Web for Public Involvement∗

Using Crowdsourcing and Text Analysis to help review Information Leaflets

for Clinical Trials

Fernando Santos†  
 ECS-FoM  
 University of Southampton  
 Southampton, UK  
 fss1g15@soton.ac.uk

Thanassis Tiropanis   
 ECS  
 University of Southampton  
 Southampton, UK.  
 t.tiropanis@southampton.ac.uk

Jeremy Wyatt  
 Wessex Institute  
 University of Southampton  
 Southampton, UK  
 j.c.wyatt@soton.ac.uk

Adam Geraghty  
 FoM  
 University of Southampton  
 Southampton, UK.  
 a.w.geraghty@soton.ac.uk

ABSTRACT

**Objective:** To assess the viability of employing a Web platform to receive feedback from a public audience on Information Leaflets (PILs) for Clinical Trials (RCTs) in the UK.

**Design and Setting:** A feasibility study on employing Amazon Mechanical Turk (MTurk) to recruit, and engage participants to review, comment, revise and assess PIL information.

**Interventions:** This study included 4 non-clinical interventions to assess the feasibility of employing a Web platform to collect feedback on PILs, objectively assess the difficulty of PIL information, and employ crowdsourcing to revise and validate proposals for improving readability of sentences that were too found to require reading skills above the expected level for general audiences.

**Outcomes:** The primary outcomes were the feasibility of the study design to employ a web platform and crowdsourcing to review, revise and validate PIL information. Secondary analysis was carried on the results to assess the association between task performance and sentence difficulty, participant reading skill, and fatigue and learning effects.

**Results**

**Conclusions**

BACGROUND

METHOD

1. Design

We conducted a four-amr feasibility study of the “Web for Public Involement” webtool previously designed to facilitate the review and correction of readability issues in RCT PILs by employing MTurk to crowdsource public feedback.

Our study assessed the feasibility of employing the webtool to:

1. Collect public feedback from MTurk participants on PIL information quality
2. Revise sentences that require higher reading skills than those expected from a general audience by employing MTruk crowdsourcing.
3. Employ MTurk participants to validate previous revisions on sentences with low readability.
4. Objectively assess the sentence difficulty and the participants reading skill.

Additional analysis of the results was done to assess the fatigue and learning effects of revising PIL sentences.

1. Participants
2. Recruitment
3. Interventions

4.1 Collecting public feedback on PIL information quality from MTurk participants

The participants in this intervention were asked to read, comment, and assess the quality of information on three PILs which presented a high number of readability issues and poor recruitment rates in their clinical trials.

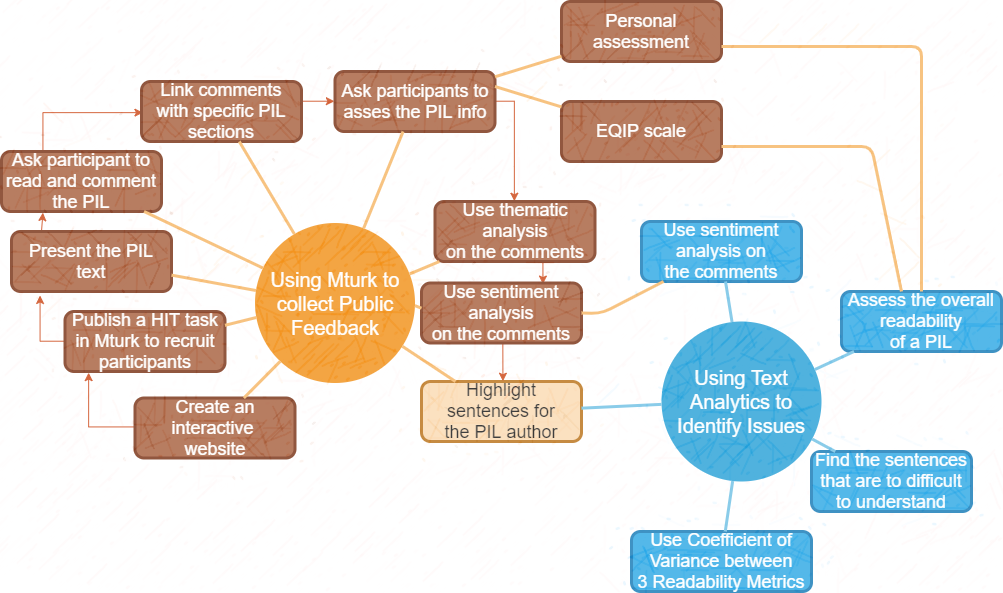


Figure 1 First Intervention Methodology

The participants were recruited by posting a Human Intelligence Task (HIT) in MTurk and offering £2 to read, comment and answer a multiple choice questionnaire about the quality of information of three PILs.

