceptability (degree of a person's tolerance for risk; (Rindfleisch 1999), and (4) solely in Quantitative Study II, overall measures of the relative perceived risks associated to the five objects (i.e. CC, THS, e-cigarettes, NRTs, cessation) based on two 100 mm visual analog scales (VAS): one VAS for overall perceived health risk to self and one for overall perceived addiction risk. Demographic characteristics were also collected as part of the survey, including age, sex, education, income, and ethnicity.

7.2.3 Analyses

Risk Perception. For Quantitative Study I, the analyses aimed at identifying the items of the PRI with the best psychometric properties. Of note, perceived health risk to self (31 items) and perceived health risk to others (3 items) were initially combined to explore the potential of forming one inclusive 34-item scale. Rasch measurement theory (RMT) analysis included targeting, fit, reliability and stability across populations, products and type of risk based on differential item functioning (DIF) (see Appendix D for details on RMT-based analyses). The following CTT-based analyses were subsequently conducted: assessment of data quality; scaling assumptions; scale-to-sample targeting; internal consistency reliability; and assessment of dimensionality (see Appendix E for details on CTT-based analyses). For Quantitative Study II, the same analyses were replicated on the item-reduced scales obtained from Quantitative Study I for the purpose of cross-validation with an independent sample.

Convergent validity was assessed by non-parametric correlations with individual items of related measures (i.e., VAS on overall Health risk and Addiction risk and the Short-term and Long-term Smoking Risks Questionnaire from (Slovic 2000)). PRI differences

between respondent groups that were expected to differ based on subject-matter considerations (known group validity) were assessed with t-tests. The group differences examined were: 1) perceived personal versus general risk among current smokers (with perceived personal risk expected to be lower); 2) current versus never smokers (with perceived risk of smoking expected to be lower for current smokers); and 3) between smokers with intention to quit and smokers with no intention to quit (with perceived risk of smoking in smokers intending to quit expected to be higher). With regards to the first two analyses, group differences were assessed for all current smokers independently of their intention to quit, and separately for current smokers with and without intention to quit. To explore the extent to which the PRI scores were influenced by the position of the object of assessment in the sequence, mean PRI scores were calculated by object and per sequence and smoking group for both PRI scales. Since the number of sequences was very large (120 possible sequences), the assessment of sequence effects was based on pairwise comparisons of objects using t-tests for independent samples.

Intent to Use. Similar to the PRI instrument, the ITUQ data were evaluated against best practices for test design comprising criteria of CTT and newer RMT methods to examine if the items would coalesce together to form a scale score for Intention to Try and Intention to Use and display acceptable psychometric characteristics. These CTT and RMT analyses were similar to those conducted for the PRI (see Appendix D and Appendix E). It was expected that, if a set of items proposed to form a scale failed to have adequate empirical psychometric scale properties, the items would then be treated as single items and would be reported as such.

RMT analyses were performed using RUMM2030 and all other analyses were performed with SPSS (version 21). All variables were described by mean and standard deviation (continuous variables) and absolute and relative frequencies (categorical variables). All statistical tests were conducted at a testwise alpha level of five percent.

7.2.4 Results

7.2.4.1 Risk Perception: Results of PRI

PRI Scale Formation and Item Reduction (Quantitative Study I)

The majority of the 34 items assessing Perceived Health Risk demonstrated no disordered thresholds, reasonable coverage of the item thresholds (88%) and good reliability as assessed by the person separation index (PSI) of 0.97 (Table 4). In total, 16 items were removed (9 based on item misfit, and 7 based on DIF). A re-analysis of the reduced 18-item Perceived Health Risk scale revealed that the scale performed appropriately: no disordered threshold, no DIF, coverage of 84% and a PSI of 0.97 (Table 4). Psychometric performance based on CTT methods was also strong: item-level missing data 3% to 7%; skewness of 0.05; Cronbach's alpha of 0.99; Corrected Item Correlations ranging from 0.89 to 0.93 (Table 4). The 11 items assessing Perceived Addiction Risk showed no disordered item thresholds, reasonable coverage of the item thresholds (80%) and good reliability PSI=0.94 (Table 4). Three items showed misfit and one item had DIF. A re-analysis of the reduced 7-item Perceived Addiction Risk scale revealed that the scale performed appropriately (Table 4). Among the seven items, three are applicable for all objects but not Cessation. One item (feeling anxiety when in a situation where people

smoke) was retained for administration only for cessation (4-item scale for Cessation and 6-item scale for all other tobacco and nicotine-containing prod-ucts). A re-analysis of the reduced 7-item Perceived Addiction Risk scale revealed that the scale performed appropriately: no disordered threshold, no DIF, coverage of 75% and a PSI of 0.93 (Table 4). Psychometric performance based on CTT methods was also strong: item-level missing data 2% to 5%; skewness of -0.41; Cronbach's alpha of 0.98; Corrected Item Correlations ranging from 0.90 to 0.93 (Table 5). For both the Perceived Health and Addiction Risk scales, the personal (PRI-P) vs general risk (PRI-G) versions performed equivalently from a psychometric point of view (i.e., no DIF).

