The Norwegian public’s ability to assess treatment claims: results of a cross-sectional study of critical health literacy [version 2; peer review: 1 approved, 2 approved with reservations]

Astrid Dahlgren, Kjetil Furuseth-Olsen, Christopher James Rose, Andrew David Oxman

Centre for Informed Health Choices, Norwegian Institute of Public Health, Postboks 222 Skøyen, Oslo, 0213, Norway

Abstract

Background: Few studies have evaluated the ability of the general public to assess the trustworthiness of claims about the effects of healthcare. For the most part, those studies have used self-reported measures of critical health literacy. Methods: We mailed 4500 invitations to Norwegian adults. Respondents were randomly assigned to one of four online questionnaires that included multiple-choice questions that test understanding of Key Concepts people need to understand to assess healthcare claims. They also included questions about intended behaviours and self-efficacy. One of the four questionnaires was identical to one previously used in two randomised trials of educational interventions in Uganda, facilitating comparisons to Ugandan children, parents, and teachers. We adjusted the results using demographic data to reflect the population. Results: A total of 771 people responded. The adjusted proportion of Norwegian adults who answered correctly was > 50% for 17 of the 30 Key Concepts. On the other hand, less than half answered correctly for 13 concepts. The results for Norwegian adults were better than the results for Ugandan children in the intervention arm of the trial and parents, and similar to those of Ugandan teachers in the intervention arm of the trial. Based on self-report, most Norwegians are likely to find out the basis of treatment claims, but few consider it easy to assess whether claims are based on research and to assess the trustworthiness of research. Conclusions: Norwegian adults do not understand many concepts that are essential for assessing healthcare claims and making informed choices. Future interventions should be tailored to address Key Concepts for which there appears to be a lack of understanding.
Introduction

Enabling people to make informed decisions about healthcare by improving their ability to think critically about such claims is an important public health initiative. Health literacy has been defined in many ways. Commonly this includes functional, interactional, and critical health literacy skills. Health literacy research efforts in Norway and elsewhere have mainly been directed at providing reliable health information and education in functional literacy, focusing on improving understanding of medical terminology and self-management of health conditions. Very little research has targeted critical health literacy, focused on enabling members of the public to make informed decisions about treatment choices, such training has rarely been offered.

Over the past decade, there has been increasing interest in enabling patients and the public to think critically about healthcare. For example, in Norway three websites have been developed that aim to empower patients and the public to assess the trustworthiness of health claims. Critical thinking has also been the focus of popular television shows that target common claims about treatment effects and illustrate how such claims can be tested using rigorous study designs. In collaboration with the Norwegian national television network (NRK), we carried out randomised trials with the aim of educating the public about the need for fair comparison of treatments. More recently, we have sought to improve critical thinking in school children through the Informed Health Choices (IHC) project, an international collaboration of researchers in Uganda, Rwanda, Kenya, the UK, and Norway. We developed an educational intervention to teach primary school children to assess healthcare claims and an educational podcast for their parents. Both interventions were shown to be effective in randomized trials in Uganda.

Few studies have been done to evaluate this ability in the general public to assess the trustworthiness of claims about the effects of healthcare and to make informed health choices. Most such studies have relied on self-assessment of critical health literacy skills. One such project is the HLS-EU survey, mapping health literacy skills in several European countries.

The Claim Evaluation Tools item bank

The Claim Evaluation Tools item bank was originally developed in English for use in the two randomised trials described above. Since the item bank was first created items have been translated and validated in several settings and languages, including Uganda (in Luganda and English), Mexico, and China.

Our starting point for developing the educational interventions and the Claim Evaluation Tools item bank was a list of Key Concepts that people need to understand and apply to assess claims about treatment effects and make informed health choices (hereafter referred to as Key Concepts) (Table 1). We define “treatment” broadly to include any intervention (action) intended to improve health, including preventive, therapeutic and rehabilitative interventions, and public health or health system interventions. The list of concepts provides a framework for researchers, teachers and others to develop interventions and measure people’s ability to assess treatment claims and make informed choices. We review and amend the list yearly.

The item bank is revised according to the Key Concepts list yearly. When we conducted this study, it included four multiple-choice questions (MCQs) for 32 of the Key Concepts shown in Table 1. The item bank also includes questions that assess intended behaviours and self-efficacy. The item bank is an open-access resource. Teachers and researchers can select items for questionnaires tailored to their specific needs.

To our knowledge only one survey has attempted to measure the ability to understand and apply any of the IHC Key Concepts in a representative sample of Norwegian adults. That study only addressed four of the Key Concepts. The purpose of this study was to map the ability of Norwegian adults to assess treatment claims and make informed health choices, using MCQs from the Claim Evaluation Tools item bank. The findings can be used to inform the development of learning resources and communication of information to patients and the public, and for international comparisons. We also wanted to take the opportunity to compare the findings of this study with the results of our previous Ugandan study. Despite any differences that may exist between these two contexts, it is our experience from previous projects that the challenges associated with assessing claims in everyday life in Uganda and Norway are very similar. However, knowledge of specific Key Concepts and overall understanding may differ. This study will consequently provide us with comparative information of understanding of the concepts in Uganda and Norway, but also explore any differences in understanding for specific Key Concepts. Furthermore, the study in Uganda included children and thus offered us the opportunity to compare our results from the adult population in Norway with those of children in Uganda.

Objective

To map the ability of Norwegian adults to assess treatment claims and make informed choices, and to compare these results to the findings of two studies in Uganda.
Table 1. The Informed Health Choices Key Concepts for assessing claims about treatment effects and making well informed treatment choices.

| 1. Claims | 2. Comparisons | 3. Choices |
|-----------|----------------|------------|
| Claims about effects that are not supported by evidence from fair comparisons are not necessarily wrong, but there is an insufficient basis for believing them. | Studies should make fair comparisons, designed to minimize the risk of systematic errors (biases) and random errors (the play of chance). | What to do depends on judgements about a problem, the relevance of the evidence available, and the balance of expected benefits, harms, and costs. |
| **1.1 It should not be assumed that treatments are safe or effective - or that they are not.** | **2.1 Comparisons of treatments should be fair.** | **3.1 Problems and options should be clear.** |
| a) Treatments can cause harms as well as benefits. | a) Comparison groups should be as similar as possible. | a) Be clear about what the problem or goal is and what the options are. |
| b) Large, dramatic effects are rare. | b) Indirect comparisons of treatments across different studies can be misleading. | **3.2 Evidence should be relevant.** |
| c) It is rarely possible to be certain about the effects of treatments. | c) The people being compared should be cared for similarly apart from the treatments being studied. | a) Attention should focus on all important effects of treatments, and not surrogate outcomes. |
| **1.2 Seemingly logical assumptions are not a sufficient basis for claims.** | d) If possible, people should not know which of the treatments being compared they are receiving. | b) There should not be important differences between the people in studies and those to whom the evidence will be applied. |
| a) Treatment may not be needed. | e) Outcomes should be assessed in the same way in all the groups being compared. | c) The treatments compared should be similar to those of interest. |
| b) Beliefs alone about how treatments work are not reliable predictors of the presence or size of effects. | f) Outcomes should be assessed using methods that have been shown to be reliable. | d) There should not be important differences between the circumstances in which the treatments were compared and those of interest. |
| c) Assumptions that fair comparisons are not relevant can be misleading. | g) It is important to assess outcomes in all (or nearly all) the people or subjects in a study. | **3.3 Expected advantages should outweigh expected disadvantages.** |
| d) An outcome may be associated with a treatment but not caused by it. | h) People's outcomes should be counted in the group to which they were allocated. | a) Weigh the benefits and savings against the harms and costs of acting or not. |
| e) More data is not necessarily better data. | **2.2 Syntheses of studies need to be reliable.** | b) Consider the baseline risk or the severity of the symptoms when estimating the size of expected effects. |
| f) Identifying effects of treatments depends on making comparisons. | a) Reviews of studies comparing treatments should use systematic methods. | c) Consider how important each advantage and disadvantage is when weighing the pros and cons. |
| g) The results of one study considered in isolation can be misleading. | b) Failure to consider unpublished results of fair comparisons may result in estimates of effects that are misleading. | d) Consider how certain you can be about each advantage and disadvantage. |
| h) Widely used treatments or those that have been used for decades are not necessarily beneficial or safe. | c) Treatment claims based on models may be sensitive to underlying assumptions. | e) Important uncertainties about the effects of treatments should be reduced by further fair comparisons. |
| i) Treatments that are new or technologically impressive may not be better than available alternatives. | **2.3 Descriptions should clearly reflect the size of effects and the risk of being misled by the play of chance.** | **2.3 Descriptions should clearly reflect the size of effects and the risk of being misled by the play of chance.** |
| j) Increasing the amount of a treatment does not necessarily increase its benefits and may cause harm. | a) Verbal descriptions of the size of effects alone can be misleading. | a) Weigh the benefits and savings against the harms and costs of acting or not. |
| k) Earlier detection of 'disease' is not necessarily better. | b) Relative effects of treatments alone can be misleading. | b) Consider the baseline risk or the severity of the symptoms when estimating the size of expected effects. |
| l) It is rarely possible to know in advance who will benefit, who will not, and who will be harmed by using a treatment. | c) Average differences between treatments can be misleading. | c) Consider how important each advantage and disadvantage is when weighing the pros and cons. |
| **1.3 Trust in a source alone is not a sufficient basis for believing a claim.** | d) Small studies may be misleading. | d) Consider how certain you can be about each advantage and disadvantage. |
| a) Your existing beliefs may be wrong. | e) Results for a selected group of people within a study can be misleading. | e) Important uncertainties about the effects of treatments should be reduced by further fair comparisons. |
| b) Competing interests may result in misleading claims. | f) Confidence intervals should be reported for estimates of effects. | **2.3 Descriptions should clearly reflect the size of effects and the risk of being misled by the play of chance.** |
| c) Personal experiences or anecdotes alone are an unreliable basis for most claims. | g) Deeming results to be “statistically significant” or “not statistically significant” can be misleading. | a) Verbal descriptions of the size of effects alone can be misleading. |
| d) Opinions alone are not a reliable basis for claims. | h) Lack of evidence of a difference is not the same as evidence of “no difference”. | b) Relative effects of treatments alone can be misleading. |
| e) Peer review and publication by a journal do not guarantee that comparisons have been fair. | **2.3 Descriptions should clearly reflect the size of effects and the risk of being misled by the play of chance.** | c) Average differences between treatments can be misleading. |
| **2.2 Syntheses of studies need to be reliable.** | a) Verbal descriptions of the size of effects alone can be misleading. | d) Small studies may be misleading. |
| a) Reviews of studies comparing treatments should use systematic methods. | b) Relative effects of treatments alone can be misleading. | e) Results for a selected group of people within a study can be misleading. |
| b) Failure to consider unpublished results of fair comparisons may result in estimates of effects that are misleading. | c) Average differences between treatments can be misleading. | f) Confidence intervals should be reported for estimates of effects. |
| c) Treatment claims based on models may be sensitive to underlying assumptions. | d) Small studies may be misleading. | g) Deeming results to be “statistically significant” or “not statistically significant” can be misleading. |
| **2.3 Descriptions should clearly reflect the size of effects and the risk of being misled by the play of chance.** | e) Results for a selected group of people within a study can be misleading. | h) Lack of evidence of a difference is not the same as evidence of “no difference”. |
Methods
Ethical statement
This study was considered for ethical approval by the Norwegian Institute of Public Health. This project was considered to not require a full ethical review (reference 18/11854), because no sensitive data was included. All information was collected through Nettskjema (a web-based survey system), ascertaining a high level of data security and safety.