The participants were then presented with a webpage containing the PIL text and asked to read and comment on it. Participants could comment on two ways, by selecting a specific section of text and clicking the “make a comment button” or by filling a general comment on the overall quality of the text.

The comments obtained in this intervention were examined with content and thematic analysis and their results compared with public feedback from a face-to-face study.

The participants then were asked to assess the quality of information of the PIL by answering a multiple choice questionnaire based on the Ensuring Quality of Patient Information (EQIP) extended guidelines **[ref]**.

Finally, the collected feedback was summarized and presented as an electronic report to the PIL authors via the Webtool and asked to assess the validity of the feedback and the potential impact to their research.

4.2 Revising sentences that require reading skills above the average level expected by general audiences

PIL sentences which required a higher reading skill level than the average reading skill expected from general audiences were identified by employing the Coefficient of Variation between three commonly use readability indexes, SMOG, ARI and Gunning Fog.

4.3 Employing MTurk to validate the proposed revisions for sentences with low readability

A selection of proposed revisions for sentences which required higher reading skills than the average level for general audiences was made by stratified random sampling. The factor of variance between SMOG, ARI and Gunning Fog indexes was computed for the readability scores of every proposed revision. A selection of 9 proposed revisions was made for each leaflet by selecting three random revisions from three levels of readability: undergraduate, graduate student and professor.

4.4 Objectively assessing the readability of a PIL and the participants reading skill level.

The Cloze procedure was employed on the selected sentences with low readability previously found. The Cloze procedure consist on replacing every nth word of a document with a blank space and asking a participant sample to complete the sentences, it assess the difficulty of a document based on the number of identical words submitted by the participants. The MTurk participants were then showed each processed sentence in a webpage and asked to complete them. The participant performance was measure as a factor of the sentence readability score, the time taken in the revision and the number of correct inputs submitted. The number of identical words submitted by the participants for each original sentence was used to validate the results from the readability indexes.

1. Outcomes and Measures
2. Sample Size
3. Randomization
4. Analysis

RESULTS

DISCUSSION

In this sample-structured document, neither the cross-linking of float elements and bibliography nor metadata/copyright information is available. The sample document is provided in “Draft” mode and to view it in the final layout format, applying the required template is essential with some standard steps.

These steps, which should require generation of the final output from the styled paper, are mentioned here in this paragraph. First, users have to run “Reference Numbering” from the “Reference Elements” menu; this is the first step to start the bibliography marking (it should be clicked while keeping the cursor at the beginning of the reference list). After the marking is complete, the reference element runs all the options under the “Cross Linking” menu.

For accuracy check of the structured paper, user can run the option **Manuscript Validation**. It informs the user of the wrong or missing values in the paper. The user must correct the paper as per validation messages and rerun **Manuscript Validation**.

Now, to generate the required layout of the paper, the user should select one of the template styles under the Define Template Style option (choose the required layout design, i.e. choose between Journals and Proceedings).

∗Article Title Footnote needs to be captured as Title Note

†Author Footnote to be captured as Author Note

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

*WebSci’20, July, 2020, Southampton, UK*

© 2018 Copyright held by the owner/author(s). 978-1-4503-0000-0/18/06...$15.00

https://doi.org/10.1145/1234567890

Some specific values are required to create a standard layout by choosing a template for the journals or proceedings. So once the user chooses one of the template layout styles, the respective Journal or Conference details dialog box (i.e. **journal/conference acronym, DOI, ISBN, copyright, year, etc.**) will appear as a prompt during the Define Template Style functionality. The user should fill these values, after which the template creates the desired layout of the paper. The user can now create a PDF of his/her manuscript using the “*Save as PDF*” option.

If the user is adding any new data, they should make sure to style it as per the instructions provided in previous sections. Carry out the steps for Cross-linking, Fundref data, adding Document History (specific to journal submission), and finally, Manuscript validation and placing the respective metadata (Bibstrip/copyright text)[[1]](#footnote-2) while applying the required template.

