# Design of Tooltips for Data Fields

## A Field Experiment of Logging Use of Tooltips and Data Correctness

Helene Isaksen[1], Mari Iversen[1(✉)], Jens Kaasbøll[1], and Chipo Kanjo[2]

[1] University of Oslo, Oslo, Norway
{helenis, mariive, jensj}@ifi.uio.no
[2] University of Malawi, Zomba, Malawi
chipo.kanjo@gmail.com

**Abstract.** Many health professionals in developing countries carry out tasks which require a higher level of education than they have. To help such undereducated health workers filling correct data in patient information systems, data fields were furnished with tooltips for guiding users. In a previous study with questionnaires and interviews, health workers preferred tooltip contents being normal values of the data with medical explanation as the second best. The experiment reported in this paper set out to test these content alternatives and also aimed at finding health workers' use of tooltips and possible effects on data correctness. In order to resemble the work setting, each of the 15 undereducated health workers participating was given a tablet PC with the patient information system and booklet of 22 cases to be entered over a period of two weeks. They were given a one hour introduction to the system. Their use of the tablet was recorded, and after completing, the participants were interviewed. The health workers opened tooltips frequently for the first cases, and thereafter the use dropped. Reasons given were that they learnt the data field during the first cases, and thereafter they did not need the tooltips so often. The number of correct data entries increased over time. The group with medical explanation tooltips performed better than the group with normal value tooltips, thus the preferred tooltip in the questionnaire gave a lower performance than the second alternative. While the experiment demonstrated that tooltips improved performance, it did not quantify the effect.

**Keywords:** Usability evaluation · Field experiment · Logging use · Learnability · Context-sensitive help · Tooltip contents · Normal data values · Formal definitions · Data quality

## 1 Introduction

Health workers in developing countries are often assigned tasks meant for those of higher cadres. As an example, undereducated staff have to do the tasks of nurses [5]. Doing work-related tasks beyond one's competence may lead to wrong data capturing and may cause fatal decision making. Training and follow ups of undereducated are often unsuccessful due to lack of supporting staff and funding. In addition, IT systems


A. Marcus and W. Wang (Eds.): DUXU 2017, Part III, LNCS 10290, pp. 63–78, 2017.
DOI: 10.1007/978-3-319-58640-3_6

are often designed for expert users, thus there is a need for providing information health workers can look up and use themselves.

There are several methods to provide additional information for users. These include users looking up information online, from external sources or by including inline information in the system. Adding inline additional information may be a solution, however, this research aim to test different content types for additional information, and to find the most effective type. Tooltips are the most common ones and have been shown several times to be effective [1, 4, 7]. Due to limitations in the software used for the experiment, textual tooltips are the basis for our research.

Our definition of tooltips is information that can be viewed when the user push a button. The information will disappear from the screen when a button is pushed, or when the user start or finish entering data into the field. The goal for tooltips, in our case, are for the users of the system to understand the medical terms and enter correct information.

Little previous research has addressed the identification of the most effective tooltips in terms of correctness of data entry. Some research has considered user-preference of expression format for tooltips. Petrie et al. [7] identified four expression formats for tooltips and asked their participants to rate the different formats based on satisfaction, understandability and preference, however the research did not opt to find the most effective tooltips. One of the end goals for tooltips are for the user to use the system effectively, therefore, a decreasing usage of help commands or tooltips is seen as a sign of system learnability [6]. Dai et al. [1] developed a software consisting of step-by-step instructions for carrying out tasks. However, these instructions would not function with tooltips, as tooltips are unsuitable for displaying sequences of instructions, since they disappear once a single task is finished. Isaksen et al. [5] conducted a survey of preferences of content types of tooltips by lower cadre health workers. The health workers preferred tooltips expressed as normal values of the data to be entered. However, their study did not explore if the tooltips actually led to more correct data entry. Their findings constitute a basis for our study.