PRI Psychometric Cross-Validation (Quantitative Study II)

The analysis of the 18-item Perceived Health Risk scale data revealed that the scale performed appropriately: no disordered threshold, no DIF, 87% of coverage and a PSI of 0.97 (Table 4). Psychometric perfromance based on CTT methods was also strong: skewness 0.02; Cronbach's alpha of 0.99; Corrected Item Correlations ranging from 0.88 to 0.92 (Table 5). The item-level missing data percentages were slightly above the acceptability criterion of 10% (between 12% and 14%). The 7-item Perceived Addiction Risk scale showed no disordered item thresholds, reasonable coverage of the item thresholds (78%) and good reliability with a PSI of 0.94 (Table 4). Psychometric perfromance based on CTT methods was also strong: skewness of -0.32; Cronbach's alpha of 0.98; Corrected Item Correlations ranging from 0.92 to 0.95 (Table 5). As for the Perceived Health Risk Scale, the item-level missing data percentages were slightly above the the acceptabilitiy criterion of 10% (between 8% and 13%).

Table 4. RMT Performance Summary for ITUQ and PRI in Quantitative Study I and II

Proposed Scale (# items)	% coverage Item threshold distribution	% items with fit residual> | 2.5 | [a]	% Items with $p(\chi^2)<0.05$ [b]	% Items with disordered thresholds	% pairs of item residual correlations>0.30	% items with p(DIF)<0.05 [b]	PSI
QSI Long Form Scales PRI							
Health Risk (34)	88	94	21	0	16/153	50	0.97
Addiction Risk (11)	80	82	18	0	3/21	9	0.94
QSI Reduced Scales PRI							
Health Risk (18)	84	61	0	0	0/153	0	0.97
Addiction Risk (7)	75	100	0	0	0/21	0	0.93
QSII Reduced Scales PRI							
Health Risk (18)	87	72	0	0	0/153	0	0.97
Addiction Risk (7)	78	86	0	0	0/21	0	0.94
QSI Core Items ITUQ [c]							
Intent to Use (4)	45	50	50	25	0/6	50	0.87

Abbreviations: $\chi 2$ = Chi-square; DIF = differential item functioning; PSI = person separation index; QSI = Quantitative Study I; QSII = Quantitative Study II.

[a] The high percentages were expected given the large sample size but are still informative when some items are much worse fitting relative to others

[b] In the statistical assessment the actual n was adjusted to 500 in order to mitigate excessive power and for parallel fit assessment based on a sample size of 500, which is deemed appropriate for the present psychometric analysis.

[c] Psychometric performance of the ITUQ was not *assessed* in Quantitative Study II.

Table 5. CTT Psychometric Performance Summary for PRI Health Risk and Addiction Risk Scales in Quantitative Study I and II

Proposed Scale (# items)	Range missing data (%)	Min-Max Sum score	Mean Sumscore (SD)	Range CITC	Ceiling/ Floor (%)	Skewness	Cronbach's alpha	Mean IIC	Range IIC
QSI									
Health Risk (18)	3-7	18-90	54.4 (22.32)	0.89-0.93	7/10	0.05	0.99	0.83	0.76-0.90
Addiction Risk (7)	2-5	6-30	20.7 (7.50)	0.90-0.93	8/20	-0.41	0.98	0.87	0.82-0.91
QSII									
Health Risk (18)	12-14	18-90	56.1 (20.46)	0.88-0.92	5/10	0.02	0.99	0.81	0.75-0.89
Addiction Risk (7)	8-13	6-30	20.6 (7.09)	0.92-0.95	6/18	-0.32	0.98	0.89	0.85-0.93

Abbreviations: CITC=corrected item-total correlation; IIC=inter-item correlation; SD=standard deviation; QSI = Quantitative Study I; QSII = Quantitative Study II.

PRI Construct Validity (Quantitative Study II)

For CC, THS 2.2, E-Cigarettes and NRT, all correlations of the VAS and PRI measures for both perceived health risk and perceived addiction risk were in the range of 0.52 to 0.68 across both types of risk (Table 6). Correlations of the 18-item Perceived Health Risk measure with all five items of the Short- and Long-Term Smoking Risks Questionnaire were all in the expected direction for both per-sonal and general risk (Table 7). Correlations were mostly weak to moderate, regardless of smoking status and type of risk, with absolute values ranging from 0.10 to 0.40 for personal risk and from 0.20 to 0.46 for general risk. Correlations were of similar magnitude across items focusing on short-term (first three items) or long-term consequences of smoking CC (last two items). This confirms that the 18-item Perceived Health Risk scale is balanced in terms of short- and long-term risks.

With respects to known group validity, all mean differences were in the expected direction. In terms of the effect sizes (Cohen's d), differences between adult smokers and adult never smokers were more pronounced than differences between personal and general risk among adult current smokers (Table 8). Regarding the differences between adult current smokers with and without intention to quit, known-group validity was confirmed as well by the perceived risk being higher for adult smokers with quitting intention.

Table 6. Convergent Validity of PRI Scales with VAS Scores (Survey 2)

Scale	CC r_s (n)	THS 2.2 r_s (n)	E-CIG r_s (n)	NRT r_s (n)
PRI-P vs. VAS Health Risk	0.58 (765)	0.65 (651)	0.65 (717)	0.54 (550)
PRI-P vs. VAS Addiction Risk	0.56 (767)	0.67 (704)	0.68 (708)	0.57 (534)
PRI-G vs. VAS Health Risk	0.52 (775)	0.61 (711)	0.62 (724)	0.52 (713)
PRI-G vs. VAS Addiction Risk	0.54 (771)	0.59 (702)	0.61 (714)	0.52 (704)

CC = conventional cigarettes; E-CIG = electronic cigarettes; n = number of study participants with both measurements; NRT = nicotine replacement therapy; PRI-P = Perceived Risk Instrument-Personal Risk; PRI-G = Perceived Risk Instrument-General Risk; r_s = *Spearman rank correlation coefficient; THS 2.2 = Tobacco Heating System 2.2; VAS = Visual Analog Scale.*