CCS CONCEPTS

• Insert CCS text here • Insert CCS text here   • Insert CCS text here

KEYWORDS

Patient Information Leaflet, Randomized Clinical Trial, Web Tool, Sentiment Analysis, Thematic Analysis, Information Quality, Text Analysis, Readability, EQIP, SMOG.

ACM Reference format:

Fernando Santos, Thanassis Tiropanis, Adam Geraghty and Jeremy Wyatt. 2020. Web for Public Involvement: Using Crowdsourcing and Text Analysis to help review Information Leaflets for Clinical Trials. In *Proceedings of 12th ACM Web Science Conference (WebSci20). ACM, Southampton, UK,* ***6 pages.* https://doi.org/10.1145/1234567890**

1. 1 Insert Heading Level 1

The updated template, user manuals, samples, and required fonts, all are available at the URL <https://www.acm.org/publications/proceedings-template>. It contains said information for all three versions of MS Word (Windows and 2 versions of Mac). There are also separate links to the user guide, which can be referred to by the user. This URL also contains some useful video links, which describe how to add the template, structure the paper, and generate the layout, in different clips. **Display Formula with Number**

 (1)

**Continuation part of Paragraph Text** The user must style this paragraph in **ParaContinue** style, which follows immediately after the **DisplayFormula** (numbered equation). The **DisplayFormula** style is applied only in case of a numbered equation. A numbered equation always has a number to its right. Insert paragraph text here. **Display Formula without Number**



The **DisplayFormulaUnnum** style is applied only in case of an unnumbered equation. An unnumbered display equation never contains an equation number to its right, and this unique property distinguishes it from a numbered equation.



Figure 1: Figure Caption and Image above the caption [In draft mode, Image will not appear on the screen]

**Theorem/Proof/Lemma.** Insert text here for the enunciation or Math statement. Insert text here for the enunciation or Math statement. Insert text here for the enunciation or Math statement. Insert text here for the enunciation or Math statement. Insert text here for the enunciation or Math statement.

....Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract.

1.1 Heading Level 2

In the below paragraph, it is explained how alt-txt value is placed in **MS Word 2010**. To add alternative text to a picture in Word 2010, follow these steps:

1. In a Word 2010 document, insert a picture.
2. Right click on the inserted picture and select the **Format Picture** option.
3. Select the **Alt Txt** option from the left-side panel options.
4. In the "Title:" and "Description:" text boxes, type the text you want to represent the picture, and then click "Close".

Below are steps to place alt-txt value in **MS Word 2013/2016**. To add alternative text to a picture in Word 2013/2016, follow these steps:

1. In a Word 2013/2016 document, insert a picture.
2. Right click on the inserted picture and select the **Format Picture** option.
3. In the settings at the right side of the window, click on the "Layout & Properties" icon (3rd option).
4. Expand **Alt Txt** option.
5. In the "Title:" and "Description:" text boxes, type the text you want to represent the picture, and then click "Close".

*1.1.1 Heading Level 3.* Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here.

*1.1.1.1 Heading Level 4.*Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here.

ACKNOWLEDGMENTS

Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here.

REFERENCES

[1] Patricia S. Abril and Robert Plant, 2007. The patent holder's dilemma: Buy, sell, or troll? *Commun. ACM* 50, 1 (Jan, 2007), 36-44. DOI: <https://doi.org/>10.1145/1188913.1188915.

[2] Sten Andler. 1979. Predicate path expressions. In *Proceedings of the 6th. ACM SIGACT-SIGPLAN Symposium on Principles of Programming Languages (POPL '79)*. ACM Press, New York, NY, 226-236. DOI:https://doi.org/10.1145/567752.567774

[3] Ian Editor (Ed.). 2007. *The title of book one* (1st. ed.). The name of the series one, Vol. 9. University of Chicago Press, Chicago. DOI:https://doi.org/10.1007/3-540-09237-4.

[4] David Kosiur. 2001. *Understanding Policy-Based Networking* (2nd. ed.). Wiley, New York, NY..

Conference Name:ACM Woodstock conference

Conference Short Name:WOODSTOCK’18

Conference Location:El Paso, Texas USA

ISBN:978-1-4503-0000-0/18/06

Year:2018

Date:June

Copyright Year:2018

Copyright Statement:rightsretained

DOI:10.1145/1234567890

RRH: F. Surname et al.

Price:$15.00

1. The existing Bibstrip data, copyright text and permission block in the sample file are dummy values, so the user needs to provide the correct values required for the submission in the metadata dialog box. [↑](#footnote-ref-2)