Optimising usability of an economic decision support tool:
Prototype of the EQUIPT tool

Short title: Optimising usability of a decision support tool

Kei Long Cheung
Caphri school of Public Health and Primary Care
Health Services Research, Maastricht University
P.O. Box 616, 6200 MD Maastricht, The Netherlands
kl.cheung@maastrichtuniversity.nl, T +31 43 38 82294, F +31 43 38 84162

Mickaël Hiligsmann
Caphri school of Public Health and Primary Care
Health Services Research, Maastricht University
Maastricht, The Netherlands

Maximilian Präger
Helmholtz Zentrum München (GmbH),
German Research Center for Environmental Health (Institute of Health Economics and Health Care Management)
Neuherberg, Germany

Teresa Jones
Health Economics Research Group,
Brunel University London
Uxbridge, United Kingdom

Judit Józwiak-Hagymásy
Syreon Research Institute,
Budapest, Hungary

Celia Muñoz
Centre for Research in Economics and Health (CRES)
University Pompeu Fabra
Barcelona, Spain

Adam Lester-George
LeLan Solutions
Bristol, UK

Subhash Pokhrel
Health Economics Research Group,
Brunel University London
Uxbridge, United Kingdom

Ángel López-Nicolásc
Polytechnic University of Cartagena
Cartagena, Spain

Marta Trapero-Bertrana
Centre for Research in Economics and Health (CRES)
University Pompeu Fabra
Barcelona, Spain

Silvia M.A.A. Evers
Caphri school of Public Health and Primary Care
Health Services Research, Maastricht University
Trimbos Institute, Netherlands Institute of Mental Health and Addiction
Maastricht, The Netherlands

Hein de Vries
Caphri school of Public Health and Primary Care
Health Promotion, Maastricht University
Maastricht, The Netherlands

1

ABSTRACT

Objectives: Economic decision-support tools can provide valuable information for tobacco control stakeholders, but their usability may impact the adoption of such tools. This study aims to illustrate a mixed-method usability evaluation of an economic decision-support tool for tobacco control, using the EQUIPT ROI tool prototype as a case study.

Methods: A cross-sectional mixed methods design was used, including a heuristic evaluation, a thinking aloud approach, and a questionnaire testing and exploring the usability of the Return of Investment tool.

Results: A total of 66 users evaluated the tool (thinking aloud) and completed the questionnaire. For the heuristic evaluation, 4 experts evaluated the interface. In total 21% of the respondents perceived good usability. A total of 118 usability problems were identified, from which 26 problems were categorised as most severe, indicating high priority to fix them prior to implementation.

Conclusions: Combining user-based and expert-based evaluation methods is recommended as these were shown to identify unique usability problems. The evaluation provides input to optimise usability of a decision-support tool, and may serve as a vantage point for other developers to conduct usability evaluations to refine similar tools prior to wide-scale implementation. Such studies could reduce implementation gaps by optimising usability, enhancing in turn the research impact of such interventions.

Keywords: Usability evaluation, decision-support tool, thinking aloud, heuristic evaluation

Acknowledgements

1 INTRODUCTION

On behalf of society, healthcare budget holders have to make choices regarding the use and diffusion of healthcare interventions and reimbursement for their costs. While the number of available health technologies is increasing, budgets are limited. Thus, healthcare policy makers need to know whether the societal benefits of a particular technology or treatment are worth the investment required to offer it (1). Consequently, health technology assessment (HTA) is increasingly important, as it informs decision-making on how to obtain the best value for money (2). HTA is a policy-oriented form of research designed to inform decision-makers on the relative effectiveness or the cost-effectiveness of a new technology in comparison with current or best practice (3). Although stakeholders' uptake of these studies is increasing (4, 5), the extent to which HTA information is used in policy making varies. For this reason, better matching of user needs and HTA information is necessary to strengthen the use of HTA (5, 6). Despite the increased amount of cost-effectiveness information, it is a challenge to enhance the uptake of such information by policy makers and other stakeholders (7, 8).

1.1 EQUIPT Return on Investment Tool

In order to stimulate use of cost-effective information, the NICE Tobacco Return on Investment Tool was developed and tested in England by Brunel University London, the National Institute of Health and Clinical Excellence, LeLan Solutions (9) and regional tobacco control organisations in 2012 (9). Internationally, 6.3 million deaths and 6.3% of all disability-adjusted life years (DALYs) lost in 2010 were estimated to be attributable to tobacco smoking (10). In 2003, the World Health Organisation's Framework Convention on Tobacco Control (11) was ratified by EU member states (12). The Return on Investment Tool is an economic decision-support tool that provides policy-makers and other stakeholders with customized information about the economic and wider returns that investment in packages of tobacco control methods may generate. As such economic decision-support tools provide valuable information for tobacco control stakeholders, the European-study on quantifying utility of investment in protection from tobacco (EQUIPT) is developing a new web-based version of the tobacco ROI tool (hereafter 'ROI tool') for use across several European countries (13). A prototype of the EQUIPT ROI tool has been developed for evaluation purposes.