The population was drawn from the National registry who provided us with a CD-ROM with respondents’ addresses. This CD was destroyed once data collection was terminated.

All respondents were given information about the purpose of the study and how the results would be managed and presented. To gain access to the questionnaire, each participant had to provide written consent electronically. The questionnaire was anonymous and once submitted, the information could not be traced back to the respondent.

Development of the questionnaires
Translation and adaption of the item bank to Norwegian.
The Claim Evaluation Tools item bank was originally developed in English for use in Uganda. For this study, a language specialist (KFO) translated all items to Norwegian. Three researchers trained in evidence-based medicine and epidemiology reviewed this translation (AD, ADO, and Atle Fretheim). The questions were not modified other than changing some of the names of people in the scenarios to more familiar names in Norwegian. For example, “Dr. Acheng” was changed to “Dr. Anker” (a more familiar name in Norwegian). The original English version was developed using plain language and this was also an important concern in the Norwegian translation.

Description of questions included in the item bank. All MCQs in the Claim Evaluation Tools item bank include a scenario with a treatment claim, and response options, with one answer being the “best” (correct) and the remaining options considered “worse” (incorrect). An example of a multiple-choice question is shown in Box 1.

In addition to knowledge questions, the item bank also includes two sets of questions scored on likert scales evaluating intended behaviour and self-efficacy associated with assessing claims.

Design of the questionnaires. For this study we decided to include two MCQs for all of the 32 Key Concepts. We also included the question sets for intended behavior and self-efficacy. Considering the large number of questions, we decided to split the questions across four questionnaires (see Additional file 1 and Additional file 5, Extended data)33,34.

We considered four of the 32 Key Concepts to be amongst the most important Key Concepts that can be understood and used by children, as well as adults. A set of 8 MCQs addressing these four Key Concepts were therefore included in all four questionnaires:

- An outcome may be associated with a treatment but not caused by it.
- The results of one study considered in isolation can be misleading.
- Widely used treatments or those that have been used for decades are not necessarily beneficial or safe.
- Comparison groups should be as similar as possible.

In order to compare the findings of the Norwegian study with the Uganda study, the first questionnaire was designed so that it was identical to the questionnaire used in Uganda and included 27 questions33,34. The second questionnaire included the remaining items that address concepts in the first (Claims) category as well as items from the second (Comparisons) category (24 questions) (Table 1). The third questionnaire included items from the third (Choices) category. This questionnaire also included three items evaluating intended behaviours regarding assessing treatment claims and agreeing to participate in a study evaluating treatments, and four items regarding self-efficacy (21 questions in total). The fourth questionnaire included MCQs addressing Key Concepts that we judged to be more difficult, mostly belonging to the Comparisons category (25 questions).

All four questionnaires also included the same set of questions regarding the participant’s sex, level of education, health professional background, training in research methods, and prior participation in research.

Recruitment of participants and administration of the questionnaires
Based on evidence from two systematic reviews that evaluated strategies to improve participation in research35,36, we developed attractive postcards personally addressed to each potential participant inviting people to take part in the study. The postcard is shown in Figure 1. The postcard included a short description of the purpose of the study as well as a URL (“link”) to one of the four online questionnaires. Participants received full information about the study when they accessed the

Box 1. Example of a multiple-choice question
Judith wants smoother skin. The younger girls in her school have smoother skin than the older girls. Judith thinks this is because the younger girls use cream on their skin to make the skin smoother.

Question: Based on this link between using cream and smooth skin, is Judith correct?
Options:
A) It is not possible to say. It depends on how many younger and older girls there are
B) It is not possible to say. There might be other differences between the younger and older girls
C) Yes, because the younger girls use cream on their skin and they have smoother skin
D) No, Judith should try using the cream herself to see if it works for her
website and were asked to provide informed consent. Evidence also suggests greater participation rates can be expected if people are informed how the research may benefit them. This was stated on the postcard and the website provided information about how the study results can be accessed.

In January and February 2019, we mailed postcards to a representative sample of 4500 adults (≥18 years) living in Norway. The sample was provided to us by the Norwegian National Registry, considering level of education, sex and geographical spread\textsuperscript{41}. The questionnaires were administrated electronically using Nettskjema, a service provided by University of Oslo\textsuperscript{42}. In addition to the first postcard, one reminder postcard was sent out to each person.

**Rasch analysis**

The questionnaires were evaluated for their psychometric properties as part of this study. Although some MCQs were tested in Norway as part of a previous pilot, this is the first full-scale psychometric evaluation of the complete available battery of the Claim Evaluation Tools Item bank\textsuperscript{30–32}.

Rasch analysis is a dynamic and practical approach to address important measurement issues required for validating an outcome\textsuperscript{39–41}. In this study we followed the fundamental steps of Rasch analysis including testing for internal construct validity (multidimensionality), invariance of the items (Item-Person Interaction), and item bias (differential item functioning), as well as testing for reliability\textsuperscript{40,41}. Rasch analysis can be used for both dichotomous and polytomous data\textsuperscript{40,42,43}. Raw data were exported from the electronic data collection service (Nettskjema) as Excel-files and entered into RUMM2030 for Rasch analysis. Results can be replicated using the open-access software WINSTEPS\textsuperscript{44}. Each questionnaire set was evaluated separately. Questionnaire 3 included also polytomous questions assessing intended behaviour and self-efficacy questions. Therefore, for questionnaire 3 specifically, additional analysis was done for each item block separately (MCQ-item block and intended behaviour and self-efficacy questions-block respectively).

