



Identifying Reliable Labeling Comprehension Questions for E-Vapor and Heated Tobacco Products: A Meta-Analytic Approach

by

Jonathan M. Gallegos, Jennifer N. Lewis, and Elizabeth Becker

Altria Client Services LLC, Richmond, Virginia, USA

SUMMARY

Background: Labeling comprehension research can be tedious and resource-constraining because it often involves iterative reliability assessments of relevant metrics (e.g., formative testing of any safety or product-use related questions), leading to increased time requirements and shrinking sample pools in the population of interest. The goal of this work was to temper the need for formative metric evaluations in labeling comprehension research by identifying labeling comprehension questions that: 1) can be reliably answered correctly and 2) are generalizable across research on electronic nicotine delivery system (ENDS) products and heated tobacco products (HTPs). We explored whether, across four labeling comprehension studies for three distinct nicotine products (one ENDS, two HTPs), we could meta-analytically identify labeling comprehension questions that consistently demonstrated high correct response rates (relative to a 100% correct response rate).

Methods: Forty-five relevant questions were identified across studies and were categorized into 13 groups/domains for analysis purposes. Correct response rates for questions within a group/domain were analyzed using a proportional meta-analysis. A proportional meta-analysis is a data-synthesis method that allows for the calculation of a pooled overall proportion from multiple individual proportions (1).

Results: Evidence from six of the tested domains ('Know to charge device only with proprietary accessories,' 'Know to not expose device to high temperatures,' 'Know to not use device if it becomes wet or immersed in a liquid,' 'Know to not use device if it appears damaged or broken,' and 'Know not to use the device with other manufacturer's items') indicates that if the domain-specific question were asked again in a future study, a high proportion of participants would be expected to answer the question correctly. These six domains can serve as benchmarks for crafting questions that engender high correct response rates for the evaluated question domains. Five question domains ('Know to not use device if it is operating in a strange manner,' 'Know that the device contains nicotine,' 'Know how long the tobacco consumable will last,' 'Know not to use the device if the battery behaves unexpectedly,' and 'Know that the product contains tobacco') generally displayed high pooled proportions and narrow confidence intervals (but wide prediction intervals), indicating robust utility in crafting similar question for new studies, albeit context specific adjustments would help bolster the likelihood of high correct response proportions. Data for the two remaining domains ('Know to keep device away from children,' and 'Know how many times you can use the tobacco consumable') indicated wide variability in correct response proportions, thus questions from these domains would likely require significant context-based adaptation in future studies to engender high correct response proportions.

*Received: 22nd January 2026 – accepted: 28th May 2026

Conclusions: This work provides evidence that labeling research can be analytically leveraged to ease the burdens of new research; in this case, facilitating labeling comprehension research by meta-analytically demonstrating correct response consistency across varied question domains, thus providing design resources for researchers who are crafting labeling questions and user guides/safety guides. [Contrib. Tob. Nicotine Res. 35 (2026) 157–166]

KEYWORDS

Labeling comprehension; meta-analysis; e-vapor; heated tobacco

ZUSAMMENFASSUNG

Hintergrund: Die Forschung zum Verständnis von Produktkennzeichnungen kann mühsam und ressourcenintensiv sein, da sie häufig iterative Zuverlässigkeitsbewertungen relevanter Metriken erfordert (z.B. formative Tests zu Fragen bezüglich Sicherheit oder Produktnutzung), was zu einem erhöhten Zeitaufwand und schrumpfenden Stichprobenpools in der Zielpopulation führt. Das Ziel dieser Arbeit war es, den Bedarf an formativen Metrikbewertungen in der Forschung zum Verständnis von Produktkennzeichnungen zu verringern, indem Fragen zum Verständnis von Produktkennzeichnungen identifiziert wurden, die: 1) zuverlässig richtig beantwortet werden können und 2) auf die gesamte Forschung zu elektronischen Nikotinabgabesystemen (ENDS) und erhitzten Tabakprodukten (HTPs) übertragbar sind. Wir untersuchten, ob wir anhand von vier Studien zum Verständnis von Produktkennzeichnungen für drei unterschiedliche Nikotinprodukte (ein ENDS, zwei HTPs) meta-analytisch Fragen zum Verständnis von Produktkennzeichnungen identifizieren konnten, die durchweg hohe richtige Beantwortungsraten aufwiesen (im Verhältnis zu einer Rate von 100%). *Methoden:* In den Studien wurden 45 relevante Fragen identifiziert und für Analysezwecke in 13 Gruppen/Bereiche kategorisiert. Die richtige Beantwortungsrate für Fragen innerhalb einer Gruppe/Bereiche wurden mittels einer proportionalen Meta-Analyse analysiert. Eine proportionale Meta-Analyse ist eine Methode zur Datensynthese, die die Berechnung eines gepoolten Gesamtanteils aus mehreren Einzelanteilen ermöglicht (1).