A unique characteristic of the ROI tool is its online availability, which allows decision-makers (and other stakeholders) to interact with the model using an interface. Due to several barriers (e.g. lack of adequate presentation format), the mere existence of HTA information does not automatically translate into usage by decision-makers. Therefore, better matching of user needs and HTA is needed to strengthen the use of HTA (5, 6). The lack of human centeredness among stakeholders may create usability problems, leading to high attrition rates, and resulting in "high tech-with-a-low impact" technologies (14). Therefore, optimising usability of the decision-support tools, such as the ROI tool, is important in enhancing their

usage. Thus, this study stresses the importance of usability evaluation and aids to the HTA discipline by illustrating methods one may use to optimise HTA decision-support tools.

1.2 Importance of Usability Evaluation

Despite the valuable information these tools may provide, users may still experience difficulties with a tool's user interface and may decide not to use it (15, 16). Usability is defined as ''the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use'' (17). Usability evaluation is important to make tools and other interventions efficient, effective, and satisfying to use (18, 19). A good user interface (UI) can make the difference between continued or discontinued intervention usage (20). Moreover, users may decide not to adopt an intervention based on perceived usability and merit. The perceived usability of a tool's UI has been demonstrated as an important determinant of an individual's intention to use an intervention (20). In the context of an online store, usability issues may affect the customers' perceptions and behaviours, with more usable websites enhancing positive attitudes towards online stores and increasing conversion rates (21).

Various usability principles determine use and should guide the design of a UI (22, 23). Usability studies provide developers with insights into the user-friendly aspects of a UI, its violation of design principles, and its specific problems (24). Hence, it is important to evaluate usability in order to identify and fix problems in—and thus optimise—an economic decision-support tool, such as the EQUIPT ROI tool.

1.3 Aims

This study aims to illustrate a mixed-method usability evaluation of an economic decision-support tool for tobacco control, using the EQUIPT ROI tool prototype as a case study. This study may help to improve the ROI tool and also serve as a vantage point for developers of similar tools to conduct a usability evaluation.

2 METHOD

This is an exploratory study to evaluate the usability of the ROI tool prototype with a mixed methods design. This section first details the ROI tool prototype and the study participants; followed by a description of the methods of the usability evaluation.

2.1 The ROI tool prototype

EQUIPT is developing a new version of the similar NICE Tobacco Return on Investment Tool (which is accessible online (9)) for use across several European countries (20). At the time of testing, the EQUIPT study had developed a prototype of the ROI tool with dummy (UK-based) data, running on the Microsoft Excel platform with a UI built in Visual Basic for Applications (VBA). The development of the prototype was based on the design plan of EQUIPT and the results of stakeholder interviews (13, 25). These interviews

revealed that it is important that the tool be easy to use, include several outputs (e.g. cost of smoking, quality of life, and effectiveness of smoking cessation interventions), and have the option to yield results in different time horizons (25). After downloading and opening the Excel file, this ROI tool prototype allows users to explore the tool's UI. Therein, users can create a "package" of tobacco control methods (which may include a mix of pharmacological and behavioural support components) and view the estimated return on investment and cost-effectiveness of such a package. At the time of the study, the Excel prototype was only functional for Windows users. The final ROI tool is intended to be a platform-agnostic, web-based tool without the need to download and open a file (see Supplementary file 1 for a video illustrating the functionalities of the tool).

A typical experience of using the ROI tool starts with a screen to select the country (or a more specific area) of interest. This allows the tool to populate the model with country-specific estimates for relevant indicators (e.g. population size, smoking prevalence, and availability and cost of different tobacco control interventions). The tool then automatically generates a "Current Package" of interventions that reflects that country's current provision of tobacco control services. The user may view and amend the parameters in the Current Package and explore the short, medium, and long-term impacts of different combinations of interventions and strategies (i.e. to explore the impact of different investment scenarios).

2.2 Participants

End-users of the tool (i.e. stakeholders) and e-Health experts were recruited. European tobacco control stakeholders were recruited through convenience sampling. Stakeholders from Germany, Hungary, Netherlands, Spain, and the UK were included, as the ROI tool is being developed for these countries. Five relevant categories of stakeholders were approached: decision-makers, professionals/service providers, evidence generators (academics and researchers), advocates of health promotion, and "other" stakeholders. This procedure included almost all those eligible who would be covered by the Centre for Disease Control and Prevention's stakeholder definition (26). It was not possible to predict *a priori* the composition of the final sample, which was determined by those who agreed to participate. All participants filled in an informed consent after reading the information letter provided prior to the study. Regarding e-health experts, three experts from the Netherlands and one from the UK were recruited through convenience sampling. This is in line with recommendations to include three to five experts for a heuristic evaluation (see below) (27, 28). An e-mail was sent to the experts, including instructions and the relevant attachments (see below).