**Survey analysis**

Each MCQ was scored as correct or incorrect. The Key Concepts for which there were two MCQs were scored as “understood” if a participant answered both MCQs correctly. For questions about intended behaviours and self-efficacy, we dichotomised the responses in the analysis, for example as likely (very likely or likely) vs not likely (very unlikely, unlikely, or don’t know). As previously mentioned, the Key Concepts are reviewed and revised annually. Over the duration of this study, two Key Concepts were revised into the new Key Concept 1.1a. Considering that the MCQs were no longer appropriately covering this concept, these MCQs were therefore taken out of the analysis.

Based on a previous survey we conducted\textsuperscript{45}, we anticipated that women and respondents with higher education levels would be more likely to respond. To address such non-random non-response, we used iterative post-stratification to match marginal distributions of sex and educational attainment level of the sample to the Norwegian population\textsuperscript{46}. We also adjusted for region of residence. Due to an error, we did not collect data on participants’ age and could not use it for post-stratification, as planned.

We obtained Eurostat data on the marginal distributions of sex and region of residence for all Norwegians, and educational attainment level for Norwegians aged 15 to 64 years\textsuperscript{47}. Participants reported their county of residence. We mapped counties to the corresponding Nomenclature of Territorial Units for Statistics (NUTS 2) regions of Norway used by Eurostat\textsuperscript{48}. Participants reported the highest level of education they attained. We mapped these to the corresponding International Standard Classification of Education (ISCED) 2011 categories used by Eurostat (levels 0–2, 3–4, and 5–8). We used multiple imputation with chained equations to account for the uncertainty introduced by missing values of the post-stratification variables sex, region of residence, and educational level\textsuperscript{49}, and iteratively post-stratified each imputed data set. Based on these adjustments, we estimated the percentage of the Norwegian population that understands each key concept, and that responded positively for each question about intended behaviours and self-efficacy. We used the R packages tidyverse, mice, mitools, and survey.

We present summaries of the results for participants and the post-stratified population estimates for understanding of the 30 Key Concepts, and the intended behaviours and self-efficacy questions. We quantified the uncertainty of our estimates using 95% confidence intervals and protected the family-wise coverage probability of the confidence intervals for the four Key Concepts included in all questionnaires via Bonferroni correction (i.e., we report a 98.75% CI for each of those concepts). For each Key Concept, we calculated the likelihood
of answering both questions (or the one question for two Key Concepts) correctly if participants randomly guessed the answer. These probabilities vary between six and 25%, depending on the number of MCQs and the number of response options (between two and four) for each MCQ.

We compared Norwegian and Ugandan adults’ understanding of the Key Concepts to that of Ugandan children in the intervention arm of our randomized trial. We compared mean test scores, and probabilities of achieving passing and mastery scores, using estimates from one year after the IHC primary school intervention for the Ugandan children and their teachers\(^\text{30,51}\) and estimates from the control group for parents in the IHC podcast trial, one year after parents in the intervention group listened to the podcast\(^\text{15,52}\). We used predetermined thresholds of at least 13 of 24 questions answered correctly for a passing score and at least 20 of 24 for mastery.

We estimated mean scores and odds ratios using generalized linear mixed-effects models (GLMMs; normal errors and identity link for mean scores, binomial errors and logit link for passing and mastery) using the lme4 R package. In the trials, children and teachers were randomized in clusters (schools), while parents were individually randomized. We modelled this clustering structure as a random intercept for each randomized unit. We did not adjust for covariates such as those used for stratified sampling in the Ugandan trials because those variables are not defined for all samples (e.g., school ownership was used in the trial on Ugandan children, but there is no analogous concept for Norwegian adults). No data were missing. It is not possible to use the lme4 package to apply post-stratification weights for the Norwegians.

Exploratory analyses
We conducted exploratory analyses to investigate associations between understanding each Key Concept and the demographic covariates sex, research training, research participation, education level, and health professional background. Based on findings from our previous survey\(^\text{12}\), we hypothesised that better understanding of the Key Concepts would be associated with having a research background or a higher education level; that there would be little difference between health professionals and others; and that there would be little difference between women and men. We used generalized linear models (GLMs; quasibinomial errors and logit link) as before, modelled the covariates as categorical variables, and used multiple imputation and post-stratification as in the main analysis.

We used data from the first questionnaire to perform exploratory analyses to investigate how Norwegians’ mean scores and achievement of passing and mastery scores are associated with the demographic covariates sex, research training, research participation, education, and health professional background. We used GLMs (normal errors and identity link for mean score; quasibinomial errors and logit link for passing and mastery) to model the outcomes in terms of the covariates, which were modelled as before. Multiple imputation and post-stratification were used as in the main analysis. The variables included in imputation were the post-stratification variables (sex, region of residence, and educational attainment); demographic variables that coded for whether participants had research training or a health professional background, and whether they had been a research participant; and mean score. We did not include passing and mastery in imputation, as they can be calculated from the mean score.

Sample size calculation
We performed three power analyses. The first two analysed power to estimate differences, of at least 5% from random guessing, in the proportion of the Norwegian population that understands the concepts. The first analysis focused on the four Key Concepts that would be included on all four questionnaires while protecting the 95% coverage of the four confidence intervals as a family, and the second analysis focused on a single prototypical Key Concept that would be included on only one questionnaire. We assumed that each concept would be probed by two MCQs, each having four options, such that we would expect that 6.25% of Norwegians would appear to “understand” a concept if all participants guessed at random. We further calculated confidence interval widths, aiming to estimate proportions with precision no worse than ±0.05 (i.e., ±5%) for those four Key Concepts and no worse than ±0.08 (i.e., ±8%) for the other Key Concepts. The third analysis estimated power to detect a difference of at least 10% in the mean score (proportion of correct answers) between Norwegians and Ugandan children, parents, and teachers. Finally, drawing on experience with our previous survey, we explored a range of response rate scenarios.

These analyses suggested that sending a total of 4500 postcards would be sufficient to achieve a margin of error no greater than ±5% and ±8% for the first two analyses, and approximately 97% power to detect differences greater than 10% in the mean score compared to results from the Ugandan trials.

Results
Participant characteristics
Table 2 shows participant characteristics across the four questionnaires\(^\text{31}\). Of the 771 respondents, 21 (2.7%) did not provide data on at least one of the covariates sex, research training, research participation, education, health professional background, and county of residence.

The validity and reliability of the questionnaires
The log describing the Rasch analysis is described in more detail in Additional file 2 (Extended data)\(^\text{4}\). Overall, the four questionnaires showed acceptable fit to the Rasch model. Three out of four questionnaires were found to be unidimensional. The questionnaires including intended behaviour and self-efficacy items suggested that this sub-scale may include more than one dimension. Separate analyses of the two item-sets in Questionnaire 3 (Key Concept MCQs and intended behaviour and self-efficacy items) suggest that the MCQ sub-test works very well, with no apparent validity issues.
Table 2. Participant characteristics across the four questionnaires.