PRI Carry-over Effects (Quantitative Study II)

For CC, THS 2.2 and e-cigarettes no differences were detected between measures of Perceived Health Risk when the object was presented first versus second or later (Table 9). However, for cessation, both personal and general Perceived Health Risk were higher when cessation was presented as the first object compared to it being presented after at

least one other object. For NRT, a similar effect was found for general risk, with the level of risk being higher when NRT was assessed first. Thus, the PRI-P seemed to be less susceptible to carry-over effects than the PRI-G. Findings from this analysis suggest that these effects identified at the level of person measures may best be accommodated by a fixed order of objects presented to the participants. The best-known product should be presented first, setting a meaningful reference point for the participant. Thereafter, tobacco products should be presented in terms of decreasing familiarity. Based on the principle of moving from use of products to their non-use, objects related to quitting smoking should be presented last, with cessation (not involving any use of NRT) to be presented as the very last object.

Item Calibration of the PRI and Production of Scoring Tables

To enable appropriate use of the PRI, a calibrated scoring was developed. Based on the unweighted sum scores of participants' responses to items forming scales, linear participant measures, or person measures, were estimated for each scale and each object, if applicable. Specifically, the participant measure was based on the weighted likelihood estimation (WLE). Given the participant raw scores and item parameters, the calibration was done with the restricted Rasch model for polytomous responses[1] (also known as rating scale model). We adopted the restricted model as this provides a more parsimonious data representation, and allows for clearer demonstrations of value. Since no differential item functioning (DIF) by study sample or by type of risk was revealed, a common calibration for PRI-P and PRI-G was possible. In addition, pooled data from Quantitative Study I and II were pooled in order to gain precision in item calibration.

For complete data, the resulting conversion table transfers sum scores to RMT logit measures, which are mapped to a 0-100 scale for convenience and to aid interpretation and reporting of results, by moving the scale range origin (i.e., zero) to 50 and adding 50. For missing data, a standard imputation rule was applied: If there is less than 50% within-person item-level missing data, the imputation rule is to replace the items with missing responses with the within-person mean computed based on the sum of available within-person item responses divided by the number of items which have responses and to round the sum total score for 18-items to the nearest whole integer before using the conversion table to derive the corresponding person measure.

The final content of the PRI is documented in a user manual (1st Edition April 2016) which was designed to provide practical information on the PRI measurement development, context of use, administration, and scoring procedures for researchers

[1] The restricted model estimates a common set of threshold parameters with differences between any pair of thresholds being equal across all items in a scale. If it is conceptually justifiable to assume that the response scale functions identically for all items and if the threshold estimates in an unrestricted model are sufficiently similar as to infer that any differences reflect random variation, it is appropriate to apply the restricted model.

interested to use the PRI (Appendix F). The PRI user manual and questionnaire have been made available to industry and academic researchers through Mapi Research Trust (https://eprovide.mapi-trust.org/login?redirectTo=https://eprovide.mapi-trust.org/instruments/perceived-risk-instrument#basic_description).

Table 7. Convergent Validity of PRI 18-Item Health Risk Scale (CC) with Items from the Short- and Long-Term Smoking Risks Questionnaire (Spearman Correlation Coefficients, Survey 2)

	PRI-P Health Risk Scale					PRI-G Health Risk Scale				
Short and Long-Term Risk Questionnaire	All (n=773)	NS (n=184)	FS (n=192)	CS IQ (n=203)	CS NIQ (n=194)	All (n=778)	NS (n=192)	FS (n=196)	CS IQ (n=197)	CS NIQ (n=193)
Item 1	-0.35	-0.26	-0.40	-0.21	-0.21	-0.30	-0.29	-0.29	-0.20	-0.33
Item 2	0.33	0.34	0.28	0.24	0.35	0.39	0.26	0.45	0.31	0.45
Item 3	-0.28	-0.27	-0.34	-0.14	-0.14	-0.29	-0.26	-0.24	-0.23	-0.25
Item 4	-0.28	-0.30	-0.37	-0.10	-0.13	-0.28	-0.27	-0.29	-0.24	-0.23
Item 5	0.30	0.18	0.18	0.28	0.37	0.41	0.29	0.39	0.36	0.46

CS IQ = current smokers with intention to quit; CS NIQ = current smokers with no intention to quit; FS = former smokers; NS = never smokers; n = number of study participants with both measurements; PRI-P = Perceived Risk Instrument-Personal Risk; PRI-G = Perceived Risk Instrument-General Risk

Item 1: There is really no risk at all for the first two years.

Item 2: Every single cigarette smoked causes a little bit of harm.

Item 3: Although smoking may eventually harm this person's health, the very next single cigarette he or she smokes will probably not cause any harm.

Item 4: Harmful effects of smoking rarely occur until a person has smoked steadily for many years.

Item 5: Smoking at the daily rate of one package of cigarettes each day will eventually harm this person's health.