2.3 Usability evaluation

Multiple methods were used to evaluate the ROI tool's usability, detailed below. Before each evaluation, a video explaining the functionalities of the ROI tool was shown (see Supplementary file 1).

2.3.1 Test-scenario

The test scenario was written out and consisted of six questions leading the users through the main functionalities of the ROI tool. While answering the questions, users performed tasks that were similar to those that would be performed by future users. In the test scenario, users were instructed to imagine being part of a national public health team with an increased budget. The goal was to explore whether it was cost-effective to invest the additional budget in the pro-active telephone support intervention. Users were asked to (1) report the percentage adult smokers, (2) report the percentage of uptake and effectiveness of the pro-active telephone support intervention, (3) report the economic return expected for every euro invested in the current package of interventions, (4) to invest the budget in the pro-active telephone support intervention, (5) to determine the cost per Quality Adjusted Life Years gained from this alternative level of investment compared with the current package, and (6) generate the narrative report and report the quasi-societal net present value after 5 years. The ROI tool prototype was piloted (n=12) with the test-scenario resulting in minor adjustments to the final ROI tool prototype and test scenario (see Supplementary file 2).

2.3.2 Questionnaire

All participating European stakeholders completed a questionnaire (see Supplementary file 3 for the phrasing of each item) as part of an EQUIPT study. This questionnaire includes items testing and evaluating the usability of the ROI tool. In order to evaluate the ROI tool quantitatively, the System Usability Scale (SUS) was used (29), which is composed of 10 statements assessing the user's strength of agreement on the intervention's usability. To enhance consistency in the questionnaire, users are asked to score each statement on a 7-point scale instead of the original SUS 5-point scale of strength of agreement (1=strongly disagree, 7=strongly agree with a statement). Final scores for the SUS can range from 0 to 100, where higher scores indicate better usability.

2.3.3 Heuristic evaluation

Heuristic evaluation is an evaluation method, using experts (rather than end-users), aimed to identify usability problems with user interfaces (28), and is based on Nielsen's recommendations (23). These recommendations are a list of recognised usability principles, which are called heuristics. Heuristics are general principles referring to common properties of usable interfaces. These are: (1) Visibility of system status, (2) Match between system and the real world, (3) User control and freedom, (4) Consistency and standards, (5) Error prevention, (6) Recognition rather than recall, (7) Flexibility and efficiency of use, (8) Aesthetic and minimalist design, (9) Help users recognize, diagnose, and recover from errors, and (10) Help and documentation. These ten heuristics (23) were translated into a series of questions (e.g. ''Are there any incidents where the ROI tool is unresponsive or slow? If so, please list them and indicate how severe they are'' – see Supplementary file 3), which were then developed into a heuristic evaluation form.

Four eHealth experts served as evaluators and were provided with the ROI tool prototype, a link to the introductory video, the heuristic evaluation form, and the slightly adjusted test scenario (tailored to the experts). Inspection of the UI was conducted independently, assessing the questions for each of the six scenarios with responses written on the heuristic evaluation form template.

The evaluators watched a video providing a brief introduction to the ROI tool. The heuristic evaluation form considered recommendations by Nielsen (23), which includes (1) name of heuristic and its general description, and (2) specific questions about the heuristics. The test scenario was used to guide the evaluators through the main functions of the ROI tool.

2.3.4 Thinking aloud procedure

The think aloud method involves participants (representing end-users) thinking aloud as they are performing a set of specified tasks (or going through a scenario) (30). The method was used to assess the stakeholders' reasoning and the source of their problems while using the ROI tool. Country-specific researchers planned a Skype meeting with each participating stakeholder. If the usage of Skype was not possible (e.g. forbidden in the institution of the stakeholder), other communication methods were used, such as telephone and face-to-face meetings. In the meetings, which lasted approximately an hour, participants were asked to share screens with the researchers. Participants were also asked to open the prototype of the ROI tool and a link leading them to an online survey, which included the questionnaire and the test scenario. The session started with the video explaining the functionalities of the ROI tool (see Supplementary file 1). After participants read the scenario, they were asked to verbalise their thoughts (think aloud) while going through the tasks of the test scenario. To encourage participants to talk freely, the researchers emphasized that the intention was to evaluate the program and not the participants' behaviour. As the meetings were online, participants were able to decide for themselves where they felt most comfortable conducting the evaluation. Participants were encouraged to speak constantly, and researchers reminded them to keep talking if they fell silent. Interruption by the observers was kept to a minimum. Researchers made field notes for each task of the scenario. Microsoft Expression Encoder 4 Screen Capture was used to capture screen display and participants' verbal comments. This allowed the researchers to review the sessions to identify problems or comments that were missed during the sessions.