| Category                | Quiz 1 | Quiz 2 | Quiz 3 | Quiz 4 | P-value  | Overall |
|-------------------------|--------|--------|--------|--------|----------|---------|
| All                     |        |        |        |        |          | 771 (100.0%) |
| Sex                     |        |        |        |        |          |         |
| Female                  | 103 (13.4%) | 102 (13.2%) | 102 (13.2%) | 94 (12.2%) | 0.63355 | 401 (52.0%) |
| Male                    | 107 (13.9%) | 109 (14.1%) | 70 (9.1%) | 83 (10.8%) | 0.56449 | 369 (47.9%) |
| Missing                 | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 1 (0.1%) | 0.45466 | 1 (0.1%) |
| Education               |        |        |        |        |          |         |
| Primary                 | 12 (1.6%) | 16 (2.1%) | 5 (0.6%) | 10 (1.3%) | 0.30396 | 43 (5.6%) |
| Secondary               | 53 (6.9%) | 53 (6.9%) | 42 (5.4%) | 49 (6.4%) | 0.95886 | 197 (25.6%) |
| Tertiary (1-2 years)    | 23 (3.0%) | 30 (3.9%) | 21 (2.7%) | 33 (4.3%) | 0.29255 | 107 (13.9%) |
| Tertiary (3-5 years)    | 71 (9.2%) | 69 (8.9%) | 58 (7.5%) | 51 (6.6%) | 0.86021 | 249 (32.3%) |
| Masters degree          | 39 (5.1%) | 36 (4.7%) | 41 (5.3%) | 30 (3.9%) | 0.49704 | 146 (18.9%) |
| PhD                     | 10 (1.3%) | 5 (0.6%) | 5 (0.6%) | 3 (0.4%) | 0.38534 | 23 (3.0%) |
| Missing                 | 2 (0.3%) | 2 (0.3%) | 0 (0.0%) | 2 (0.3%) | 0.67033 | 6 (0.8%) |
| Research training       |        |        |        |        |          |         |
| No                      | 126 (16.3%) | 145 (18.8%) | 105 (13.6%) | 126 (16.3%) | 0.66839 | 502 (65.1%) |
| Yes                     | 84 (10.9%) | 66 (8.6%) | 65 (8.4%) | 52 (6.7%) | 0.35023 | 267 (34.6%) |
| Missing                 | 0 (0.0%) | 0 (0.0%) | 2 (0.3%) | 0 (0.0%) | 0.050443 | 2 (0.3%) |
| Research participant    |        |        |        |        |          |         |
| No                      | 146 (18.9%) | 157 (20.4%) | 121 (15.7%) | 126 (16.3%) | 0.97134 | 550 (71.3%) |
| Yes                     | 63 (8.2%) | 52 (6.7%) | 50 (6.5%) | 51 (6.6%) | 0.80132 | 216 (28.0%) |
| Missing                 | 1 (0.1%) | 2 (0.3%) | 1 (0.1%) | 1 (0.1%) | 0.67033 | 5 (0.6%) |
| Medical education       |        |        |        |        |          |         |
| No                      | 175 (22.7%) | 168 (21.8%) | 133 (17.3%) | 147 (19.1%) | 0.96116 | 623 (80.8%) |
| Yes                     | 35 (4.5%) | 43 (5.6%) | 37 (4.8%) | 28 (3.6%) | 0.57834 | 143 (18.5%) |
| Missing                 | 0 (0.0%) | 0 (0.0%) | 2 (0.3%) | 0 (0.0%) | 0.061528 | 5 (0.6%) |
| County                  |        |        |        |        |          |         |
| Akershus                | 27 (3.5%) | 28 (3.6%) | 16 (2.1%) | 21 (2.7%) | 0.72691 | 92 (11.9%) |
| Aust-Agder              | 4 (0.5%) | 2 (0.3%) | 2 (0.3%) | 2 (0.3%) | 0.87829 | 10 (1.3%) |
| Buskerud                | 7 (0.9%) | 7 (0.9%) | 14 (1.8%) | 8 (1.0%) | 0.162 | 36 (4.7%) |
| Finnmark                | 5 (0.6%) | 2 (0.3%) | 3 (0.4%) | 1 (0.1%) | 0.48594 | 11 (1.4%) |
| Hedmark                 | 2 (0.3%) | 3 (0.4%) | 4 (0.5%) | 3 (0.4%) | 0.76607 | 12 (1.6%) |
| Hordaland               | 16 (2.1%) | 29 (3.8%) | 9 (1.2%) | 20 (2.6%) | 0.048787 | 74 (9.6%) |
| Møre og Romsdal         | 8 (1.0%) | 3 (0.4%) | 6 (0.8%) | 9 (1.2%) | 0.24526 | 26 (3.4%) |
| Nordland                | 9 (1.2%) | 9 (1.2%) | 5 (0.6%) | 6 (0.8%) | 0.89401 | 29 (3.8%) |
| Oppland                 | 3 (0.4%) | 9 (1.2%) | 5 (0.6%) | 5 (0.6%) | 0.40473 | 22 (2.9%) |
| Oslo                    | 38 (4.9%) | 29 (3.8%) | 38 (4.9%) | 26 (3.4%) | 0.27531 | 131 (17.0%) |
| Østfold                  | 11 (1.4%) | 9 (1.2%) | 14 (1.8%) | 7 (0.9%) | 0.37221 | 41 (5.3%) |
| Rogaland                | 25 (3.2%) | 35 (4.5%) | 18 (2.3%) | 20 (2.6%) | 0.40053 | 98 (12.7%) |
| Sogn og Fjordane        | 3 (0.4%) | 2 (0.3%) | 4 (0.5%) | 1 (0.1%) | 0.53125 | 10 (1.3%) |
| Telemark                | 7 (0.9%) | 5 (0.6%) | 4 (0.5%) | 9 (1.2%) | 0.47175 | 25 (3.2%) |
| Troms                    | 6 (0.8%) | 8 (1.0%) | 3 (0.4%) | 7 (0.9%) | 0.64403 | 24 (3.1%) |
| Trøndelag               | 19 (2.5%) | 14 (1.8%) | 13 (1.7%) | 15 (1.9%) | 0.84835 | 61 (7.9%) |
| Vest-Agder               | 6 (0.8%) | 10 (1.3%) | 6 (0.8%) | 11 (1.4%) | 0.4576 | 33 (4.3%) |
| Vestfold                | 12 (1.6%) | 6 (0.8%) | 7 (0.9%) | 6 (0.8%) | 0.53705 | 31 (4.0%) |
| Missing                 | 2 (0.3%) | 1 (0.1%) | 1 (0.1%) | 1 (0.1%) | 0.67033 | 5 (0.6%) |
Overall, the analysis of the sub-test consisting of items evaluating intended behaviour and self-efficacy items also shows satisfactory fit. However, two of the intended behaviour items show disordered thresholds. We did not observe any important local dependency in Questionnaire 2, 3 and 4. However, in Questionnaire 1, twelve correlations were observed. Two questionnaires were underpowered for this analysis (Questionnaire 1 and 4). However, based on the available data, most MCQs in all questionnaires showed acceptable fit to the ICC-curve, and no important differential item functioning by gender was identified in any of the questionnaires. Five MCQs showed evidence of under discriminating with one ability class deviating: two MCQs in Questionnaire 1 (q11 and q19.1), two MCQs in Questionnaire 2 (q3 and q18.3), and one MCQ in Questionnaire 4. In the two questionnaires that were adequately powered, Cronbach’s alpha was 0.6 and 0.7, respectively.

Norwegians’ understanding of the Key Concepts
The mean width of the confidence intervals for the four concepts common to all questionnaires was ±4.6% (range ±3.6% to ±5.9%). One concept (An outcome may be associated with a treatment but not caused by it) was not estimated to the desired precision. The mean width of the other confidence intervals was ±7.92% (range ±3.37% to ±11.9%; 15 of the 30 concepts had confidence intervals wider than ±8%).

Figure 2 shows estimates of the percentage of Norwegian adults who understand each Key Concept. The adjusted proportion of Norwegian adults who answered correctly five of the Key Concepts was > 80% and more than 50% understand 17 concepts. On the other hand, less than half answered correctly for 13 concepts, and for seven of those concepts, the proportions of correct answers were no better than if people had randomly guessed.

Intended behaviours and self-efficacy
Intended behaviours and self-efficacy are presented in Figure 3. The results suggest that most Norwegian adults are likely to question the basis of treatment claims (75%; 95% CI 66–85%), find out if treatment claims are based on research evidence (70%; 95% CI 59–81%), and say “yes” if asked to participate in a study comparing treatments (67%; 95% CI 63–72%). However, only about one out of five Norwegian adults find it easy to assess the relevance of studies that compare treatments (21%; 95% CI 9–33%), assess whether a treatment claim is based on a study comparing treatments (18%; 95% CI 12–25%), find studies that compare treatments (18%; 95% CI 9–28%), or assess the trustworthiness of the results of studies that compare treatments (16%; 95% CI 6–26%).

Comparison of Norwegian and Ugandan adults to Ugandan children
The results for Norwegian and Ugandan adults are compared to results for Ugandan children in Table 3 to Table 5. The mean score for Norwegian adults who participated in the survey (Questionnaire 1) was 17% higher (95% CI 14–20%) than the mean score for Ugandan children in the intervention arm of the randomised trial of the IHC primary school resources one year after the intervention and similar to the mean score for Ugandan teachers in the intervention arm of the trial (Table 3)3. The scores for Ugandan teachers in the control group were similar to those of the children in the intervention group, and the scores of Ugandan parents in the control group of the IHC podcast trial34 were 16% lower (95% CI -19 to -13%).

Nearly all the Norwegian adults (209 out of 210) who participated in the survey and Ugandan teachers in the intervention group (77 out of 78) had a passing score (Table 4). This was 15% more than the Ugandan children in the intervention group. However, because only one Norwegian and one Ugandan teacher in the intervention group did not have a passing score, the estimated differences may be unreliable for those comparisons. The proportion of Ugandan teachers in the control group with a passing score (88%) was similar to that of the children (84%), and 47% fewer (95% CI 56% to 37%).

Three quarters (66%; 95% CI 56–74%) of the Norwegian adults who participated in the survey and 70% (95% CI 59–80%) of the Ugandan teachers in the intervention group had a mastery score (Table 5). Compared to Ugandan children in the intervention group, the Odds for achieving a mastery score by the Norwegian participants was 3.3 (95% CI 3.5–7.9%), whereas the Odds for achieving a mastery score was 0.24 (95% CI 0.15–0.4) for the Ugandan parents compared to the children.