Ergebnisse: Die Ergebnisse aus sechs der untersuchten Bereiche ('Wissen, dass das Gerät nur mit Originalzubehör aufgeladen werden darf', 'Wissen, dass das Gerät keinen hohen Temperaturen ausgesetzt werden darf', 'Wissen, dass das Gerät nicht verwendet werden darf, wenn es nass wird oder in eine Flüssigkeit eingetaucht wird', 'Wissen, dass das Gerät nicht verwendet werden darf, wenn es beschädigt oder defekt erscheint' und 'Wissen, dass das Gerät nicht mit Produkten anderer Hersteller verwendet werden darf') deuten darauf hin, dass bei einer erneuten Befragung zu den gruppenspezifischen Fragen in einer zukünftigen Studie zu erwarten wäre, dass ein hoher Anteil der Teilnehmenden die Frage richtig beantworten würde. Diese sechs Bereiche können als Maßstab für die Formulierung von Fragen dienen, die hohe richtige Beantwortungs-

raten für die bewerteten Fragegruppen erzielen. Fünf Fragegruppen ('Wissen, dass das Gerät nicht verwendet werden darf, wenn es sich ungewöhnlich verhält', 'Wissen, dass das Gerät Nikotin enthält', 'Wissen, wie lange das Tabak-Verbrauchsmaterial reicht', 'Wissen, dass das Gerät nicht verwendet werden darf, wenn sich der Akku unerwartet verhält' und 'Wissen, dass das Produkt Tabak enthält') wiesen im Allgemeinen hohe gepoolte Anteile und enge Konfidenzintervalle (jedoch breite Vorhersageintervalle) auf, was auf einen robusten Nutzen bei der Erstellung ähnlicher Fragen für neue Studien hindeutet, wenngleich kontextspezifische Anpassungen dazu beitragen würden, die Wahrscheinlichkeit hoher Anteile korrekter Antworten zu erhöhen. Die Daten für die beiden verbleibenden Bereiche ('Wissen, dass das Gerät von Kindern ferngehalten werden muss' und 'Wissen, wie oft das Tabak-Verbrauchsmaterial verwendet werden kann') zeigten eine große Variabilität bei den Anteilen korrekter Antworten; daher würden Fragen aus diesen Gruppen in zukünftigen Studien wahrscheinlich eine erhebliche kontextbezogene Anpassung erfordern, um hohe Anteile korrekter Antworten zu erzielen.

Schlussfolgerungen: Diese Arbeit liefert Belege dafür, dass die Forschung zu Produktkennzeichnungen analytisch genutzt werden kann, um den Aufwand für neue Forschungsvorhaben zu verringern; in diesem Fall wird die Forschung zum Verständnis von Produktkennzeichnungen erleichtert, indem meta-analytisch die Konsistenz der richtigen Antworten über verschiedene Fragegruppen hinweg aufgezeigt wird, wodurch Designressourcen für Forscher bereitgestellt werden, die Fragen zu Produktkennzeichnungen sowie Gebrauchsanweisungen und Sicherheitshinweise erstellen. [Contrib. Tob. Nicotine Res. 35 (2026) 157–166]

RESUME

Contexte: La recherche sur la compréhension des étiquettes peut s'avérer fastidieuse et exigeante en ressources, car elle implique souvent des évaluations itératives de la fiabilité des indicateurs pertinents (par exemple, des tests formatifs portant sur des questions liées à la sécurité ou à l'utilisation du produit), ce qui entraîne un allongement des délais et une réduction de la taille des échantillons au sein de la population étudiée. L'objectif de ce travail était de réduire le recours aux évaluations formatives des indicateurs dans la recherche sur la compréhension des étiquettes en identifiant des questions de compréhension des étiquettes qui: 1) peuvent recevoir une réponse correcte de manière fiable et 2) sont généralisables à l'ensemble des recherches sur les systèmes électroniques de délivrance de nicotine (ENDS) et les produits du tabac chauffés (HTP). Nous avons cherché à déterminer si, à travers quatre études sur la compréhension de l'étiquetage portant sur trois produits nicotiques distincts (un ENDS, deux HTP), nous pouvions identifier, par méta-analyse, des questions de compréhension de l'étiquetage présentant systématiquement des taux de réponses correctes élevés (par rapport à un taux de réponses correctes de 100 %).

Méthodes: Quarante-cinq questions pertinentes ont été identifiées dans l'ensemble des études et classées en 13

groupes/domaines à des fins d'analyse. Les taux de réponses correctes pour les questions au sein d'un groupe/domaine ont été analysés à l'aide d'une méta-analyse proportionnelle. Une méta-analyse proportionnelle est une méthode de synthèse des données qui permet de calculer une proportion globale combinée à partir de plusieurs proportions individuelles (1).

Résultats: Les données issues de six des domaines testés ('Savoir ne recharger l'appareil qu'avec des accessoires d'origine', 'Savoir ne pas exposer l'appareil à des températures élevées', 'Savoir ne pas utiliser l'appareil s'il est mouillé ou immergé dans un liquide', 'Savoir ne pas utiliser l'appareil s'il semble endommagé ou cassé' et 'Savoir ne pas utiliser l'appareil avec des produits d'autres fabricants') indiquent que si la question spécifique à chaque domaine était posée à nouveau dans une étude future, on pourrait s'attendre à ce qu'une forte proportion de participants y réponde correctement. Ces six domaines peuvent servir de références pour élaborer des questions suscitant des taux de réponses correctes élevés pour les domaines de questions évalués. Cinq domaines de questions ('Savoir qu'il ne faut pas utiliser l'appareil s'il fonctionne de manière inhabituelle', 'Savoir que l'appareil contient de la nicotine', 'Savoir combien de temps durera le consommable à base de tabac', 'Savoir qu'il ne faut pas utiliser l'appareil si la batterie se comporte de manière inattendue' et 'Savoir que le produit contient du tabac') ont généralement affiché des proportions globalisées élevées et des intervalles de confiance étroits (mais des intervalles de prédiction larges), ce qui indique une utilité solide pour l'élaboration de questions similaires dans le cadre de nouvelles études, même si des ajustements spécifiques au contexte contribueraient à renforcer la probabilité d'obtenir des proportions élevées de réponses correctes. Les données relatives aux deux domaines restants ('Savoir qu'il faut garder l'appareil hors de portée des enfants' et 'Savoir combien de fois on peut utiliser le consommable à base de tabac') ont révélé une grande variabilité dans les proportions de réponses correctes; par conséquent, les questions issues de ces domaines nécessiteraient probablement une adaptation contextuelle importante dans les études futures afin d'obtenir des proportions élevées de réponses correctes.