The objectives for this research is to compare two content types for tooltips and find out whether there is a difference between them in terms of correctness of data entry. We also wish to see if the tooltips actually affect the correctness. Our research is, therefore, an experiment to find out how often the users use the tooltips, and if they can be seen as successful. By successful tooltip, we mean that they have opened the tooltip, and that they enter the correct data.

## 2 Tooltip Contents

Through interviews with professionals within Antenatal care (ANC) systems, Isaksen et al. [5] identified four content types for tooltips for medical terms. These content types were normal values, the formal definition, treatment, and procedure to find measurements. They found that normal values were the most preferred among health workers of different cadres, with formal definitions as the runner up. Therefore, this study will focus on these two alternatives.

Tooltips containing formal definitions, or explanations, explain medical terms. An example from the study is "Occurs when the woman has hypertension and proteinuria. It can happen at any point after week 20 of pregnancy.", which is the explanation of pre-eclampsia.

Normal values in the tooltips provide either a range of normal values or signs of the given condition. For example, pre-eclampsia has the following normal value tooltips: "Signs: Diastolic blood pressure above 90 and protein in urine.".

Below are some examples from the experiment, showing both versions of the tooltip (Table 1).

**Table 1.** Examples of the two content types

| Data element | Normal value | Explanation |
|---|---|---|
| Pre-eclampsia | Signs: diastolic blood pressure above 90 and protein in urine | Occurs when the woman has hypertension and proteinuria. It can happen at any point after week 20 of pregnancy |
| Diastolic blood pressure | Diastolic blood pressure should be between 60 and 80 | Diastolic blood pressure is the minimum blood pressure |
| Fundal height | Normal fundal height measurement:<br>20 weeks = 17–20 cm<br>28 weeks = 25.5–28.5 cm<br>36 weeks = 33–35 cm<br>40 weeks = 36–38 cm | Measurement from the public bone to the top of the uterus. This is done to assess how far into the pregnancy the woman is |

## 3 Technology Description

In order to conduct the experiment, we utilized a generic software package called District Health Information System 2 (DHIS2). The DHIS2 package can either be run through a web browser or through Android apps. For our study the Tracker Capture (TC) android app was used for hosting the testing program. The TC enables the end users to track people or objects over a period of time, and follow up each individual case. The TC can be tailored in the web version for different purposes, and one can create specific programs. For our research the two first authors created two shortened antenatal care programs, and added data elements, skip logics, tooltips and options sets. The data elements were chosen based on Malawian health passports. The programs used exactly the same data elements and order, but the tooltips had different content types.

In Malawian health passports, blood pressure is registered in a single field, labeled either "Blood pressure" or just "BP", and is not marked diastolic and systolic. Therefore, we wanted to check the participants' ability to cope with unusual order of data fields, and chose to list diastolic and systolic in the opposite order of how one usually writes them (see Fig. 1).

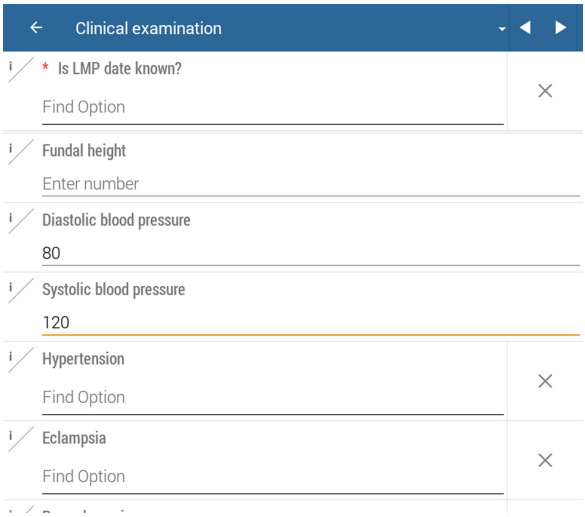


**Fig. 1.** Example of diastolic and systolic data elements in Tracker Capture

The data elements were assigned to stages, like "Previous pregnancies" and "First antenatal care visit", and categories, like "Family history" and "Clinical examination". "Previous pregnancies" stood out by being the only one which contained checkboxes for different data elements. This was done for the program to resemble the health passports, where information is entered for all previous pregnancies in one page, rather than separate pages for each pregnancy (Fig. 2).