2.4 Data analysis

Analyses were conducted using SPSS Statistics 22 software. In order to get an overall picture of the usability of the ROI tool, we calculated the SUS score (29). Scores on the SUS items were recoded to a score between '0' and '6', summed, and then multiplied by (10/6). This resulted in a SUS score between 0 and 100 (29).

For the heuristic evaluation, data in the templates were assembled in one document. The content of this document was categorised and analysed according to Nielsen's heuristics. For each heuristic, a list of usability violations was created and sorted according to severity in each country. A discussion between

two independent researchers (KLC & MP) was planned to discuss categorisations according to the heuristics. Five country-specific researchers then independently rated each usability problem for its severity (five-point Likert scale ranging from '0' no usability problem to '4' severe usability problem). As the mean scores of the severity ratings of the four evaluators were found to be reliable (31), we averaged the severity ratings of the five researchers to indicate the severity of the specific usability problem.

Observations and field notes from the thinking aloud approach were reviewed and comments (i.e. relevant thoughts concerning problems during tasks or when tasks went noticeably smoothly) were identified by the country-specific researchers. In each country, two independent researchers reviewed the comments and identified violations of the heuristics; if researchers disagreed, consensus was reached through discussion. Two researchers compiled the five resulting lists of usability problems and categorised the usability violations according to the heuristics. Again, five country-specific researchers independently rated each usability problem for its severity (five-point Likert scale ranging from '0' no usability problem to '4' severe usability problem). These scores were then averaged to indicate the severity of the specific usability problem.

3 RESULTS

3.1 Sample characteristics of European stakeholders

Of the 151 European stakeholders approached in the five countries, 66 respondents agreed to take part, resulting in a 44% response rate. About 41% respondents were female, and the average age of stakeholders was 49.4 years. Of the 66 respondents, 23% were Dutch, 24% Hungarians, 21% Germans, 21% Spanish and 11% British. Respondents were from different categories of stakeholders: 18 (27%) decision-makers, 23 (35%) professional service providers, 13 (20%) evidence generators, 2 (3%) advocates of health promotion, and 10 (15%) other stakeholders.

3.2 Usability evaluation

The mean of the SUS score was 51.80, with a standard deviation of 18.79. In total, 14 participants (21%) had a ≥68 SUS score (SUS score of 68 being an average usability score of interfaces (32)). The think aloud procedure revealed a total number of 118 distinct usability problems. The heuristic evaluation method revealed 28 usability problems, while 14 of these overlapped with the think aloud procedure. We found 3 cosmetic problems, with a score between 0 and 0.9 (e.g. old-fashioned layout with main hyperlinks in capital letters), 27 minor problems with a score between 1.0 and 1.9 (e.g. one is not able to enlarge the screen of the tool, and the tool shows former smokers estimates before current smokers estimates, which may be counterintuitive), and 57 major problems with a score between 2.0 and 2.9 (e.g. buttons do not look clickable, and the tool is unable to print/export the graphics).

3.3 Severe usability problems

It was determined that 26 usability problems should be fixed with high priority; those with a severity rating of 3.0 and up (see Table 1). A complete overview of the usability problems, the mean severity ratings, and the categorization of the problem types based on Nielsen's heuristics is provided in Supplementary file 4.

4 DISCUSSION

This study illustrated, via a mixed-method design, the evaluation of an economic decision-support tool, using the EQUIPT ROI tool as a case study. This resulted in the identification of a rich list of usability problems regarding the ROI tool. Literature indicates that it is preferable to use a mixed method for usability evaluation, especially user-based and expert-based methods, as these two methods complement each other. The think aloud method provides a rich source of data and shows insights into why certain aspects are usability problems. The heuristic evaluation is cheap and less time consuming, relies less on task selection, and may better capture different usability problems (28).

4.1 Usability of the ROI tool

Overall, the results indicated low usability of the ROI tool UI at the time of this study, as the SUS score fell below the average usability score of interfaces (32). Hence, it is necessary to improve the usability of the UI, and thus the tool. For this purpose, the list of usability problems was given to the tool development team so that the UI could be revised. The development team and researchers conducting this usability study discussed the usability problems, and based on the severity ratings, some issues were prioritized to implement changes to the UI. No high priority usability problems (usability catastrophes) were identified in the heuristics 'Visibility of system status' and 'Flexibility and efficiency of use'. Most severe problems were found in the heuristics 'User control and freedom' and 'Consistency and standards'. Designers of such decision-support tools may thus be especially aware of the need to minimise instances where important changes cannot be easily undone or where inconsistencies arise concerning language use or functionality (23). Implications of the findings were to fix these problems, especially those with high priority. The user interface needed further development and solutions were proposed by the team and implemented for the final version of the ROI tool. For instance, as the option for displaying different types of results was not plainly evident, the results menu is now displayed in the 'open' state, so the full list of options is clearly visible. For the full list of resolutions of the major problems, see Table 2.