Conclusions: Ce travail démontre que la recherche sur l'étiquetage peut être exploitée de manière analytique pour alléger la charge de travail liée aux nouvelles recherches; dans le cas présent, il facilite la recherche sur la compréhension de l'étiquetage en démontrant, par une méta-analyse, la cohérence des réponses correctes à travers divers domaines de questions, fournissant ainsi des ressources de conception aux chercheurs qui élaborent des questions relatives à l'étiquetage ainsi que des guides d'utilisation et de sécurité. [Contrib. Tob. Nicotine Res. 35 (2026) 157–166]

INTRODUCTION

Labeling comprehension research is critical to the developmental life cycle of a product because it helps organizations: 1) assess whether prospective users of their products understand how a product functions and whether safety information related to product use is comprehended (e.g.,

not to use a device if its battery is leaking); and in some cases 2) ensure that products meet regulatory standards – for instance, market authorization of electronic nicotine delivery systems (ENDS) and heated tobacco products (HTPs) by the U.S. Food and Drug Administration (FDA) requires evidence demonstrating that “users (and consumers) are able to understand the labeling and instructions for use and use the product in accordance with those instructions” (2).

Labeling comprehension research typically involves giving prospective users paper or digital resources related to the product of interest (like user guides and/or safety guides) and asking a series of questions to assess comprehension of product use and safety information. The labeling comprehension assessment is often extensive, including many questions (e.g., upwards of 20) to comprehensibly evaluate whether a prospective user understands key safety and usability information related to product use. The methodology of labeling comprehension research can be tedious and resource-constraining because it often requires formative testing – an iterative process of multiple independent evaluations used to ensure that the relevant labeling comprehension interview questions are easily comprehended by participants, and that the questions consistently and accurately assess the intended metric (3). Stated differently, formative testing precedes the labeling comprehension study, and is needed to ensure that the required interview metrics are easily comprehended by the intended users and, critically, reliably assess an intended construct or domain (e.g., does the question ‘According to the materials, who should this product be kept away from’ successfully evaluate comprehension that a product should be kept away from children and pets?).

The iterative nature of these assessments increases the time required to conduct labeling comprehension research and reduces the needed sample pools in the population of interest (because respective samples for a labeling comprehension study and formative testing are drawn from the same population).

Extant literature offers little guidance on how to temper these methodological issues. Although WILLIS (3) offers robust guidance on how to conduct formative/cognitive testing, the resource constraints of metric-reliability assessments (i.e., formative testing) are minimally discussed, highlighting that “in one sense, a questionnaire could be tested forever” (p. 31), while stressing the value of a limited number of testing rounds. Similarly, there are varied examples of literature discussing guidelines and best practices for developing, but not evaluating, product instructions and safety information guides (4–7).

The goal of this work was to temper the need for formative metric evaluations in labeling comprehension research by identifying product-safety and product-instruction labeling comprehension questions that: 1) reliably demonstrate comprehension of target labeling information (i.e., are reliably answered correctly); and 2) are generalizable across research on ENDS and HTPs. We explored whether, across four labeling comprehension studies for three distinct nicotine products (one ENDS, two HTPs), we could meta-analytically identify labeling comprehension questions that consistently demonstrated high correct response rates (relative to a 100% correct response rate). We describe our meta-analytic approach below.

METHOD

Meta-analysis

A proportional meta-analysis is a data-synthesis method that allows for the calculation of a pooled overall proportion from multiple individual proportions (1). Proportion-based questions that assessed similar topics across four studies were identified and grouped into domains, or categories, for analysis purposes. Thus, each question category, or domain, included multiple questions that addressed a similar subject matter, and each question was sourced from one of four unique studies. One meta-analysis was performed for each question domain to assess the proportions of correct responses, across studies. See Table 1 for a list of question domains, along with the questions that each domain includes.

Study search strategy and inclusion criteria

A search was conducted within the regulatory study database of Altria Client Services LLC (ALCS) to identify studies that asked participants labeling comprehension questions. These are questions used to assess prospective product users' comprehension of a product's instructions for use and safety information. To be included in the current analysis, studies needed to meet the following two inclusion criteria: 1) studies needed to include labeling comprehension questions, with responses coded as correct or incorrect; and 2) studies needed to be classified as either formative – a study that evaluates the accuracy, clarity, and consistency of labeling comprehension questions – or summative – a study that evaluates comprehension of product instructions and safety information. Each study included in the current analysis was conducted by ALCS as a part of an independent Premarket Tobacco Product Application (PMTA) to fulfill regulatory requirements for ENDS or HTP candidate products.¹

Final study sample pool and labeling comprehension question selection

A total of six studies were identified in the ALCS database, however only four studies met inclusion criteria. Two formative (N = 11; N = 16) and two summative studies (N = 70; N = 201), across three distinct nicotine products (two HTPs and one ENDS), were included in the current research. Forty-five relevant questions were identified across these four studies and were categorized into 13 groups/domains for analysis purposes. Questions were grouped into domains based on conceptual equivalence of the underlying comprehension construct, not surface level wording (e.g., questions that examine comprehension regarding not exposing a product to high temperatures were grouped into a domain). Each question within a domain was sourced from a unique study, such that no two questions within a domain were sourced from the same study (Table 1).

¹ Formative and summative studies included in this analysis were independent from one another, thus each study was a unique and independent evaluation of labeling comprehension questions.

Outcome variable

Proportion of correct responses. The proportion of correct responses for a question, within a question domain, describes the number of participants within a study that answered the questions correctly.