In order to register the informant's behavior in the system, an analytic tool called UXcam was utilized. UXcam is a tool used for improving user experiences in applications, through screen recordings, emphasizing the touches on the screen. The recordings are stored on UXcam's server and are accessible through their web page. The tool was added to the TC code, enabling us to watch and analyze the informants behavior on the screen. The tablets could be traced by the tablet's own ID, as well as the profession of the participant using the tablet. This gave us an impression of their progress throughout the experiment. However, there were risks using this additional software, as we were dependent on the participants being connected to internet when doing their tasks. UXcam is only able to send recordings if connected to the internet, meaning we were at risk of not getting all of the recordings. Thus we equipped each of the tablets with sim cards and preloaded internet bundles. To ensure that the internet bundle was only used for the experiment, an app called "Applocker" was installed, blocking the usage of all other applications.

For the study, 30 tablets were bought, one for each participant. The two first authors installed the TC on all the tablets, making sure the system was running.

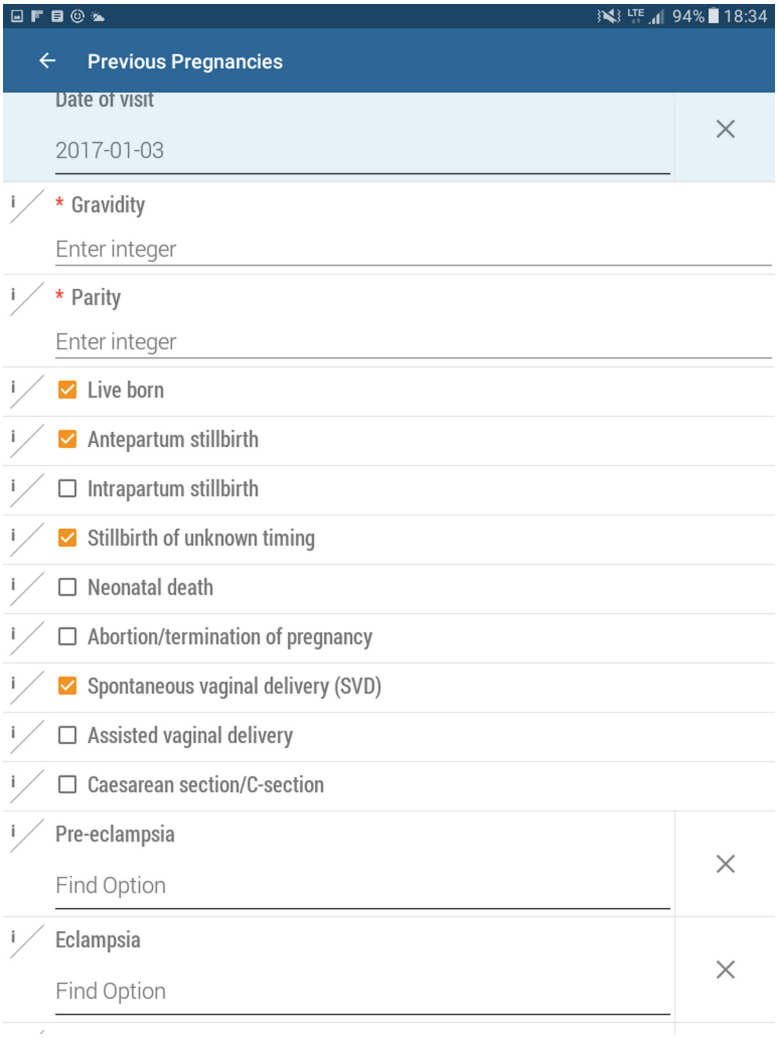


**Fig. 2.** Here is "Live born", "Antepartum stillbirth", Stillbirth of unknown timing" and Spontaneous vaginal delivery (SVD)" checked, meaning that the woman has experienced these in her previous pregnancies.