4.2 Thinking aloud vs. Heuristic evaluation

Literature showed that the heuristic evaluations detect relatively large numbers of minor usability problems (28, 33), and there was an indication that the method detects more low- and high-severe problems than the think-aloud method (33). However, this study did not yield such trend, as experts identified problems

with a variety of severity scores. This indicates the need for more research regarding the differences between these methods in different contexts. Literature also indicates that both heuristic evaluation and think aloud methods are effective in revealing usability problems, but that they capture different problems (34). Moreover, there is indication that more than half of the problems identified by the heuristic evaluation, may not be detected by the think-aloud method (35). Findings of this study are consistent with prior studies, as the heuristic evaluation yielded 14 more unique usability problems, which is half of the detected problems with the heuristic evaluation method. Yet, the differences between the perspectives of these methods are unclear, indicating the need for more research. In addition, this study suggests the need to evaluate usability in the future development of decision-support tools. The usability of these tools may not be optimal while it may be paramount for adoption, in line with important innovation characteristics to enhance adoption outlined in the Diffusion of Innovation theory (36).

In line with our findings and previous literature, we recommend combining user-based evaluation methods (e.g. think-aloud method) with expert-based evaluation methods (e.g. heuristic evaluation). The think-aloud method and heuristic evaluation are the most popular methods used in usability studies. Each method has its advantages and disadvantages. However, when used in tandem, they complement one another, resulting in their collective application being more powerful than when applied in isolation (28). While the think-aloud method may provide a rich source of data (including the cognitions behind it), verbal protocols may not may not capture all thought processes. Results may be impacted by the user characteristics, the selected task, and the extent to which the users are representative to the end-users. The heuristic evaluation may therefore identify different usability problems due to the use of experts and the relatively unstructured approach (rather than capturing cognitions during the selected tasks) (28). It is recommended that design of any interactive intervention follow an iterative process of initial design (prototype) evaluation, and redesign (28), using a combination of evaluation methods (such as the methods used in this study). This iterative process enhances the match between user needs and the intervention, and the identification of problems in early development stages (which may avoid unnecessary high adjustment costs).

In practice, the usability evaluation plan depends on several factors, including time and financial constraints, and the availability of end-users. There seems to be an increasing need to make economic models open access and available to the community. In view of this potential, future trend, optimising usability of these economic models will be an important challenge in the field of HTA. This study may therefore serve as a vantage point to enhance usability of these online available models.

4.3 Limitations

This study has several limitationsOne limitation is the potential lack of generalisability of the results, as participants were limited to the stakeholders of tobacco control. Findings regarding the problems identified in each method are tool specific. Another limitation is that this sample may not reflect stakeholders of

tobacco control, as we cannot rule out participation bias. It is feasible that participating stakeholders were more familiar with economic decision-support tools or health economic terminology. Moreover, this study evaluated and tested the usability of the prototype. It is therefore uncertain to what extent the final ROI tool has major usability problems. Furthermore, due to time constraints, the scenario for the usability evaluation did not cover all functionalities of the ROI tool prototype. More usability problems may have been detected if stakeholders were asked to explore other metrics and functionalities.

4.4 Conclusion

In conclusion, despite carefully designed interfaces of decision-support tools, evaluating the usability enabled the development of a tool that is better aligned to the requirements of its stakeholders. This study used a mixed-method design to detect usability flaws, and may serve as a vantage point for other developers to conduct usability evaluations to refine similar tools prior to wide-scale implementation. Combining user-based and expert-based evaluation methods is recommended as these were shown to identify unique usability problems. This study aids the development of an interactive intervention and reduces implementation gaps by optimising usability, enhancing in turn the research impact of such interventions.

References

1.	Park AL, McDaid D, Weiser P, Von Gottberg C, Becker T, Kilian R, et al. Examining the cost effectiveness of interventions to promote the physical health of people with mental health problems: a systematic review. BMC public health. 2013;13:787.

2.	Nicod E, Kanavos P. Commonalities and differences in HTA outcomes: a comparative analysis of five countries and implications for coverage decisions. Health policy. 2012;108(2):167-77.

3.	Drummond MF, Schwartz JS, Jönsson B, Luce BR, Neumann PJ, Siebert U, et al. Key principles for the improved conduct of health technology assessments for resource allocation decisions. International journal of technology assessment in health care. 2008;24(03):244-58.