Associations between participant characteristics and their responses
We did not find strong associations between gender, health professional background, or having participated in research and how well participants did (Table 6). Having 1–2 years of tertiary school education (ISCED levels 5–8) was associated with a mean score that was 11% higher (95% CI 4.0–19%), and a higher likelihood of having a mastery score (OR 3.5; 95% CI 1 to 12). Norwegian men were more likely than women to state that they are likely to find out the basis of treatment claims (OR 3.0; 95% CI 1.1 to 8.2), and more likely to find it easy to assess the trustworthiness of the results of studies that compare treatments (OR 3.8; 95% CI 1 to 14). People who have participated in research may be more likely to find it easy to find research based on studies that compare treatments (OR 5.7; 95% CI 1.9 to 17). People with secondary school education (ISCED levels 3–4) may be less likely than people with no more than primary school education (ISCED levels 0–2) to find it easy to assess the trustworthiness of results of studies that compare treatments (OR 0.057; 95% CI 0.0049 to 0.68). Besides these associations, we did not find evidence suggesting that gender, education level, health professional background, or prior participation in research were associated with intended behaviours or self-efficacy. Associations between participant characteristics and their understanding of specific Key Concepts, intended behaviour and self-efficacy are reported in Additional file 3 (Extended data)35.

Discussion
According to Statistics Norway, 24.7% of Norwegian adults have a primary school education, 41.7% secondary school education, 22.4% tertiary school education, 7.3% have a
master’s degree, and 0.7% of the population have a PhD. As we anticipated, participants in our study had a somewhat higher educational level than the general population\(^5\). In Norway, approximately 18% of the population (15–74) have an educational background in health or welfare\(^5\). This is comparable to participants in this study.

Participants in our survey had a good understanding of the 12 Key Concepts that were addressed in our randomised trial of the IHC primary school intervention\(^2\). Their understanding was comparable to Ugandan teachers in the intervention arm of our randomized trial of a primary school intervention, and better than that of Ugandan children in the intervention arm.

![Figure 2](image-url)  
**Figure 2.** Estimates of the percentage of Norwegian adults who understand each Key Concept.*

* Vertical lines indicate expected results if participants guessed at random. Understanding is based on both questions being answered correctly for each concept. † There was only one question for these concepts. ‡ Confidence intervals have been Bonferroni-corrected.
Table 4. The compared probability of passing in the Norwegian and Ugandan samples.

| Sample                     | Probability (95% CI) | Difference (95% CI) | Odds ratio (95% CI) | P value |
|----------------------------|----------------------|---------------------|---------------------|---------|
| Ugandan Children (Intervention) | 3160 / 3943         | 84% (81% to 88%)   | 0                   | 1       |
| Ugandan Parents (Control)    | 101 / 256           | 37% (28% to 47%)   | -47% (-56% to -37%)| 0.11 (0.073 to 0.17) | <0.0001 |
| Ugandan Teachers (Control)   | 50 / 59             | 88% (76% to 94%)   | 3.4% (-8.5% to 9.9%)| 1.3 (0.58 to 3) | 0.4984  |
| Ugandan Teachers (Podcast)   | 77 / 78             | 99% (94% to 100%)  | 15% (9.7% to 15%)   | 21 (3 to 150)  | 0.0023  |
| Norwegian Adults             | 209 / 210           | 100% (98% to 100%) | 15% (13% to 16%)    | 54 (7.5 to 400) | <0.0001 |

Table 3. The mean test scores of the Norwegian and Ugandan samples.

| Sample                      | n     | Mean score (95% CI) | Difference (95% CI) | P-value |
|-----------------------------|-------|---------------------|---------------------|---------|
| Ugandan Children (Intervention) | 3943  | 69% (67% to 71%)    | 0                   |         |
| Ugandan Parents (Control)   | 256   | 53% (48% to 57%)    | -16% (-19% to -13%) | <0.0001 |
| Ugandan Teachers (Control)  | 59    | 68% (61% to 75%)    | -0.59% (-5.6% to 4.4%) | 0.82    |
| Ugandan Teachers (Podcast)  | 78    | 86% (81% to 92%)    | 17% (14% to 21%)    | <0.0001 |
| Norwegian Adults            | 210   | 86% (81% to 91%)    | 17% (14% to 20%)    | <0.0001 |

Figure 3. The Norwegian population’s intended behaviours and self-efficacy.

We estimate that more than 80% of Norwegian adults understand these five concepts:

- Increasing the amount of a treatment does not necessarily increase its benefits and may cause harm.
- Competing interests may result in misleading claims.
- Personal experiences or anecdotes alone are an unreliable basis for most claims.
The people being compared should be cared for similarly apart from the treatments being studied.

Weigh the benefits and savings against the harms and costs of acting or not.

On the other hand, Norwegian adults appear to do no better than if they were to randomly guess the answers to questions about these seven Key Concepts:

- Beliefs alone about how treatments work are not reliable predictors of the presence or size of effects.
- Widely used treatments or those that have been used for decades are not necessarily beneficial or safe.
- Comparison groups should be as similar as possible.
- People’s outcomes should be counted in the group to which they were allocated.
- Results for a selected group of people within a study can be misleading.
- Deeming results to be “statistically significant” or “nonsignificant” can be misleading.

Considering that people who responded to our survey had a somewhat higher educational level than the general population, the ability of our respondents may be higher than in the general population.

Based on self-report, most Norwegians are likely to find out what the basis of a treatment claim is and to find out if a treatment claim is based on research. However, they do not consider it easy to assess the relevance of a claim, whether it is based on research, or its trustworthiness. They also do not consider it easy to find relevant studies.

Comparison with other studies

In the previous Norwegian study, four Key Concepts were evaluated. In both studies the majority responded correctly to the question “Weighing the benefits and harms”. However, in the present study, the respondents were less likely to respond correctly to “Relative effects of treatments can be misleading” (29% vs 64%) and “The use of p-values may be misleading” (19% vs 51.5%). In contrast, more responded correctly to “An outcome may be associate with a treatment but not caused by it” (64% vs 30%). It is difficult to interpret these differences, other than that these evaluations were done using different question sets. Consequently, one explanation may be difference in the questions’ difficulty level. The gender distribution in the studies were identical, however a higher percentage in the present study had at least one year of education beyond secondary school (68% vs 52%). We did not find gender or having a health professional background to be a good predictor of participants’ understanding of the Key Concepts in any of studies. This is consistent with other studies and findings that both health professionals and patients feel challenged finding, appraising, and applying relevant evidence for use in health decisions.

### Table 5. The compared probability of mastery in the Norwegians and Ugandan samples.

| Sample                          | Probability (95% CI) | Difference (95% CI) | Odds ratio (95% CI) | P value |
|---------------------------------|----------------------|---------------------|---------------------|---------|
| Ugandan Children (Intervention) | 1138 / 3943 (28.9%)  | 27% (22% to 32%)    | 0                   | 1       |
| Ugandan Parents (Control)      | 27 / 256 (10.5%)     | 8.1% (5.1% to 13%)  | -18% (-22% to -14%) | 0.24    |
| Ugandan Teachers (Control)     | 13 / 59 (22.0%)      | 19% (9.8% to 33%)   | -8% (-17% to 5.9%)  | 0.63    |
| Ugandan Teachers (Podcast)     | 53 / 78 (67.9%)      | 70% (59% to 80%)    | 44% (32% to 53%)    | 6.6     |
| Norwegian Adults               | 133 / 210 (63.3%)    | 66% (56% to 74%)    | 39% (29% to 48%)    | 5.3     |

### Table 6. Associations between demographic covariates and Norwegians’ mean scores and achievement of passing and mastery scores.

| Intercept | Male | Research participant | ISCED Levels 3-4 | ISCED Levels 5-8 | Medical education |
|-----------|------|----------------------|------------------|------------------|------------------|
| Mean score| 75.0% (68.0% to 82.0%) | -2.3% (-6.6% to 2.1%) | 0.6% (-5.5% to 6.7%) | 5.0% (-2.9% to 13.0%) | 11.0% (4.0% to 19.0%) | -0.4% (-5.2% to 4.4%) |
| Passing   | 9.8e+33 (6.9e+32 to 4.9e+34) | 6.7e-23 (6.4e-24 to 7.1e-22) | 3.8e-23 (5.9e-24 to 2.4e-22) | 2e+22 (2.2e+21 to 1.8e+23) | 2e+22 (3e+21 to 1.3e+23) | 0.28 (0.063 to 1.2) |
| Mastery   | 0.53 (0.17 to 1.7) | 0.81 (0.36 to 1.8) | 1.7 (0.54 to 5.1) | 0.89 (0.26 to 3.1) | 3.5 (1 to 12) | 1.3 (0.42 to 4.1) |

ISCED, International Standard Classification of Education.
There are few other studies to which we can compare these results. The European health literacy survey included all domains of health literacy, responses were given as self-assessments, and results are reported as overall scores. Consequently, comparison with our study is difficult. They concluded that a little more than one tenth (12%) of those surveyed had insufficient health literacy and almost one half (47%) had limited (insufficient or problematic) health literacy. Both our survey and the European health literacy survey examined associations with education and, unsurprisingly, found that people with higher education have higher health literacy skills. However, the European health literacy survey reported that this result varies by country.