Table 8. Known-Group Validity: Comparison of Perceived Health Risk Score for CC between Different Groups (Survey 2)

	Smoking Status Group	n	Mean	SD	t (df)	*p*-value	Cohen's d
Differences between personal and general risk							
PRI-P	CS (all)	397	1.26	2.88	2.5 (785)	0.013	0.18
PRI-G	CS (all)	390	1.77	2.88			
PRI-P	CS NIQ	194	0.93	2.96	1.21 (385)	0.227	-
PRI-G	CS NIQ	193	1.29	2.93			
PRI-P	CS IQ	203	1.58	2.76	2.42 (398)	0.016	0.24
PRI-G	CS IQ	197	2.25	2.76			
Differences between current smokers and never smokers							
PRI-P	CS (all)	397	1.26	2.88	6.28 (579)	<.001	0.53
	NS	184	3.05	3.80			
	CS NIQ	194	0.93	2.96	6.08 (376)	<.001	0.62
	NS	184	3.05	3.80			
	CS IQ	203	1.58	2.76	4.39 (385)	<.001	0.44
	NS	184	3.05	3.80			
PRI-G	CS (all)	390	1.77	2.88	7.53 (580)	<.001	0.68
	NS	192	3.65	2.69			
	CS NIQ	193	1.29	2.93	8.22 (383)	<.001	0.84
	NS	192	3.65	2.69			
	CS IQ	197	2.25	2.76	5.06 (387)	<.001	0.51
	NS	192	3.65	2.69			
Differences between CS IQ and CS NIQ							
PRI-P	CS IQ	203	1.58	2.76	2.28 (395)	0.023	0.23
	CS NIQ	194	0.93	2.96			
PRI-G	CS IQ	197	2.25	2.76	3.33 (388)	0.001	0.34
	CS NIQ	193	1.29	2.93			

CS IQ = current smokers with intention to quit; CS NIQ = current smokers with no intention to quit; FS = former smokers; NS = never smokers; n = number of study participants with valid data; PRI-P = Perceived Risk Instrument-Personal Risk; PRI-G = Perceived Risk Instrument-General Risk; SD = standard deviation. Cohen's d indicated for p-values <0.05

Table 9. Assessment of Carry-Over Effects (Perceived Health Risk Scale RMT Logit Measure, Survey 2)

Sequence	n	Mean (logit)	SD	t (df)	*p*-value	Cohen's d
PRI-P						
CC first	159	2.08	2.98	0.18 (771)	0.860	-
CC subsequently	614	2.13	3.24			
THS 2.2 first	149	0.62	3.19	-0.45 (716)	0.650	-
THS 2.2 subsequently	569	0.48	3.17			
E-CIG first	142	-0.25	3.42	0.39 (724)	0.696	-
E-CIG subsequently	584	-0.12	3.34			
NRT first	110	-1.35	2.85	-0.42 (554)	0.672	-
NRT subsequently	446	-1.49	3.22			
CESS first	115	-0.05	2.52	-2.66 (584)	0.008	0.29
CESS subsequently	471	-0.84	2.91			
PRI-G						
CC first	162	2.89	2.75	-1.89 (776)	0.060	-
CC subsequently	616	2.41	2.91			
THS 2.2 first	149	0.50	2.97	0.62 (714)	0.537	-
THS 2.2 subsequently	567	0.66	2.97			

(table continues)

Sequence	n	Mean (logit)	SD	t (df)	*p*-value	Cohen's d
E-CIG first	143	-0.09	3.21	-0.35 (726)	0.723	-
E-CIG subsequently	585	-0.19	3.03			
NRT first	140	-0.21	2.85	-2.10 (716)	0.037	0.20
NRT subsequently	578	-0.82	3.17			
CESS first	156	0.95	2.76	-4.41 (765)	<0.001	0.40
CESS subsequently	611	-0.15	2.80			

CC = Conventional Cigarettes; CESS = Cessation; E-CIG = E-cigarettes; n = number of study participants with valid data; NRT = Nicotine Replacement Therapy; PRI-P = Perceived Risk Instrument-Personal Risk; PRI-G = Perceived Risk Instrument-General Risk; SD = Standard deviation; THS 2.2 = Tobacco Heating System 2.2. Cohen's d indicated for p-values <0.05

7.2.4.2 *Intent to Use: Results of ITUQ*

The pilot study indicated that the ITUQ items provided little evidence that they could form a scale. Therefore some adjustments were made on how the items were entered in the psychometric analysis: only 4 items were included in the analysis, the 2 items on Intention to Try and the 2 items on Intention to Use regularly in order to limit the number of structural missing data (while all participants were asked to complete the two items on Intention to Try implying complete data, only participants who expressed some intention to try (i.e., did not respond "definitely not" to the Intention to Try items) were asked to complete items capturing Intention to Use). When focusing on the four key items of the ITUQ that are common across the different products that were assessed (i.e. CC, THS, and NRT), results showed that these items had serious psychometric problems of one type or another when considering them part of a scale (Table 4). Of all the issues, the most serious issues were that two of the four items misfit (Item Chi-Square and F-statistics comparing expected and observed responses), and the remaining two showed significant DIF. Since all items in the ITUQ had issues which prevent a measure to be successfully constructed, no item-reduction strategy would result in an appropriate final instrument. Furthermore, because the ITUQ items did not converge into forming a scale, CTT analyses were not applicable. It was thus concluded that the items constituting the ITUQ are to be considered as standalone single categorical items, and for that reason, no psychometric validation was performed on the ITUQ in Survey 2. The final content of the ITUQ was documented in a user guide (version 1.0) designed to provide practical information on the ITUQ administration (Appendix G).