Outcome statistics

Pooled proportion of correct responses.

A combined estimate of the overall proportion of correct responses from the individual correct response proportions from each study (1).

Prediction Interval: PI.

Describes the expected range in true effects (i.e., proportion of correct responses) in similar future studies (1).

Cochran's Q.

A test statistic that describes whether any variability that exists between proportions, across studies, is greater than that which is expected by chance (i.e., whether the proportions across studies are heterogeneous (8)).

Inconsistency Index: I^2 and Tau^2 .

I^2 describes the degree of variability that is attributable to real differences across proportions in a set of studies, rather than chance (1, 8). I^2 can be thought of as the percentage of total variation in proportions that is not attributable to chance. Tau^2 estimates the between-study variance – the spread of the true underlying proportions (1).

RESULTS

A proportional meta-analysis was performed to assess the proportion of correct responses for each question domain. Table 2 provides pooled proportions (including 95% confidence intervals), tests of heterogeneity (Cochran's Q), and metrics for quantifying heterogeneity (I^2 and Tau^2) for each question domain. See [Supplemental Material](#) (Appendix A and Appendix B) for an overview of study demographics and for forest plots of each meta-analysis.

The correct response rates for six of the 13 question domains (i.e., 'Know to charge device only with proprietary accessories,' 'Know to not expose device to high temperatures,' 'Know to not use device if it becomes wet or immersed in a liquid,' 'Know to not use device if it appears damaged or broken,' 'Know not to use the device with other manufacturer's items,' and 'Know that the device heats the tobacco') did not demonstrate statistically significant variability across studies (heterogeneity), indicating consistency in correct response proportions, across studies, for these six domains. Specifically, the pooled correct response proportions for these six question domains were $\geq 96\%$, with corresponding 95% confidence intervals ranging from 93% to 100%, an $I^2 \leq 44\%$, and a $Tau^2 \leq 0.004$ – indicating that the proportions of correct responses for questions within these six domains were generally high (near a 100% correct response rate) and exhibited moderate to no variability (heterogeneity). The prediction intervals ranged between 85% and 100% for these six question domains suggesting that if the questions from each domain were asked again, in a new study, correct response rates would be expected to fall between 85–100%.

Table 1. Labeling comprehension questions and respective user guide information by question domain.

Study	Question	Answer in user guide
<i>Know to charge device only with proprietary accessories</i>		
HTP #1 Formative study	Jonah lost the AC adapter and USB cable charger for his device, but he has another charger that he uses for his phone. According to the materials, what should Jonah do or not do?	Charge only with the dedicated and recommended USB cable and AC adapter. Charge only with the provided adapter and USB Cable
HTP #2 Formative study	Jonah lost the charger for his device, but he has another USB cable that he uses for his phone. According to the materials, what should Jonah do or not do?	Be sure to use the supplied USB type-C cable. CHARGE WITH INCLUDED CABLE
HTP #1 Summative study	Jonah lost the AC adapter and USB cable charger for his device, but he has another charger that he uses for his phone. According to the materials, what should Jonah do or not do?	Charge only with the dedicated and recommended USB cable and AC adapter. Charge only with the provided adapter and USB Cable
ENDS #1 Summative study	Dave's ENDS #1 battery needs charging, but he cannot find his ENDS #1 USB charger. According to the Guide, should Dave try to charge his ENDS #1 e-vapor device with any other charging device?	Do not attempt to charge the ENDS #1 battery with any other charging device.
<i>Know to not use device if it is operating in a strange manner</i>		
HTP #1 Formative study	Julio noticed that his device does not appear to be vibrating or pulsating correctly. According to the materials, what should Julio do or not do?	Do not use or charge your device or its accessories if any of them: - Appears BROKEN or DAMAGED - Operates in an erratic manner
HTP #2 Formative study	Julio noticed that the display was glitchy and there were flashing and changing icons. According to the materials, what should Julio do or not do?	To reduce the risk of injury, do not use your device or items if they: - Appear damaged or broken - Are operating in an erratic manner.
HTP #1 Summative study	Julio noticed that his device does not appear to be vibrating or pulsating correctly. According to the materials, what should Julio do or not do?	Do not use or charge your device or its accessories if any of them: - Appears BROKEN or DAMAGED - Operates in an erratic manner
ENDS #1 Summative study	Rose noticed that her ENDS #1 e-vapor device was operating erratically. According to the Guide, what should Rose do when she noticed that her ENDS #1 e-vapor device was operating erratically?	To reduce the risk of injury, do not use your ENDS #1 e-vapor device or items if they: - Appear broken or damaged - Are operating in an erratic manner
<i>Know to not expose device to high temperatures</i>		
HTP #1 Formative study	Omar left his device sitting in his car, with outside temperatures reaching over 100 degrees. According to the materials, what should Omar do or not do?	Do not use or charge your device or its accessories if any of them: - Has been exposed to EXCESSIVE HEAT DO NOT leave or charge the device near heat sources, such as: - Radiators or stoves - Direct sunlight - Inside a vehicle on a hot day
HTP #2 Formative study	Omar left his device sitting on top of a hot radiator while charging it. What could potentially happen?	Do not use, store or charge any items in or near heat sources such as radiators, stoves, hot appliances, or fires. High or low temperatures may cause battery fluid leakage, thermal burns, battery explosion or fire.
HTP #1 Summative study	Omar has left his device sitting in his car, with outside temperatures reaching over 100 degrees. According to the materials, what should Omar do or not do?	Do not use or charge your device or its accessories if any of them: - Has been exposed to EXCESSIVE HEAT DO NOT leave or charge the device near heat sources, such as: - Radiators or stoves - Direct sunlight - Inside a vehicle on a hot day.
ENDS #1 Summative study	Tim realized that he had left his ENDS #1 e-vapor device on top of a hot radiator, and it felt unusually hot to the touch when he picked it up. According to the Guide, what should Tim do if device felt unusually hot to the touch?	To reduce the risk of injury, do not use your ENDS #1 e-vapor device or items if they: - Have been exposed to excessive heat - Are unusually hot to the touch.