**Strengths and limitations**

The strength of this study is that we provide new evidence on people’s ability to assess treatment claims using multiple-choice questions. To our knowledge, few such objective surveys have been conducted in Europe. Our work complements studies in which participants assessed their own abilities.

Many conceptualisations of critical thinking exist. Not all are based on explicit criteria, and unlike the Key Concepts used in this survey, few if any are subject to annual revision. The questionnaires we administered were based on a framework that has been developed from methodological literature and input from multiple disciplines and people with methodological expertise.

We validated the questionnaires using robust methods. Although the results of the Rasch analysis are promising (Additional file 2, Extended data), it suggests the potential for improvements. Across all four questionnaires, we found that only five MCQs warranted improvement. However, Rasch analysis of two of the questionnaires was underpowered, so the validity and reliability of these should be assessed again in future studies.

Our analysis of the questionnaire including intended behaviours and self-efficacy suggests that intended behaviour and self-efficacy items measure two different dimensions. This might be because intended behaviour is more complex and dependent on self-efficacy, values, knowledge, and other factors. Except for Questionnaire 1, we did not identify any important dependencies between items.

We did not collect information on age for this study, and thus the association between age and ability to assess treatment claims should be explored in future studies. In our previous study conducted in Norway, younger respondents had a higher proportion of correct responses and higher total scores. This may suggest that there may have been improvements over time in the ability of both health professionals and patients to assess treatment claims.

Another possible limitation of our study is the low response rate, which we anticipated. Our sample was similar to the general population in terms of the percentage of health professionals, but people with a higher education were over-represented among respondents. We addressed this issue using prespecified post-stratification in the survey analyses but did not address this in the comparisons to Ugandans. The results of those analyses therefore cannot be assumed to apply to the general population of Norwegian adults.

**Implications and future research**

The results of this study can inform the development and evaluation of educational interventions that address Key Concepts that Norwegians appear to poorly understand. Up to now, few such interventions have been evaluated. There is a need to evaluate interventions for health professionals as well as for the general public to help ensure that they can think critically about treatment claims and choices. The results also can inform the development and evaluation or strategies for improving communication of information about the effects of treatments by researchers, health professionals, and others.

Studies like this one in other countries would help to map similarities and differences in people’s abilities across different countries and settings. Such information could help to determine the extent to which interventions should be tailored to address different Key Concepts for different populations.

Health professionals and others who communicate health information should be aware that patients may not be able to think critically about treatment claims and may therefore struggle to process information necessary to making informed decisions.

Our Rasch analysis suggests that the MCQs can be used in Norway. Most MCQs performed well, and this evaluation is the first step in developing a calibrated item bank that can be used for Computer Assisted Testing. The results of our Rasch analysis can be used to improve the multiple-choice questions that did not perform well.

**Conclusion**

Norwegian adults’ understanding of Key Concepts for assessing treatment claims and making choices varies from over 80% who understand five Key Concepts to 20% or less who understand seven Key Concepts. We did not find strong evidence that gender, being a health professional, or having participated in research are associated with the ability to assess treatment claims, intended behaviours, or self-efficacy. We did find evidence that people with higher education have a better understanding of Key Concepts.

Understanding the need for systematic reviews of fair comparisons as the basis for trustworthy treatment claims has the potential to reduce waste and harm from trusting and acting on misleading claims, and from not trusting and acting on reliable claims. Future interventions should be tailored to address these essential Key Concepts as well as other Key Concepts for which there appears to be a lack of understanding.
Data availability
Underlying data
Harvard Dataverse: Replication Data for: Raw data Norwegian Claim Study_2019. https://doi.org/10.7910/DVN/R3DHA5

This project contains the following underlying data:
- Quiz-1-data.xlsx
- Quiz-2-data.xlsx
- Quiz-3-data.xlsx
- Quiz-4-data.xlsx

With respect to the potential ability to triangulate identities, access to data is restricted. Ethical approval was granted with an informed consent including a statement where people were guaranteed to be anonymous. Prospective data users can contact the authors to access the data, which can be rejected or approved by the data owner based on certain conditions (must be for legitimate research purposes and requestors are required to sign a data use agreement).

Zenodo: multinormal/fhi.informed-health-choices-norway.2019: Version 1.0. https://doi.org/10.5281/zenodo.366996

This project contains the following underlying data:
- Uganda-children-one-year-follow-up.csv
- Uganda-parents-immediate.csv
- Uganda-parents-one-year-follow-up.csv
- Uganda-teachers-one-year-follow-up.csv

Extended data
Figshare: Additional file 1. Claim Evaluation Tools_tests administrated in Norway_2019. https://doi.org/10.6084/m9.figshare.11439651.v1

Figshare: Additional file 2 Log Rasch analysis Norwegian Claim Study_2019.docx. https://doi.org/10.6084/m9.figshare.10109840.v1

Figshare: Additional file 3. Supplementary tables and figures_Norwegian Claim Study_2019. https://doi.org/10.6084/m9.figshare.11439915.v1

Figshare: Additional file 4. STROBE_checklist_cross-sectional_Norwegian Claim Study_2019. https://doi.org/10.6084/m9.figshare.11439753.v2

Figshare: Additional file 5. English versions of Claim Evaluation Tools_tests administrated in Norway_2019. https://doi.org/10.6084/m9.figshare.11652342.v1

Reporting guidelines
Figshare: Additional file 4. STROBE_checklist_cross-sectional_Norwegian Claim Study_2019. https://doi.org/10.6084/m9.figshare.11439753.v2

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

Software availability
Source code available from: https://github.com/multinormal/fhi.ihc.norway.2019

Archived source code at time of publication: https://doi.org/10.5281/zenodo.3606259

License: MIT

Acknowledgments
We would like to thank Atle Fretheim for his feedback on the Norwegian translation of the Claim Evaluation Tools Item bank assuring face validity and applicability to the Norwegian context of the translated content. Professor Atle Fretheim, Centre for Informed Health Choices, Norwegian Institute of Public Health has given his permission for his name and affiliations to be included in this publication.

References
1. Chalmers I, Glasziou P, Badenoch D, et al.: Evidence Live 2016: Promoting informed healthcare choices by helping people assess treatment claims. BMJ. Reference Source
2. Sørensen K, Van den Broucke S, Fullam J, et al.: Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health. 2012; 12: 80. PubMed Abstract | Publisher Full Text | Free Full Text
3. Austvoll-Dahlgren A: Improving health literacy: identifying barriers and facilitators to people's ability to obtain health information, and the development and evaluation of an intervention to improve health literacy. Oslo: Unipub; 1 b. (fære pag.); ill. p. 2012. Reference Source
4. Cusack L, Del Mar CB, Chalmers I, et al.: Educational interventions to improve people's understanding of key concepts in assessing the effects of health interventions: a systematic review protocol. Syst Rev. 2016; 5: 37. PubMed Abstract | Publisher Full Text | Free Full Text
5. Nordheim LV, Gunderson MW, Espehaug B, et al.: Effects of School-Based Educational Interventions for Enhancing Adolescents Abilities in Critical Appraisal of Health Claims: A Systematic Review. PLoS One. 2016; 11(8): e0161485. PubMed Abstract | Publisher Full Text | Free Full Text
6. Austvoll-Dahlgren A, Nsangi A, Semakula D: Interventions and assessment tools addressing key concepts people need to know to appraise claims about treatment effects: a systematic mapping review. Syst Rev. 2016; 5(1): 215. PubMed Abstract | Publisher Full Text | Free Full Text
7. Austvoll-Dahlgren A, Johansen M: The patient as participant and knowledge manager. Nor J Epidemiol. 2013; 23(2): 225–230. Reference Source
8. Austvoll-Dahlgren A, Nsangi A, Semakula D, et al.: Measuring peoples' understanding of the effects of treatments: a review of outcome measures. PROSPERO CRD42015016449. 2015.
9. Oxman AD, Chalmers I, Austvoll-Dahlgren A, et al.: Key Concepts for assessing
claims about treatment effects and making well-informed treatment choices [version 2; peer review: 3 approved]. F1000Res. 2019; 8:567.