8 CONCLUSIONS

The qualitative and quantitative evaluations of the PRI confirmed the content validity and psychometric performance of the 18-item Perceived Health Risk and 7-item Perceived Addiction Risk scales. The two stand-alone items assessing Perceived Harm to Others, as part of the PRI, provide additional descriptive (without a scoring rule) information on the potential impact of tobacco and nicotine-containing products, as perceived by consumers. Both Personal vs General versions of the PRI performed equivalently from a psychometric point of view. These results provide empirical evidence to support that the two PRI scales measure their respective concepts of interest (i.e., perceived negative risk of product use to physical health, and perceived negative risk of being addicted to using a product, respectively) in the context of studies designed to compare risk perception of different types of product (i.e., conventional cigarettes, nicotine replacement therapy, RRPs), and provide a reference point for comparisons between studies and populations (i.e., adult smokers with the intention to quit, adult smokers without intention to quit, adult former smokers and adult never smokers). Subsequent to the PBA01 Scale Development and Validation Project, the PRI was implemented in a series of studies (i.e., THS-PBA-03-US, THS-PBA-05-RRC-US, THS-PBA-05-RRC2-US and THS-PBA-05-REC-US) conducted in the context of PMI's PBA program to study consumer perception and behavior towards potential RRPs. The PRI-P version (assessing personal perceived risks for the individual respondents) was prefered over the PRI-G version (assessing

perceived risks for the product user in general) for these particular studies as it was considered more relevant to assess the potential effect of THS marketing on the tobacco use behavior of consumer groups with various smoking history (current tobacco users and non-users including former and never users). Since it has recognized that current measures to assess risk perception lack evidence of validity and reliability (Rees 2009), the present development fills an important gap. Based on the structured development process following best practice guidelines (Aaronson 2002); (FDA 2009), (Mokkink 2010), and the amount of validation data, the PRI can be a valuable self-report instrument that provides a scientifically rigorous method to quantify the perceived risks of tobacco and nicotine-containing products and related behaviors. With increasing numbers of researchers incorporating the PRI into their studies, we envision a rapidly expanding knowledge-base, informing further interpretation of risk perception data. Such data will provide meaningful information on 1) the effects of risk perception on tobacco and nicotine-containing product use behavior among current tobacco users, 2) the effects on product use initiation among non-users, and 3) the effects of risk communication on consum-er understanding.

Even though qualitative research established content and face validities of the items constituting the ITUQ, the qualitative and quantitative evaluations of the potential instrument confirmed that *Intent to Use* concept does not behave as a functional scale. Given the psychometric problems for all items in the instrument, the items must be used individually. The items have to be intepreted as stand-alone global measures providing a profile of various aspects of product trial and use following the recommandations of the IOM report on Scientific standards for studies on modified risk tobacco products (IOM (Institute of Medicine) 2012). Because the measure did not yield an instrument with a scale and an associated scoring, items related to *Intent to Use* are henceforth combined into the ITUQ, allowing for the administration of all items or a selection depending on the study objectives.

9 REFERENCES

Aaronson N, Alonso J, Burnam A, Lohr KN, Patrick DL, Perrin E, Stein RE. Assessing health status and quality-of-life instruments: attributes and review criteria. Qual Life Res. 2002; 11(3):193-205.

Andrich D. An elaboration of Guttman scaling with Rasch models for measurement. Sociol Methodol. 1985; 15:33-80.

Andrich D. An index of person separation in latent trait theory, the traditional KR-20 index, and the Guttman scale response pattern. Education Research and Perspectives. 1982; 9:1:95-104.

Andrich D, de Jong JHAL, Sheridan BE. Applications of latent trait and latent class models in the social sciences. Rost J, Langeheine R, editors. 4. ed.: Waxmann Publishing Co.; 1997. Diagnostic opportunities with the Rasch model for ordered response categories. p. 59-70.

Andrich D, Hagquist C. Real and artificial differential item functioning. J Educ Behav Stat. 2012.

Borland R, Yong HH, King B, Cummings KM, Fong GT, Elton-Marshall T, Hammond D, McNeill A. Use of and beliefs about light cigarettes in four countries: findings from the International Tobacco Control Policy Evaluation Survey. Nicotine Tob Res. 2004; 6 Suppl 3:S311-21.

Budd G, Preston D. College student's attitudes and beliefs about the consequences of smoking: development and normative scores of a new scale. Journal of the American Academy of Nurse Practitioners. 2001; 13(9):421-7.

Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. 1951; 16(3):297-334.

DeVellis RF. Scale development: theory and applications. 26 USA: Sage Publications; 2003.

FDA. Guidance for Industry – Modified Risk Tobacco Product Applications (Draft Guidance) http://www.fda.gov/downloads/TobaccoProducts/GuidanceComplianceRegulatoryInformation/UCM297751.pdf2012.

FDA. Patient reported outcome measures: use in medical product development to support labelling claims 2009. Available from: http://www.fda.gov/downloads/Drugs/Guidances/UCM193282.pdf.

Gilliard J, Bruchon-Schweitzer M. Development and validation of a multidimensional smoking behaviour questionnaire. Psychol Rep. 2001; 89(3):499-509.

Guttman L. Some necessary conditions for common-factor analysis. Psychometrika. 1954; 19(2):149-61.

Halpern-Felsher BL, Biehl M, Kropp RY, Rubinstein ML. Perceived risks and benefits of smoking: differences among adolescents with different smoking experiences and intentions. Prev Med. 2004; 39(3):559-67.

Hamilton WL, Norton G, Ouellette TK, Rhodes WM, Kling R, Connolly GN. Smokers' responses to advertisements for regular and light cigarettes and potential reduced-exposure tobacco products. Nicotine Tob Res. 2004; 6 Suppl 3:S353-62.

Hays RD, Anderson R, Revicki D. Psychometric considerations in evaluating health-related quality of life measures. Qual Life Res. 1993; 2(6):441-9.

Hays RD, Hayashi T. Beyond internal consistency reliability: rationale and user's guide for Multitrait Analysis Program on the microcomputer. Behav Res Methods Instrum Comput. 1990; 22(2):167-75.