4.	van Velden ME, Severens JL, Novak A. Economic evaluations of healthcare programmes and decision making. Pharmacoeconomics. 2005;23(11):1075-82.

5.	Oliver K, Innvar S, Lorenc T, Woodman J, Thomas J. A systematic review of barriers to and facilitators of the use of evidence by policymakers. BMC health services research. 2014;14:2.

6.	Macintyre S, Chalmers I, Horton R, Smith R. Using evidence to inform health policy: case study. BMJ: British Medical Journal. 2001;322(7280):222.

7.	Drummond M. Economic evaluation in health care: Is it really useful or are we just kidding ourselves? Australian Economic Review. 2004;37(1):3-11.

8.	Garrido MV. Health technology assessment and health policy-making in Europe: current status, challenges and potential: WHO Regional Office Europe; 2008.

9.	NICE. The NICE Tobacco Return on Investment Tool, from http://www.nice.org.uk/usingguidance/implementationtools/returnoninvesment/TobaccoROITool.jsp. 2012.

10.	Lim SS, Vos T, Flaxman AD, Danaei G, Shibuya K, Adair-Rohani H, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. The lancet. 2013;380(9859):2224-60.

11.	WHO "WHO Framework Convention on Tobacco Control. Available from: http://www.who.int/fctc/en/ Accessed: 17 December 2015. 2003.

12.	European Commission. "Public Health: Tobacco Policy." Available from: http://ec.europa.eu/health/tobacco/policy/index_en.htm. Accessed: 17 December 2015. 2012.

13.	Pokhrel S, Evers S, Leidl R, Trapero-Bertran M, Kalo Z, De Vries H, et al. EQUIPT: protocol of a comparative effectiveness research study evaluating cross-context transferability of economic evidence on tobacco control. BMJ open. 2014;4(11):e006945.

14.	van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. Journal of medical Internet research. 2011;13(4):e111.

15.	Eysenbach G. The law of attrition. Journal of medical Internet research. 2005;7(1).

16.	Horsky J, Schiff GD, Johnston D, Mercincavage L, Bell D, Middleton B. Interface design principles for usable decision support: a targeted review of best practices for clinical prescribing interventions. Journal of biomedical informatics. 2012;45(6):1202-16.

17.	9241-11 I. Ergonomic requirements for office work with visual display terminals (VDT)s – Part 11 Guidance on usability. 1998.

18.	Bevan N. Usability is quality of use. Advances in human factors ergonomics. 1995;20:349-.

19.	Wichansky AM. Usability testing in 2000 and beyond. Ergonomics. 2000;43(7):998-1006.

20.	Cho V, Cheng TE, Lai WJ. The role of perceived user-interface design in continued usage intention of self-paced e-learning tools. Computers & Education. 2009;53(2):216-27.

21.	Becker S, Mottay FE. A global perspective on web site usability. Software, IEEE. 2001;18(1):54-61.

22.	Ahern DK. Challenges and opportunities of eHealth research. American journal of preventive medicine. 2007;32(5):S75-S82.

23.	Nielsen J. Heuristic evaluation. Usability inspection methods. 1994;17(1):25-62.

24.   Voncken-Brewster V, Moser A, Van Der Weijden T, Nagykaldi Z, De Vries H, Tange H. Usability evaluation of an online, tailored self-management intervention for chronic obstructive pulmonary disease patients incorporating behavior change techniques. JMIR Research Protocols. 2013;2(1):e3.

25.   Cheung KL, Evers SM, Hiligsmann M, Vokó Z, Pokhrel S, Jones T, et al. Understanding the stakeholders' intention to use economic decision-support tools: A cross-sectional study with the Tobacco Return on Investment tool. Health Policy. 2015.

26.   Centre for Disease Control and Prevention. Introduction to program evaluation for public health programs: A self-study guide. In: Services USDoHaH, editor. http://www.cdc.gov/eval/guide/CDCEvalManual.pdf2011.

27.   Nielsen J, Landauer TK, editors. A mathematical model of the finding of usability problems. Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems; 1993: ACM.

28.   Jaspers MW. A comparison of usability methods for testing interactive health technologies: methodological aspects and empirical evidence. International journal of medical informatics. 2009;78(5):340-53.

29.   Brooke J. SUS-A quick and dirty usability scale. Usability evaluation in industry. 1996;189(194):4-7.

30.   Jaspers MW, Steen T, van Den Bos C, Geenen M. The think aloud method: a guide to user interface design. International journal of medical informatics. 2004;73(11):781-95.

31.   Nielsen J, editor Reliability of severity estimates for usability problems found by heuristic evaluation. Posters and short talks of the 1992 SIGCHI conference on Human factors in computing systems; 1992: ACM.