Table 1. Labeling comprehension questions and respective user guide information by question domain. (Continued)

Study	Question	Answer in user guide
<i>Know to not use device if it becomes wet or immersed in a liquid.</i>		
HTP #1 Formative study	Amaya dropped her device in a sink full of water. According to the materials, what should Amaya do or not do?	Do not use or charge your device or its accessories if any of them: - Becomes wet or has been immersed in any liquid.
HTP #2 Formative study	Amaya dropped her device in a sink full of water. According to the materials, what should Amaya do or not do?	To reduce the risk of injury do not use your device or items if they: - Become wet or immersed in any liquid.
HTP #1 Summative study	Amaya dropped her device in a sink full of water. According to the materials, what should Amaya do or not do?	Do not use or charge your device or its accessories if any of them: - Becomes wet or has been immersed in any liquid.
ENDS #1 Summative study	Liz's ENDS #1 e-vapor device was immersed in water when she dropped it in the sink while washing dishes. According to the Guide, what should Liz do if ENDS #1 e-vapor device was immersed in water?	To reduce the risk of injury, do not use your ENDS #1 e-vapor device or items if they: - Become wet or are immersed in any liquid
<i>Know to not use device if it appears damaged or broken</i>		
HTP #1 Formative study	Lee noticed that the sliding door at the top of his device is broken. According to the materials, what should Lee do or not do?	Do not use or charge your device or its accessories if any of them: - Appears BROKEN or DAMAGED - Operates in an erratic manner
HTP #2 Formative study	Lee dropped his device, and it's no longer working properly. According to the materials, what should Lee do or not do?	To reduce the risk of injury, do not use your device or items if they: - Appear damaged or broken
HTP #1 Summative study	Lee noticed that the sliding door at the top of his device is broken. According to the materials, what should Lee do or not do?	Do not use or charge your device or its accessories if any of them: - Appears BROKEN or DAMAGED - Operates in an erratic manner
ENDS #1 Summative study	John dropped his ENDS #1 e-vapor device and it broke. According to the Guide, what should John do to reduce the risk of injury after it broke?	To reduce the risk of injury, do not use your ENDS #1 e-vapor device or items if they: - Appear damaged or broken.
<i>Know to keep device away from children</i>		
HTP #1 Formative study	According to the materials, who should this product be kept away from?	KEEP THE DEVICE, ITS ACCESSORIES AND THE TOBACCO STICKS OUT OF REACH OF CHILDREN AND PETS
HTP #2 Formative study	According to the materials, who should this product be kept away from?	KEEP OUT OF REACH (device, accessories, and tobacco sticks) OF CHILDREN AND PETS To reduce the risk of choking or poisoning: - Keep the device and items out of the reach of children.
HTP #1 Summative study	According to the materials, who should this product be kept away from?	KEEP THE DEVICE, ITS ACCESSORIES AND THE TOBACCO STICKS OUT OF REACH OF CHILDREN AND PETS
ENDS #1 Summative study	According to the Guide, should Sam leave his spare cartridges on his short table when his young children are in the room?	KEEP OUT OF REACH (device, accessories, and tobacco sticks) OF CHILDREN AND PETS To reduce the risk of choking or poisoning: - Keep the ENDS #1 e-vapor device and items out of the reach of children. Do not leave ENDS #1 items unattended where children or pets could gain access to them.
<i>Know that the device contains nicotine</i>		
HTP #1 Formative study	According to the materials, what addictive chemical does this product contain?	The compatible tobacco sticks contain tobacco and nicotine . Nicotine is an addictive substance.
HTP #2 Formative study	According to the materials, what addictive chemical does this product contain?	To reduce the risk of choking or poisoning: - Each capsule contains nicotine.
HTP #1 Summative study	According to the materials, what addictive chemical does this product contain?	The compatible tobacco sticks contain tobacco and nicotine . Nicotine is an addictive substance.

Table 1. Labeling comprehension questions and respective user guide information by question domain. (Continued)