Heishman S, Singleton E, Moolchan E. Tobacco Craving Questionnaire: Reliability and validity of a new multifactorial instrument. Nicotine and Tobacco Research. 2003; 5(5):645-54.

Hobart J, Cano S, Warner T, Thompson A. What sample sizes for reliability and validity studies in neurology? Journal of Neurology. 2012; 259(12):2681–94.

IOM. Scientific standards for studies on modified risk tobacco products Washington DC: The National Academies Press; 2012.

IOM (Institute of Medicine). Scientific standards for studies on modified risk tobacco products. Washington, DC: The National Academies Press. 2012.

Jeffries S, Catley D, Okuyemi K, Nazir N, McCarter K, Grobe J, Ahluwalia J. Use of a Brief Smoking Consequences Questionnaire for Adults (SCQ-A) in African American Smokers. Psychology of Addictive Behaviors. 2004; 18(1):74-7.

Kerr C, Nixon A, Wild D. Assessing and demonstrating data saturation in qualitative inquiry supporting patient-reported outcomes research. Expert Rev Pharmacoecon Outcomes Res. 2010; 10(3):269-81.

Kropp RY, Halpern-Felsher BL. Adolescents' beliefs about the risks involved in smoking "light" cigarettes. Pediatrics. 2004; 114(4):e445-51.

Lewis-Esquerre J, Rodrigue J, Kahler C. Development and validation of an adolescent smoking consequences questionnaire. Nicotine and Tobacco Research. 2005; 7(1):81-90.

Likert R. A technique for the measurement of attitudes. Arch Psychol. 1932; 140:5-53.

Lyna P, McBride C, Samsa G, Pollak KI. Exploring the association between perceived risks of smoking and benefits to quitting: who does not see the link? Addict Behav. 2002; 27(2):293-307.

McHorney CA, Ware JE, Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. Med Care. 1993; 31(3):247-63.

McHorney CA, Ware JEJ, Lu JF, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. Medical Care. 1994; 32(1):40-66.

Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res. 2010; 19(4):539-49.

Morrell HE, Song AV, Halpern-Felsher BL. Predicting adolescent perceptions of the risks and benefits of cigarette smoking: a longitudinal investigation. Health Psychol. 2010; 29(6):610-7.

Ng M, Freeman MK, Fleming TD, Robinson M, Dwyer-Lindgren L, Thomson B, Wollum A, Sanman E, Wulf S, Lopez AD, Murray CJ, Gakidou E. Smoking prevalence and cigarette consumption in 187 countries, 1980-2012. JAMA. 2014; 311(2):183-92.

Oncken C, McKee S, Krishnan-Sarin S, O'Malley S, Mazure CM. Knowledge and perceived risk of smoking-related conditions: a survey of cigarette smokers. Prev Med. 2005; 40(6):779-84.

Parascandola M, Augustson E, O'Connell ME, Marcus S. Consumer awareness and attitudes related to new potential reduced-exposure tobacco product brands. Nicotine Tob Res. 2009; 11(7):886-95.

Park ER, Ostroff JS, Rakowski W, Gareen IF, Diefenbach MA, Feibelmann S, Rigotti NA. Risk perceptions among participants undergoing lung cancer screening: baseline results from the National Lung Screening Trial. Ann Behav Med. 2009; 37(3):268-79.

Pederson LL, Nelson DE. Literature review and summary of perceptions, attitudes, beliefs, and marketing of potentially reduced exposure products: communication implications. Nicotine Tob Res. 2007; 9(5):525-34.

Peretti-Watel P, Constance J, Guilbert P, Gautier A, Beck F, Moatti JP. Smoking too few cigarettes to be at risk? Smokers' perceptions of risk and risk denial, a French survey. Tob Control. 2007; 16(5):351-6.

Prochaska J, Di Clemente C. Transtheoretical therapy: Toward a more integrative model of change. Psychotherapy: Theory, Research and Practice. 1982; 19(3):276-88.

Rees V, Kreslake J, Cummings K. Assessing Consumer Responses to Potential Reduced-Exposure Tobacco Products: A Review of Tobacco Industry and Independent Research Methods. Cancer Epidemiology Biomarkers & Prevention. 2009; 18(12):3225-40.

Rindfleisch A, Crockett DX. Cigarette smoking and perceived risk : a multidimensional investigation. J Public Policy Mark. 1999; 18(2):159-71.

Royal College of Physicians. Harm reduction in nicotine addiction: helping people who can't quit. A report by the Tobacco Advisory Group of the Royal College of Physicians. London: RCP. 2007.

Slovic P. What does it mean to know a cumulative risk? Adolescents' perceptions of short-term and long-term consequences of smoking. J Behav Decis Mak. 2000; 13(2):259-66.

Song AV, Glantz SA, Halpern-Felsher BL. Perceptions of second-hand smoke risks predict future adolescent smoking initiation. J Adolesc Health. 2009; 45(6):618-25.

Tilleczek KC, Hine DW. The meaning of smoking as health and social risk in adolescence. J Adolesc. 2006; 29(2):273-87.

Ware JE, Harris WJ, Gandek B, Rogers BW, Reese PR. MAP-R for Windows: Multitrait / multi-item analysis program-revised users's guide version 1 Boston MA: Health Assessment Lab; 1997.

Weinstein N, Marcus S, Moser R. Smokers' unrealistic optimism about their risk. Tobacco control. 2005; 14:55–9.

Weitkunat R, Coggins CRE, Sponsiello-Wang Z, Kallischnigg G, Dempsey R. Assessment of cigarette smoking in epidemiologic studies. Contributions to Tobacco Research 2013; 25:638-48.