32.   Brooke J. SUS: a retrospective. Journal of Usability Studies. 2013;8(2):29-40.

33.   Jeffries R, Miller JR, Wharton C, Uyeda K, editors. User interface evaluation in the real world: a comparison of four techniques. Proceedings of the SIGCHI conference on Human factors in computing systems; 1991: ACM.

34.   Lai T-Y, Larson EL, Rockoff ML, Bakken S. User Acceptance of HIV TIDES—tailored interventions for management of depressive symptoms in persons living with HIV/AIDS. Journal of the American Medical Informatics Association. 2008;15(2):217-26.

35.   Bailey RW, Allan RW, Raiello P, editors. Usability testing vs. heuristic evaluation: A head-to-head comparison. Proceedings of the Human Factors and Ergonomics Society Annual Meeting; 1992: SAGE Publications.

36.   Rogers EM. Diffusion of innovations: Simon and Schuster; 2010.

Tables

Table 1. Severe usability problems

Usability principle	Usability problems	Severity score
Match between system and the real world	1. Some input data (e.g. uptake or cost of interventions) are not country-specific. For Netherlands, usually estimates used are from 15 plus instead of 16 plus.	3.2
User control and freedom	2. Narrative report provides the wrong estimates (generation of the report goes wrong)	4
	3. Discount rate and threshold cannot be modified correctly.	3.6
	4. It is too intricate to find the different types of the results. The button of the indicators (BCA, ABD, ICER, NPV) is not evident.	3.4
	5. Details of certain interventions could not be found easily , not intuitive (e.g. telephone support)	3.2
	6. Changing the values of parameters: sometimes not possible (errors), sometimes too little steps when clicking on "+" and "-", recognizing that figures could be filled into the fields by Keypad	3.2
Consistency and standards	7. Problems with the understanding of the structure of the tool, not clear what the three buttons represent (they should be placed next to its description)	3.4
	8. For Hungary: the tool does not classify each intervention to the right group. (e.g. top-level interventions were classified to the pharmacotherapy group)	3.2
	9. The logic of the tool is difficult to understand at first use.	3.2
	10. Difficulties in finding the real analysis part of the tool	3.2
	11. It is not clear for the users that the packages of the interventions are compared with each other. (The users think that one intervention can be compared with another.)	3.2
	12. The button to go back to the overview from the individual interventions was not clear.	3
	13. You can see when an intervention is Dominant, but not the corresponding figure of savings (Results).	3
	14. Terminology in part not uniform (it would be better to use precise and consistent usage, combinations of words, e.g. cost-benefit ratio vs. benefit-cost ratio).	3
Error prevention	15. Sometimes interviewees want to leave the results section and to perform some analyses elsewhere in the tool, although the analysis would have to be performed within the results section. Results may imply an end-state after an analysis and changing something may imply that the analysis needs to be redone.	3.2
	16. Settings and results of the analysis should be saved within the tool, e.g. such that unintended loss of calculation results by leaving the results section is avoided	3
Recognition rather than recall	17. The mechanism of the three packages was not (fully) understood, especially when and where to calculate an alternative package and what exactly the baseline package means	3.4
	18. It is not clear that the users have to run the model to get the results (Change from Interventions to Results).	3.2
	19. More detailed definition is needed to distinguish between interventions aimed at the whole smoking population and interventions targeted at smokers who are making quit attempts (Interventions).	3
	20. Uncertain and unclear how to add the investment	3

Aesthetic and minimalist design	21. Sometimes it is not clear to which absolute number a figure in percent relates to	3
Help users recognize, diagnose, and recover from errors	22. Some runtime errors appeared, tool crashed several times, especially when some estimates in the results section were changed, added a new intervention, selected a region, or generating the narrative report	4
	23. No solutions were provided in the error alerts	3.4
Help and documentation	24. Help buttons didn't provide any information	3.2
	25. Need of a glossary to understand the tool faster and better, lack of some definitions, e.g. quasi-societal, reach, and uptake	3
	26. There is not enough help available in the tool. It would be good if the definitions of the interventions, the calculations of the cost of interventions, and the source of the data appear.	3

a) Represents the number of users experiencing this problem of the think aloud procedure.