Pérez Gaxiola G, Austvoll-Dahlgren A: Medir la habilidad de la población general para evaluar afirmaciones to assess claims about treatment effects: Validating a test in Mandarin from Claim Evaluation Tools. 2017; e1040.

Austvoll-Dahlgren A, Oxman AD, Nsangi A, et al.: Key concepts that people need to understand to assess claims about treatment effects. J Eval Based Med. 2015; 8(3):112–25.

Osman AD, Austvoll-Dahlgren A, Garrant A, et al.: Understanding of key concepts relevant to assessing claims about treatment effects: a survey of Norwegian adults. IHC Working Paper. 2017.

Dahlgren A: Additional file 1. Claim Evaluation Tools_tests administrated in Norway. 2019. figshare. Journal contribution. 2019. http://www.doi.org/10.6084/m9.figshare.11439651.v1

Dahlgren A: Additional file 5. English versions of Claim Evaluation Tools_tests administrated in Norway. 2019. figshare. Dataset. 2020. http://www.doi.org/10.6084/m9.figshare.11652342.

Dahlgren A: Additional file 6. Strategic to improve recruitment to randomised trials. Cochrane Database Syst Rev. 2018; 2(2):MR000013.

Tennant A, Conaghan PG: The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis Rheum. 2007; 57(8):1358–62.

Pytlab Group: Introductory Rasch Analysis Using RUMM2020. The Section of Rehabilitation Medicine. University of Leeds, 2016.

Oxman AD, Austvoll-Dahlgren A, Garratt A, et al.: Establishing a library of resources to help people understand key concepts in assessing treatment claims-The "Critical thinking and Appraisal Resource Library" (CARL). PLoS One. 2017; 12(7): e0178666.

Tennant A, Conaghan PG: Establishing a library of resources to assess claims about treatment effects. Cochrane Database Syst Rev. 2017; 49. e013185.

Carling C, Kristoffersen DT, Herrin J, et al.: Strategies to improve recruitment to randomised trials. Cochrane Database Syst Rev. 2018; 2(2):MR000013.

Psylab Group: Additional file 5. English versions of Claim Evaluation Tools_tests administrated in Norway. 2019. figshare. Dataset. 2020. http://www.doi.org/10.6084/m9.figshare.11652342.

Psylab Group: Introductory Rasch Analysis Using RUMM2020. The Section of Rehabilitation Medicine. University of Leeds, 2016.

Psylab Group: Additional file 5. English versions of Claim Evaluation Tools_tests administrated in Norway. 2019. figshare. Journal contribution. 2019. http://www.doi.org/10.6084/m9.figshare.11439651.v1

Psylab Group: Introductory Rasch Analysis Using RUMM2020. The Section of Rehabilitation Medicine. University of Leeds, 2016.
56. Norway S: Befolkningens utdanningsnivå 2019. (07.11.2019).
Reference Source
57. Norway S: Helse- og sosialpersonell. 2019.
Reference Source
58. Akl EA, Oxman AD, Herrin J, et al.: Using alternative statistical formats for presenting risks and risk reductions. Cochrane Database Syst Rev. 2011; 2011(3): CD006776.
PubMed Abstract | Free Full Text
59. Robinson EJ, Kerr CEP, Stevens AJ, et al.: Lay public’s understanding of equipoise and randomisation in randomised controlled trials. Health Technol Assess. Research Support, Non-U.S. Gov’t. NHS R&D HTA Programme; Report No.: 1366-5278 (Linking) Contract No.: 8. 2005; 9(3): 1-192, iii-iv.
PubMed Abstract | Publisher Full Text
60. Austvoll-Dahlgren A, Helseth S: Public health nurses’ barriers and facilitators to the use of research in consultations about childhood vaccinations. Scand J Caring Sci. 2012; 26(2): 271–8.
PubMed Abstract | Publisher Full Text
61. Treweek S, Flottorp S, Fretheim A, et al.: [What do general practitioners do to keep themselves up to date?]. Tidsskr Nor Laegeforen. 2005; 125(3): 304–6.
PubMed Abstract
62. Heiwe S, Nilsson-Kajermo K, Olsson M, et al.: Evidence-based practice among Swedish medical social workers. Soc Work Health Care. 2013; 52(10): 947-58.
PubMed Abstract | Publisher Full Text
63. Glenton C, Nilsen ES, Carlsen B: Lay perceptions of evidence-based information—a qualitative evaluation of a website for back pain sufferers. BMC Health Serv Res. 2006; 6: 34.
PubMed Abstract | Publisher Full Text | Free Full Text
64. Oxman AD, Martinez GL: Comparison of the Informed Health Choices Key Concepts to other frameworks that are relevant to learning how to think critically about treatment claims, comparisons, and choices: protocol for a mapping review. IHC Working Paper. 2018.
Reference Source
65. Aronson JK, Barends E, Boruch R, et al.: Key concepts for making informed choices. Teach people to think critically about claims and comparisons using these concepts, urge Andrew D. Oxman and an alliance of 24 researchers — they will make better decisions. Nature. 2019; 572(7769): 303–306.
Publisher Full Text
66. Dahlgren A: Additional file 4. STROBE_checklist_cross-sectional_Norwegian Claim Study_2019. figshare. journal contribution. 2019.
http://www.doi.org/10.6084/m9.figshare.11439753.v2
Susie Sykes
Institute of Health and Social Care, London South Bank University, London, UK

Many thanks for the opportunity to review this paper. It is a very interesting paper and explores an important area of enquiry.

The paper sets itself up as a cross-sectional study of critical health literacy when in fact it is a cross-sectional study of critical appraisal skills. The two are different and the conceptual basis of the study therefore needs clarification. The literature offering conceptual clarity on critical health literacy has not been drawn upon and therefore key components of the concept including the important element of social and political action are overlooked. The paper should either more accurately reflect critical health literacy as a concept or refer instead to critical thinking or critical appraisal skills.

The authors should consider literature such as:[1],[2],[3],[4],[5].

The authors are correct that little has been done to measure CHL but do not refer to work other than the HLS_EU. They should consider:
[6],[7],[8]
The last of these was conducted in Norway and would seem extremely relevant.
The purpose of comparing the Norwegian and Ugandan populations is unclear to me. The rationale for such a comparison is not clearly made. No reference to the comparison is made in the discussion or conclusion so the value of the comparison within the results section is not clear. The different context between the two countries is also not properly explored. A statement is given that: it is our experience from previous projects that the challenges associated with assessing claims in everyday life in Uganda and Norway are very similar but this is insufficient.

References
1. Nutbeam D, Lloyd JE: Understanding and Responding to Health Literacy as a Social Determinant of Health. Annu Rev Public Health. 2021; 42: 159-173 PubMed Abstract | Publisher Full Text
2. Abel T, McQueen D: Critical health literacy in pandemics: the special case of COVID-19. Health Promot Int. 2021; 36 (5): 1473-1481 PubMed Abstract | Publisher Full Text
3. Sykes S, Abel T: Critical Health Literacy. 2023. Publisher Full Text
4. Sykes S, Wills J, Rowlands G, Popple K: Understanding critical health literacy: a concept analysis. *BMC Public Health*. 2013; **13**: 150 PubMed Abstract | Publisher Full Text
5. Chinn D: Critical health literacy: a review and critical analysis. *Soc Sci Med*. 2011; **73** (1): 60-7 PubMed Abstract | Publisher Full Text
6. Chinn D, McCarthy C: All Aspects of Health Literacy Scale (AAHLS): developing a tool to measure functional, communicative and critical health literacy in primary healthcare settings. *Patient Educ Couns*. 2013; **90** (2): 247-53 PubMed Abstract | Publisher Full Text
7. Barsell DJ, Everhart RS, Perrin PB: Refining the Factor Structure of the All Aspects of Health Literacy Scale. *Am J Health Behav*. 2020; **44** (2): 118-128 PubMed Abstract | Publisher Full Text
8. Haugen ALH, Riiser K, Esser-Noethlichs M, Hatlevik OE: Developing Indicators to Measure Critical Health Literacy in the Context of Norwegian Lower Secondary Schools. *Int J Environ Res Public Health*. 2022; **19** (5). PubMed Abstract | Publisher Full Text

*Is the work clearly and accurately presented and does it cite the current literature?*
No

*Is the study design appropriate and is the work technically sound?*
Yes

*Are sufficient details of methods and analysis provided to allow replication by others?*
Yes

*If applicable, is the statistical analysis and its interpretation appropriate?*
I cannot comment. A qualified statistician is required.