WHO. Guidelines for Controlling and Monitoring the Tobacco Epidemic. ed.: World Health Organization; 1998. Monitoring tobacco use. p. 76-101.

WHO. Monograph: advancing knowledge on regulating tobacco products. Geneva: World Health Organization. 2009.

WHOQOL Group. The World Health Organistion Quality of Life Assessment (WHOQOL): Development and general psychometric properties. Social science & medicine. 1998; 46(12):1569-85.

Wilson N, Borland R, Weerasekera D, Edwards R, Russell M. Smoker interest in lower harm alternatives to cigarettes: national survey data. Nicotine Tob Res. 2009; 11(12):1467-73.

Wright BD, Masters GN. Rating Scale Analysis: Pluribus Press; 1982.

Zeller M, Hatsukami D. The strategic dialogue on tobacco harm reduction: a vision and blueprint for action in the US. Tob Control. 2009; 18(4):324-32.

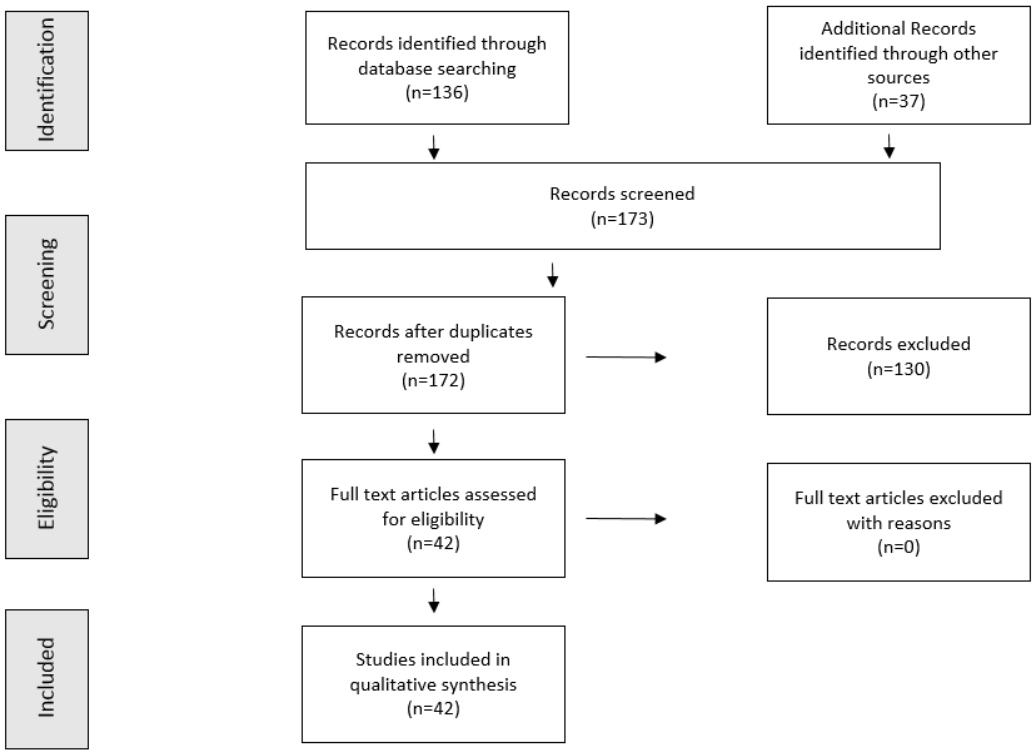
10 APPENDICES

Appendix A: Search Term used in the Literature Review

Topic	EMTREE Terms[a]	Title/Abstract Terms[a]
Tobacco-related	Cigarette smoke Cigarette smoking Smoking Tobacco Tobacco dependence Tobacco smoke	Cigarette(s) Nicotine Smokeless tobacco Smoking Tobacco
AND		
Survey or Questionnaire or Question or Item	Health surveys Item Question(s) Questionnaire(s)	Instrument Item(s)s Measure Question(s) Questionnaire Response(s) Scale Survey(s) Tool
AND		
Risk-Perception	Health risk Health risk and understanding Perceived risk Perception Risk Risk assessment Risk awareness Risk-benefit assessment Risk-perception	Attitude Perceived risk Perception of risk Risk Risk perception

[a] Disjunction (OR) at topic-level.

Appendix B: Four Phase Flow Chart of the Literature Review



Appendix C: CVs of the Key Opinion Leaders and Methodological Consultants

CV Stephen Crawford Key Opinion Leader

CV David Hodgins Key Opinion Leader

CV Adam Jaffe Key Opinion Leader

CV Don Saunders Key Opinion Leader

CV Stefan Cano Methodological consultant

CV Thomas Salzberger Methodological consultant

Appendix D: Rasch Measurement Theory Analyses: Properties, Definitions and Acceptability Criteria