Study	Question	Answer in user guide
<i>Know not to use the device with other manufacturer's items</i>		
HTP #1 Formative study	Annette ran out of tobacco sticks, but she has sticks for a different brand's device. According to the materials, what should Annette do or not do?	Use only genuine tobacco sticks and accessories intended for use with this device.
HTP #2 Formative study	Annette ran out of HTP #2 capsules, but she has a capsule for a different brand's device. According to the materials, what should Annette do or not do?	Do not use any items with any other manufacturer's heated tobacco device items.
HTP #1 Summative study	Annette ran out of tobacco sticks, but she has sticks for a different brand's device. According to the materials, what should Annette do or not do?	Use only genuine tobacco sticks and accessories intended for use with this device.
<i>Know that the product contains tobacco</i>		
HTP #1 Formative study	According to the materials, what is inside the stick?	The compatible tobacco sticks contain tobacco and nicotine .
HTP #2 Formative study	According to the materials, what is the capsule filled with?	Insert a real tobacco filled capsule into the palm-sized device to get started.
HTP #1 Summative study	According to the materials, what is inside the stick?	The compatible tobacco sticks contain tobacco and nicotine .
<i>Know that the device heats the tobacco</i>		
HTP #1 Formative study	According to the materials, what does the device do to the tobacco?	The device heats the tobacco to a high temperature.
HTP #2 Formative study	According to the materials, what does the device do to the tobacco?	Your HTP #2 system provides a new experience by heating real tobacco without burning it.
HTP #1 Summative study	According to the materials, what does the device do to the tobacco?	The device heats the tobacco to a high temperature.
<i>Know how many times you can use the tobacco consumable</i>		
HTP #1 Formative study	According to the materials, how many times can you use a stick?	ONLY USE THE TOBACCO STICK ONCE
HTP #2 Formative study	According to the materials, how many times can you use a capsule?	Each HTP #2 capsule is single use and should not be re-heated, even if a session is stopped early. HTP #2 capsules may not be heated more than once.
HTP #1 Summative study	According to the materials, how many times can you use a tobacco stick?	ONLY USE THE TOBACCO STICK ONCE
<i>Know how long the tobacco consumable will last</i>		
HTP #1 Formative study	According to the materials, after preheating, approximately how long will the experience last?	A session will last about 5 minutes.
HTP #2 Formative study	According to the materials, after preheating, approximately how long will the experience last?	Capsules last about 7 minutes or 20 puffs.
HTP #1 Summative study	According to the materials, after preheating, approximately how long will the experience last?	A session will last about 5 minutes.
<i>Know not to use the device if the battery behaves unexpectedly</i>		
HTP #1 Formative study	The battery in Annie's device appears to be leaking, what should Annie do?	DO NOT use or charge the device if the battery appears to be leaking. - DO NOT touch with bare hands. - DO NOT touch eyes. - If the liquid is swallowed, immediately call a doctor. - In case of contact with skin, wash hands with plenty of soap and water.
HTP #1 Summative study	The battery in Annie's device appears to be leaking. According to the materials, what should Annie do or not do?	DO NOT use or charge the device if the battery appears to be leaking. - DO NOT touch with bare hands. - DO NOT touch eyes. - If the liquid is swallowed, immediately call a doctor. - In case of contact with skin, wash hands with plenty of soap and water.
ENDS #1 Summative study	Al's ENDS #1 battery no longer holds a charge. According to the Guide, should he try to service or repair it himself?	To reduce the risk of injury: - Your e-vapor device and items contain no user-serviceable parts. Do not attempt to open, service, or repair; doing so may cause electric shock, thermal burns or fire.

Note: Information descriptions in the "Answer in User Guide" column are formatted as they appeared in the product materials from this research.

The correct response rates for a seventh question ('Know that the product contains tobacco') did not demonstrate statistically significant variability across studies (heterogeneity) either, but this finding should be interpreted with caution considering that Tau^2 is likely highly inflated in these analyses and is consequently widening the prediction interval $PI = [0.28, 1.00]$. Tau^2 tends to become inflated in proportional meta-analyses when the number of included studies is small (< 10 ; $N = 3$ for 'Know that the product contains tobacco' (9)), when study proportions near 100% (10) which is the case for the questions in the 'Know that the product contains tobacco' domain), and due to the bounded nature of proportions (cannot exceed 100% (11)). Thus, the wide PI is likely strongly overstating uncertainty, suggesting that the range of correct response proportions for this domain, in future studies, is likely much narrower. By contrast, the correct response rates for the remaining six question domains (i.e., 'Know to not use device if it is operating in a strange manner,' 'Know to keep device away from children,' 'Know that the device contains nicotine,' 'Know how many times you can use the tobacco consumable,' 'Know how long the tobacco consumable will last,' 'Know not to use the device if the battery behaves unexpectedly') demonstrated statistically significant variability (heterogeneity) across studies – indicating variation in correct response proportions across studies for these six domains. The pooled correct response proportions for these six domains were $\geq 86\%$, with corresponding 95% confidence intervals CI ranging from 46% to 100%, $I^2 \geq 82\%$, and $Tau^2 \geq 0.021$ – indicating that the proportions of correct responses for questions within these six domains were generally high (near a 100% correct response rate) and exhibited high variability (heterogeneity). Prediction

intervals ranged between 0% and 100%, suggesting wide expected variability in correct response proportions if the questions within these domains were asked within a new study.

DISCUSSION

Labeling comprehension research is necessary in the developmental lifecycle of a product, but it is also tedious and resource constraining because it often requires extensive formative testing. This project examined whether the need for formative metric evaluations could be tempered by identifying product-safety and product-instruction labeling comprehension questions that: 1) reliably demonstrate comprehension of target labeling information (i.e., are reliably answered correctly); and 2) are generalizable across research on ENDS and HTPs.

The data provide robust evidence that most of the assessed questions are easily comprehended and are consistently answered correctly across the included labeling comprehension research on HTPs and ENDS. Evidence from six of these domains ('Know to charge device only with proprietary accessories,' 'Know to not expose device to high temperatures,' 'Know to not use device if it becomes wet or immersed in a liquid,' 'Know to not use device if it appears damaged or broken,' 'Know not to use the device with other manufacturer's items,' and 'Know that the device heats the tobacco') indicates that if the domain-specific question were asked again in a future study, a high proportion of participants would be expected to answer the question correctly. These six domains can serve as benchmarks for crafting questions that engender high correct

Table 2. Results of the proportional meta-analyses by question domain.