*Are all the source data underlying the results available to ensure full reproducibility?*
Yes

*Are the conclusions drawn adequately supported by the results?*
Yes

*Competing Interests:* No competing interests were disclosed.

*Reviewer Expertise:* public health and health promotion

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.
Thank you for a very innovative research study around critical health literacy competencies and a comparative study across two countries. The educational intervention at primary school level and the podcast for parents are novel. Below are some suggestions that may help to improve your manuscript:

Title:
○ The title of your paper could be amended to more accurately reflect your study. The use of the word 'claims' in a health care setting can very easily be confused with health insurance claims! This will then need to follow through your entire manuscript. Perhaps consider using the word 'statement' rather than 'claims' so your title might be something like the following:

The Norwegian public's ability to assess statements about health treatments: ...

Abstract:
○ For the above reason the Background section of the Abstract needs to be rephrased.

Research Design:
○ Surrounding the research design, the rationale for comparing Uganda with Norway is not clearly explained to the reader. Why are Uganda and Norway comparable? A justification is necessary.

Methods and Analysis:
○ While a rationale can be understood for a comparison of adults in one country with parents and teachers in another country (all are adults), the rationale for comparing adults in one country with primary school children in another country remains a puzzle. This is so especially because the quizzes collect data on education level attained and so can be controlled for. Again this needs to be explained more clearly to the reader.

○ The structure of the Methods section is difficult for the reader to follow. For example the methods section begins with explaining through Table 1 the three ‘Health Choices Key Concepts’ of ‘Claims, Comparisons and Choices’ and then moves on to the four ‘concepts’ immediately following Box 1. The use of the identical term ‘concept’ to describe both of the above is very confusing to the reader. It is not clear how the three ‘Health Choices Key Concepts’ link or map to these four ‘concepts’. Furthermore the label of ‘concepts’ does not seem to be accurate for the four ‘concepts’. Are these concepts? Perhaps ‘contentions’ is a more accurate label. This entire section needs to be redrafted.
○ It is not clear from the manuscript if the quizzes went through a plain English audit process. This should be stated.

○ For an English journal publication, it might be best to move the English version of the quizzes to be Additional File 1, rather than the Norwegian version.

Conclusions:
○ The conclusion stated in the Abstract ‘This can result in poorly informed decisions, underuse of effective interventions, and overuse of ineffective or harmful interventions’ does not align with the conclusions at the end of the manuscript. Indeed it is difficult to see how these conclusions set out in the Abstract can be made from this particular study.

Two typing errors were noted:
1. Ethical Statement, second sentence
2. Page 13 should read ‘including a statement where people were guaranteed…’

Thank you for a very interesting piece of research and I wish you every success enhancing your paper.

Is the work clearly and accurately presented and does it cite the current literature?
Partly

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
I cannot comment. A qualified statistician is required.

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Health Literacy, Value measurement in health care

I confirm that I have read this submission and believe that I have an appropriate level of
expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Astrid Dahlgren

Thank you for considering our paper and for this useful feedback. We have revised the manuscript and considered all suggested revisions carefully. Our revisions and comments are provided below:

1. Suggestion to change the name “Claim”.

   **Our response:** we understand that this term may be understood differently based on context. However, this term has been evaluated for relevance in several country contexts and be considered as the most precise term to be used. The Claim Evaluation Tools is also the name of the item bank this study was based on and thus cannot be changed.

2. Suggestion to add rationale for comparing Uganda and Norway

   **Our response:** the rationale for this comparison has now been added.

3. Suggestion to add rationale for comparing children and adults.

   **Our response:** the rationale for this comparison has now been added.

4. Suggestion to restructure the methods section describing the Key Concepts list and the claim Evaluation Tools item bank.

   **Our response:** these sections have now been revised accordingly to improve clarity.

5. Suggestion to clarify whether the tests went through a plain English audit process. The test was based on a set of MCQS translated from the English to Norwegian.

   **Our response:** statement now added to the manuscript.

6. Suggestion to replace the Norwegian version of the tests with the English version in the additional file.

   **Our response:** F1000 publishes international studies reporting evidence from a range of contexts. The evaluation instruments used in this study was designed and adapted to a Norwegian setting. Consequently, our study describes the validation and results of the Norwegian tests. Responses and psychometric properties of the tests may be different for
the English versions of these tests. Thus, we believe it to be most appropriate to include the tests that was used in the study. However, examples of the questions in English has been provided in the manuscript. Other studies have used the English versions of the MCQs in other contexts (such as in Uganda).

7. Suggestion to revise the conclusion taking into account: the conclusion stated in the Abstract 'This can result in poorly informed decisions, underuse of effective interventions, and overuse of ineffective or harmful interventions' does not align with the conclusions at the end of the manuscript. Indeed, it is difficult to see how these conclusions set out in the Abstract can be made from this particular study.

Our response: the sentence has been deleted.

8. Typos were corrected were noted

**Competing Interests:** No competing interests were disclosed.

---

Reviewer Report 17 March 2021

[https://doi.org/10.5256/f1000research.24149.r80622](https://doi.org/10.5256/f1000research.24149.r80622)

© 2021 Lühnen J. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Julia Lühnen

1 Institute for Health and Nursing Science, Medical Faculty, al Faculty, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany
2 Clinic for Internal Medicine I, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany
3 Institute for Health and Nursing Science, Medical Faculty, al Faculty, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany
4 Clinic for Internal Medicine I, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany

The authors provided a very interesting and highly relevant paper. Critical health literacy and the ability to assess the trustworthiness of claims about healthcare are prerequisites for informed choices. For the assessment of public health literacy, only self-reported measures are used in most of the studies.

The authors performed a cross-sectional study to assess the Norwegian public's ability to assess treatment claims. The questionnaires, assessing understanding, are based on the key concepts people need to understand to assess healthcare claims, developed in the IHC project. The key concepts and the item bank are regularly up-dated, open access resources. The item bank
included at least four multiple-choice questions for 32 Key Concepts. From this item pool four questionnaires were designed and randomly assigned to the participants.

771 out of 4500 invited Norwegian adults responded. Rasch analysis were performed to evaluate the psychometric properties, the percentage of the Norwegian population that understands each key concept was calculated, the results were compared to the results of a previous study in Uganda and analyses were conducted to explore associations between understanding of the key concepts and demographic characteristics. The results are reported in detail. The authors’ discussed strength and limitations of their work and provided implications for practice and future research.

I have only few comments to optimize the manuscript:

- The description of the results in the abstract and on page 7 last paragraph is confusing. Maybe there better way providing the results than using “...more than half understood x...” and “...less than half...”?

- Page 6 in the paragraph “Rasch analysis”: There is a repetition of the sentence “Rasch analysis can be used for both dichotomous and polytomous data.”

- The authors used the STROBE checklist. Looking at reporting guidelines for online surveys (e.g. CHERRIES, Eysenbach 2004), some aspect are not entirely addressed such as the number of items in each questionnaire, number of screens/pages, completion rate (besides covariates) or the handling of incomplete questionnaires.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
I cannot comment. A qualified statistician is required.

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

**Competing Interests:** The reviewer is involved in the translation and validation of the IHC Claim evaluation tool for its use in German speaking countries.

**Reviewer Expertise:** evidence-based health information, informed decision-making, trainings in evidence-based medicine
I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 14 Jul 2021

Astrid Dahlgren

Thank you for considering our paper and for this useful feedback. We have revised the manuscript and considered all suggested revisions carefully. Our revisions and comments are provided below:

1. The description of the results in the abstract and on page 7 last paragraph is confusing. Maybe there better way providing the results than using “...more than half understood x...” and “...less than half...”?

**Our response:** these sections have now been revised accordingly to improve clarity.

1. Page 6 in the paragraph “Rasch analysis”: There is a repetition of the sentence “Rasch analysis can be used for both dichotomous and polytomous data.”

**Our response:** typo is corrected.

1. The authors used the STROBE checklist. Looking at reporting guidelines for online surveys (e.g. CHERRIES, Eysenbach 2004), some aspect are not entirely addressed such as the number of items in each questionnaire, number of screens / pages, completion rate (besides covariates) or the handling of incomplete questionnaires.

**Our response:** a description of the number of items per questionnaire has now been added to the description of the questionnaires in the methods section. Responses were mandatory in the software we used and thus all questionnaires were complete (thus it was not necessary to handle any incomplete questionnaires).

For missing values on specific questions, we used multiple imputation with chained equations to account for the uncertainty introduced by missing values of the post-stratification variables sex, region of residence, and educational level. This is stated under the subheading “Survey analysis”

**Competing Interests:** No competing interests were disclosed.
The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com