Psychometric Property Definition	Evaluation(s) and Criteria for Acceptability
Targeting: refers to the extent to which the range of the target construct measured by each of the scales (i.e. perceived health risk and perceived addiction risk) matches the range of that target construct in the study sample. Better targeting equates to a greater ability to interpret the psychometric data with confidence (Wright 1982).	This involves examination of the relative distributions of the item locations and the person measurements and of the plot of the person-item location distributions, which shows the item locations and the person measurements on a common scale. There is no specific criterion. Essentially, the item locations should cover the sample adequately and the sample should cover the item locations adequately.
Fit: The items of each of the proposed instrument's scales must work together (fit) as a conformable set both conceptually and statistically. Otherwise, it is inappropriate to sum item responses to reach a total score and consider the total sale score as a measure of the target construct. When items do not work together (misfit) in this way, the validity of the scale is questionable (Wrigh 1982)	The following statistical and graphical indicators of fit were investigated (Andrich 1997): • Item discrimination: Fit residuals summarize the difference between observed and expected responses to an item across all people (item-person interaction). Fit residuals should ideally lie within ± 2.5. Fit residuals lying outside this range imply misfit of the observed data to the Rasch model. Negative values indicate overdiscriminating; positive values underdiscriminating items. Due to the large sample size in Surveys 1 and 2, it was to be expected to find a substantial number of item misfits, but this indicator was still considered helpful as some items were expected fitting much worse than others. • Item fit: Chi-squared values summarize the difference between observed and expected responses to an item for groups (or 'class intervals') of people with relatively similar levels of ability (item-trait interaction). A chi-squared value with a low likelihood (p-value) implies that the discrepancy between the observed responses and the expected value is large relative to chance for that item. • Item response ordering: This involves the examination of the category probability curves (CPCs) and the threshold probability curves (TPCs) which

Psychometric Property Definition	Evaluation(s) and Criteria for Acceptability
	show the ordering of the thresholds for each item. A threshold marks the location on the latent continuum where two adjacent response categories are equally likely. The ordering of the thresholds should reflect the intended order of the categories, i.e. ordered sequentially from less ('no risk') to more ('high risk'). Correct ordering supports the assumption that the response categories work as intended. Disordered thresholds indicate that the response categories for a particular item, are not working as intended, and therefore that the scoring function for that item is not valid. • Local independence: This involves an examination of item residual correlations. Correlations between the residuals should be low (<0.30). If residuals for item pairs are correlated >0.30, this indicates that the response to one item depends on the response to the other item, i.e., the items are locally dependent (Andrich 1985).
Precision and reliability: refer to the extent to which scale scores reflect random error (Andrich 1982)	This was assessed using the person separation index (PSI), which is an internal reliability statistic comparable to Cronbach's alpha. The PSI quantifies the error associated with the measurements of individuals in the sample (Andrich 1982). The PSI ranges from 0 (all error) to 1 (no error), with values closer to 1 indicating higher reliability. A low PSI implies that scale items are not able to reliably separating individuals on the scale they define.
The stability (and therefore comparability): of PRI measures across different factors will be based on tests of invariance (key criterion of successful measurement), implying that items mean the same to different participant groups under different conditions.	This is assessed by means of a test for differential item functioning (DIF) (Andrich 2012). Invariance was assessed according to demographic criteria (age, gender, education) as well as across different tobacco and nicotine-containing products, different subpopulations based on smoking status and across the application of the scales to personal risk and risk in general. DIF is assessed by comparing observed residuals (i.e., the difference between expected responses under the assumption of no DIF and actually observed responses) across groups of participants

Psychometric Property Definition	Evaluation(s) and Criteria for Acceptability
	defined by the DIF factor investigated (e.g., males versus females) and classified in several class intervals along the latent continuum measured by the scale.

Appendix E: Classical Test Theory Analyses: Properties, Definitions and Acceptability Criteria

Psychometric property Definition	Evaluation(s) and Criteria for Acceptability
Data quality: refers to the extent to which the scale items are accepted by the participants and, consequently, yield usable responses.	Missing data are indicative of a lack of acceptability and/or a lack of applicability of the items from the perspective of the participant. Item-level missing data should be <10% (WHOQOL Group 1998).
Scaling assumptions: refers to the extent to which it is legitimate to sum a set of item scores, without weighting or standardisation, to produce a single total score (DeVellis 2003, Hays 1990).	Summing scale item scores is considered legitimate, when the items: • Are approximately parallel (i.e. they measure at the same point on the scale). This criterion is satisfied when items have similar mean scores (Likert 1932). • Contribute similarly to the variation of the total score (i.e. they have similar variances), otherwise they should be standardized. This criterion is satisfied when items have similar standard deviations (McHorney 1994). • Measure a common underlying construct otherwise combining them to produce a single score is not appropriate (Guttman 1954). This criterion is satisfied when items have adequate corrected item-total correlation (ITC ≥ 0.30) (Ware 1997). • Contain a similar proportion of information concerning the construct being measured. Otherwise items should be given different weights (Likert 1932). This criterion is satisfied when items have similar ITCs (Ware 1997).
Scale-to-sample targeting: The extent to which the range of the construct measured by the scale matches the range of that variable in the study sample. Adequate targeting provides greater confidence in making judgments about the performance of the scale	Scale scores should span the entire range; floor (proportion of the sample at the minimum scale score for the scale) and ceiling (proportion of the sample at the maximum scale score) effects should be low (<15%) (McHorney 1993) and skewness should be between ± 1 (Hays 1993) There are no published criteria for item level targeting.

Psychometric property Definition	Evaluation(s) and Criteria for Acceptability
when interpreting results. Poor targeting implies that measurement precision is limited. People with extreme scores represent a sub-sample in which changes within and differences between individuals will be underestimated.	
Precision and reliability: are intimately linked and essentially around the issues of the extent to which scale scores are associated with random error. High reliability indicates that scores are associated with little random error, i.e. are consistent.	Internal consistency reliability estimates the random error associated with total scores from the intercorrelations among the items.(Cronbach 1951). The recommended level for adequate scale internal consistency is Cronbach's alpha coefficient ≥ 0.80 (Cronbach 1951) and item-total correlations > 0.30 (WHOQOL Group 1998).

Appendix F: PRI User Manual

PRI User Manual

Appendix G: ITUQ User Guide

ITUQ User Guide