Severity score: 0-0.9 cosmetic problem, 1-1.9 minor usability problem, 2-2.9 major usability problem, 3-4 usability catastrophe

Table 2: Major usability problems and resolutions

Usability principle	Usability problems	Resolutions
Match between system and the real world	27. Some input data (e.g. uptake or cost of interventions) are not country-specific. For Netherlands, usually estimates used are from 15 plus instead of 16 plus.	Country-specific data will be updated in the final web tool.
User control and freedom	28. Narrative report provides the wrong estimates (generation of the report goes wrong)	Redevelopment of the template narrative report
	29. Discount rate and threshold cannot be modified correctly.	Discount rates and threshold will include more options.
	30. It is too intricate to find the different types of the results. The button of the indicators (BCA, ABD, ICER, NPV) is not evident.	The menu is now displayed in the 'open' state, so the full list of options is visible
	31. Details of certain interventions could not be found easily , not intuitive (e.g. telephone support)	Balloon pop-ups were added to the interface to highlight that more interventions details are visible on clicking the intervention name.
	32. Changing the values of parameters: sometimes not possible (errors), sometimes too little steps when clicking on "+" and "-", recognizing that figures could be filled into the fields by Keypad	Errors in the code were fixed and the increment size of the +/- buttons was increased
Consistency and standards	33. Problems with the understanding of the structure of the tool, not clear what the three buttons represent (they should be placed next to its description)	The Home page was structured accordingly
	34. For Hungary: the tool does not classify each intervention to the right group. (e.g. top-level interventions were classified to the pharmacotherapy group)	This coding error was fixed
	35. The logic of the tool is difficult to understand at first use.	'Breadcrumbing' of session progress added and access to each part of the tool facilitated via hotlinking and drpo-down menus in the navigation bar
	36. Difficulties in finding the real analysis part of the tool	More information was provided on the front page explaining the layout of the tool
	37. It is not clear for the users that the packages of the interventions are compared with each other. (The users think that one intervention can be compared with another.)	Different terms were used for "Current Package" and "Alternative Package", now referring to the "Current Investment Scenario" and "Prospective Investment Scenario".
	38. The button to go back to the overview from the individual interventions was not clear.	The prototype was limited regarding its space; better navigation cues will be implemented for the Web tool
	39. You can see when an intervention is Dominant, but not the corresponding figure of savings (Results).	A range of additional outcomes were included in the Results Menu, including cost/saving breakdowns
	40. Terminology in part not uniform (it would be better to use precise and consistent usage, combinations of words, e.g. cost-benefit ratio vs. benefit-cost ratio).	Terminology was made more uniform.
Error prevention	41. Sometimes interviewees want to leave the results section and to perform some analyses elsewhere in the tool, although the analysis would have to be performed within the results section. Results may imply an end-state after an analysis and changing something may imply that the analysis needs to be redone.	Warnings added when users attempt to leave the Results Area alerting that changes to the analysis will be lost
	42. Settings and results of the analysis should be saved within the tool, e.g. such that unintended loss of calculation results by leaving the results section is avoided	This coding error was fixed in a subsequent release
Recognition rather than recall	43. The mechanism of the three packages was not (fully) understood, especially when and where to calculate an alternative package and what exactly the baseline package means	Terminology was updated to refer "Zero Investment", "Current Investment" and "Prospective Investment" scenarios as more relevant nomenclature may help to clarify the mechanism; descriptions of each scenario were

			also added to the template narrative report for further clarity.
	44.	It is not clear that the users have to run the model to get the results (Change from Interventions to Results).	Home page added to explain the layout of the tool
	45.	More detailed definition is needed to distinguish between interventions aimed at the whole smoking population and interventions targeted at smokers who are making quit attempts (Interventions).	The Demographics section of the tool has been re-worked to make the two types of intervention explicit and a further explanation was added to the template narrative report
	46.	Uncertain and unclear how to add the investment	The home page was added which may enhance the clarity of adding new investments
Aesthetic and minimalist design	47.	Sometimes it is not clear to which absolute number a figure in percent relates to	Control tips were added to help provide clarity
Help users recognize, diagnose, and recover from errors	48.	Some runtime errors appeared, tool crashed several times, especially when some estimates in the results section were changed, added a new intervention, selected a region, or generating the narrative report	Coding errors were resolved in subsequent releases.
	49.	No solutions were provided in the error alerts	The tool was still in development and the error alerts were due to system errors; in-tool warnings have been added and the system errors have been resolved.
Help and documentation	50.	Help buttons didn't provide any information	The tool was still in development at the time of user testing and informative help buttons were added
	51.	Need of a glossary to understand the tool faster and better, lack of some definitions, e.g. quasi-societal, reach, and uptake	Glossary will be provided in the supplementary materials; in the Web tool, the manual will electronically linked to the tool and guidance videos will also be provided
	52.	There is not enough help available in the tool. It would be good if the definitions of the interventions, the calculations of the cost of interventions, and the source of the data appear.	Informative menus were added in subsequent releases.

Supplementary files

Supplementary file 1: Video

Supplementary file 2: Test scenario

Supplementary file 3: Questionnaire

Supplementary file 4: Usability violations of the ROI tool detected from the thinking aloud procedure and heuristic evaluation