Question domain	Pooled proportion and 95% CI	Prediction interval PI	Q	P -value for Q	I^2	Tau^2
Know to charge device only with proprietary accessories	0.99 [0.95, 1.00]	[0.85, 1.00]	5.34	0.15	43.8%	0.004
Know to not use device if it is operating in a strange manner	0.86 [0.68, 0.98]	[0.19, 1.00]	23.68	<.0001*	87.3%	0.036
Know to not expose device to high temperatures	0.96 [0.93, 0.98]	[0.90, 0.99]	2.12	0.55	0.0%	0.000
Know to not use device if it becomes wet or immersed in a liquid	0.98 [0.94, 1.00]	[0.88, 1.00]	3.89	0.27	23.0%	0.002
Know to not use device if it appears damaged or broken	0.97 [0.93, 1.00]	[0.87, 1.00]	3.92	0.27	23.0%	0.002
Know to keep device away from children	0.86 [0.46, 1.00]	[0.00, 1.00]	23.07	0.0001*	91.3%	0.123
Know that the device contains nicotine	0.95 [0.73, 1.00]	[0.00, 1.00]	10.97	0.004*	81.8%	0.051
Know not to use the device with other manufacturer's items	0.99 [0.95, 1.00]	[0.88, 1.00]	1.94	0.38	0.0%	0.000
Know that the product contains tobacco	0.91 [0.77, 1.00]	[0.28, 1.00]	4.60	0.10	56.5%	0.015
Know that the device heats the tobacco	1.00 [0.97, 1.00]	[0.90, 1.00]	0.03	0.99	0.00%	0.000
Know how many times you can use the tobacco consumable	0.93 [0.56, 1.00]	[0.00, 1.00]	23.85	0.0001*	91.6%	0.124
Know how long the tobacco consumable will last	0.89 [0.64, 1.00]	[0.00, 1.00]	10.99	0.004*	81.8%	0.051
Know not to use the device if the battery behaves unexpectedly	0.97 [0.86, 1.00]	[0.33, 1.00]	12.29	0.002*	83.7%	0.021

Note: Asterisks (*) indicate a statistically significant Q value (i.e., statistically significant heterogeneity). I^2 and Tau^2 describe the observed heterogeneity. P -values are formatted into two decimal places, except for lower values that necessitate additional figures for reporting accuracy.

response rates for the evaluated question domains. Five question domains ('Know to not use device if it is operating in a strange manner,' 'Know that the device contains nicotine,' 'Know how long the tobacco consumable will last,' 'Know not to use the device if the battery behaves unexpectedly,' and 'Know that the product contains tobacco') generally displayed high pooled proportions and narrow *CI*s (but wide *PI*s), indicating utility in crafting similar question for new studies, albeit context specific adjustments would help bolster the likelihood of high correct response proportions for these domains. Data for the two remaining domains ('Know to keep device away from children,' and 'Know how many times you can use the tobacco consumable') indicated wide variability in correct response proportions, thus questions from these domains would likely require significant context-based adaptation in future studies to engender high correct response proportions. Generally, items from domains that exhibited heterogeneity are still useful from an item-benchmarking standpoint, but caution should be exercised when using items from these domains by including modifications that ground the item in a device-specific context. Relatedly, it is worth noting that high correct response rates are not unambiguously positive, and these may at times reflect item simplicity, such that an item/question may be sufficiently intuitive to elicit a correct response but may be less sensitive at detecting nuanced misunderstandings. Indeed, the current research reduces the burdens of labeling comprehension research, but it should not be used as a platform to discourage robustness in the item-crafting step of future labeling comprehension research.

Together, the current findings identified varied labeling comprehension questions that can serve as benchmarks in future labeling comprehension evaluations of ENDS and HTPs to reduce the number of questions that need to be tested, thus shortening study-session time, reducing the cost of the study, and reducing or eliminating the potential need for iterative testing if the question domains of interest are those that were tested here. In a similar vein, it is important to highlight the utility of the user guide content that is tied to each tested question domain in the current study (Table 2). Answering the tested questions correctly is a function of both the structure and content of the questions and the user guide information that describes the answers to the questions. Thus, the content and structure of the user guide information that includes the questions' answers can also be used in future labeling comprehension evaluations of ENDS and HTPs to promote the clarity of information in user guides.

There are broader implications for these data that extend beyond labeling comprehension research on HTPs and ENDS, insofar as many of the questions tested here are applicable to other devices not categorized as ENDS or HTPs. Aside from five question domains ('Know that the device contains nicotine,' 'Know that the product contains tobacco,' 'Know that the device heats the tobacco,' 'Know how many times you can use the tobacco consumable,' 'Know how long the tobacco consumable will last'), all domains included here are applicable to a variety of electronic devices, particularly those that are hand-held and contain a battery, like cell-phones, video gaming consoles, smartwatches, digital cameras, and so on. Accordingly,

labeling comprehension research for these product categories may also benefit from including the questions, and user guide content, from the current work in their methodology (as applicable to device features), in terms of reducing the burden of formative testing. Specifically, the research burdens of labeling research, which often accompany other resource intensive research like usability testing, can lead to a "check mark" approach to this type of work – mainly organizations conducting narrowly focused research that is not exhaustive in evaluating comprehension of all product information. By providing researchers with benchmarks for item and user guide content creation, the current work creates capacity to more exhaustively evaluate comprehension of information related to product-specific features across the spectrum of handheld-battery powered devices. The applicability of this work to devices different from HTPs and ENDS also provides an avenue for researchers across product domains to publish their labeling comprehension work, thus increasing the data pools for, and generalizability of, future meta-analytic work.

A different broader implication involves how the current work facilitates future regulatory research. As noted at the outset, some labeling comprehension research is necessary if market authorization for the product of interest is regulated by a third-party agency, like the FDA. When considering HTPs and ENDS – potentially reduced harm products that may support the transition of adults who smoke to smoke-free products – the current research provides tools for facilitating regulatory authorization. At minimum, this work provides a resource-saving tool-kit for labeling comprehension regulatory research that allows researchers more time to explore comprehension of novel labeling information (like feature-specific labeling information), an outcome that is consistent with FDA's PMTA framework that encourages applicants to rely on existing scientific evidence (while) demonstrating consumers' comprehension of labeling (see FDA, p. 55424-55425, 2021 (2)). At most, the current work may help to streamline regulatory submission by allowing researchers to focus on labeling information related to key product features rather than labeling content that has already been evaluated in the literature.

It is worth noting that the list of question domains that was considered in the current research is not exhaustive, in part because they were developed to assess products from specific categories (ENDS and HTPs) from a single company (ALCS); however, it is critical for organizations to consider conducting and publishing similar work on their products to facilitate analyses that could further expand the current list of labeling comprehension questions. The goal of this approach would be to allow for the continued improvement of labeling comprehension methodology through a cross-company and cross-product validation process that would engender an increasing list of metrics that have question reliability and are generalizable across product categories. This process would 1) reduce resource burdens or constraints for organizations conducting product research, 2) improve communication (to consumers) and evaluation of product-safety and product-instruction information, and 3) promote a culture of reciprocal item benchmarking within the product research literature. Labeling comprehension product research in isolation has a limited scope, but cross-company and cross-product

benchmarking of items increases item reliability and generalizability and facilitates further meta-analytic work. Labeling comprehension research is rarely published because it is product specific, but this work provides evidence that labeling research can be analytically leveraged to ease the burdens of new research – In this case, facilitating labeling comprehension research by meta-analytically demonstrating correct response consistency across varied question domains, and providing design resources for researchers who are crafting labeling questions and user guides/safety guides.

SUPPLEMENTARY INFORMATION

This publication contains supplementary material available at the [Journal's website](#).

FUNDING INFORMATION

This work was funded by Altria Client Services LLC.

AUTHOR CONTRIBUTIONS

Jonathan M. Gallegos, Jennifer N. Lewis, and Elizabeth Becker are the primary authors of this work.

CONFLICTS OF INTEREST

Jonathan M. Gallegos, Jennifer N. Lewis, and Elizabeth Becker are employees of Altria Client Services LLC. The current work does not represent the views or perspectives of Altria Client Services LLC and is not intended to have binding implications on past, current, or future research conducted by Altria Client Services LLC or research that may be conducted in support of regulatory applications or scientific publications.

REFERENCES

1. Barker, T.H., C.B. Migliavaca, C. Stein, V. Colpani, M. Falavigna, E. Aromataris, and Z. Munn: Conducting Proportional Meta-Analysis in Different Types of Systematic Reviews: A Guide for Synthesisers of Evidence; *BMC Med. Res. Methodol.* 21 (2021) 189. DOI: 10.1186/s12874-021-01381-z
2. U.S. Food and Drug Administration (FDA), Department of Health and Human Services: Premarket Tobacco Product Applications and Recordkeeping Requirements; Federal Register, 2021, Docket No. FDA-2019-N-2854. Available at: <https://www.federalregister.gov/documents/2021/10/05/2021-21011/premarket-tobacco-product-applications-and-record-keeping-requirements> (assessed June 2026)
3. Willis, G.B.: *Cognitive Interviewing: A Tool For Improving Questionnaire Design*; Sage Publications, Thousand Oaks, CA, USA, 2005. ISBN: 0-7619-2804-9

4. Singer, J.P., G.M. Balliro, and N.D. Lerner: *Manufacturer's Guide to Developing Consumer Product Instructions*; edited by T.P. Smith, U.S. Consumer Product Safety Commission (CPSC) Washington, DC, USA, 2003. Available at: <https://www.cpsc.gov/s3fs-public/pdfs/guide.pdf> (assessed June 2026)
5. Li, D., T. Cassidy, and D. Bromilow: *The Design of Product Instructions*; Chapter 5 *in: Advances in Industrial Design Engineering*, edited by D.A. Coelho, IntechOpen Limited, London, UK, 2013. DOI: 10.5772/54011
6. McCarthy, R.L., J.P. Finnegan, S. Krumm-Scott, and G.E. McCarthy: *Product Information Presentation, User Behavior, and Safety*; *Proceedings of the Human Factors Society Annual Meeting* 28 (1984) 81–85. DOI: 10.1177/154193128402800124
7. Braun, C.C., S.A. Glusker, R.S. Holt, and N.C. Silver: *Adding Consequence Information to Product Instructions: Changes in Hazard Perceptions*; *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 39 (1995) 346–349. DOI: 10.1177/154193129503900511
8. Jadotte, Y.T., A. Moyer, and J. Gurevitch: *Meta-Analysis*; *in: The Cambridge Handbook of Research Methods and Statistics for the Social and Behavioral Sciences*, Cambridge University Press, Cambridge, UK, 2023, pp. 583–605. DOI: 10.1017/9781009010054.028
9. Borenstein, M., L.V. Hedges, J.P. Higgins, and H.R. Rothstein: *Introduction to Meta-Analysis*; John Wiley & Sons, Toronto, Canada, 2021. DOI: 10.1002/9780470743386
10. Hamza, T.H., H.C. van Houwelingen, and T. Stijnen: *The Binomial Distribution of Meta-Analysis was Preferred to Model Within-Study Variability*; *J. Clin. Epidemiol.* 61(2008) 41–51. DOI: 10.1016/j.jclinepi.2007.03.016
11. Barendregt, J.J., S.A. Doi, Y.Y. Lee, R.E. Norman, and T.Vos: *Meta-Analysis of Prevalence*; *J. Epidemiol. Community Health* 67 (2013) 974–978. DOI: 10.1136/jech-2013-203104.

Corresponding author:

*Jonathan Gallegos, PhD
Altria Client Services LLC
601 East Jackson Street
Richmond, VA 23219
USA
E-mail: Jonathan.M.Gallegos@altria.com*