## 4 Method

In order to get a better understanding of the health worker's use of the tooltips, we decided to carry out an experiment. We chose to conduct the experiment in natural settings, as this could introduce issues which the participants would not encounter in a lab [2]. It was also important to test over time, in order to see their evolvement. We also wanted to see if they learned anything from the tooltips.

As mentioned, the tablets contained either a program with tooltips containing normal values, or explanations, and these were given to the participants randomly.

### 4.1 Informants

We chose participants of cadres lower than nurses and higher than community health workers, with ANC experience. A total of 30 people participated in this experiment, however, some of them turned out to be nurses of different degrees. The initial idea was to do 15 participants in South Africa and 15 in Malawi. However, due to misunderstandings and time constraints, the distribution was 20 in Malawi and 10 in South Africa.

This article will include results from the first 15 participants from Malawi, as the experiment extends past the deadline for final version of this paper. The participants in Malawi were recruited by the fourth author, either by appointments or by asking acquaintances and other participants if they knew anyone in the respective cadres.

### 4.2 Cases

To ensure that the participants used every part of the system and the provided tooltips, the two first authors created a total of 22 cases. Data from these cases was entered into the TC app by the participants over a period of eleven days, two cases a day.

**Enrollment date:** Today's date
**First name:** Pika
**Last name:** Chula
**Date of birth:** 14th march 1985
**Marital status:** Single
**Mobile number:** 123 123 245

**Previous pregnancies:**
Date of visit: today's date
Pika has had two embryos removed, and has given birth to three babies. One of them was delivered through an incision in the abdomen, but, unfortunately, died before the onset of labour. Pika doesn't remember much of it because she was in a coma. The two other were born in normal manners.

**First visit:**
Date of visit: today's date
Pika's father react to blinking lights and often get seizures, while her mother has a disorder of metabolism which makes her drink a lot of water and produce large amounts of urine. Pika herself often experience difficulties breathing due to spasms in the bronchi of the lungs, and is in addition allergic to antibiotics in general. She cannot remember her LMP, but her fundal height is 25 cm, her blood pressure is 120/90 and she has protein in her urine. She has been given malaria prophylaxis and iron supplements.

**Fig. 3.** An example of a case from the booklet

The cases contained information about fictive pregnant women, often quite sick and having lost multiple children. However, it was not written straightforward, but was instead disguised as symptoms, or resembling the information the participants could find in the tooltips. Examples are ".. lost the child in week 38, before the onset of labour", which indicates an antepartum stillbirth, or ".. has abnormally high blood pressure and protein in the urine", which indicate pre-eclampsia.

Several of the cases contained similar information, and these were distributed evenly over the period. This was to see is the participants learned the different expressions from one day to another (Fig. 3).

### 4.3 Introducing the Experiment

The experiment started with a brief introduction about who we were, where we came from, and that we wanted to work on improving the usability of a system. We did not inform them about the testing of the tooltips, to make sure we wouldn't affect the results. We then introduced them to the tablets and the TC, explaining what the application did, using a modified question suggestion approach [5]. This included making them aware of the tooltips, informing them that they could use these if they were in doubt regarding what information to enter. We also presented them with the same example case, similar to the next 22 cases they would solve.

The participants in Malawi were situated in groups of three, four or five people, enabling them to cooperate and discuss the matter as they would have in a normal work situation. This also gave us the opportunity to observe what each of them did, and to evaluate their technical skills. The observation enabled us to adapt the information given during the introduction, and to give proper follow-up on each participant. Also, a lot of the explaining of the different elements and tasks was repeated in Chichewa, the local language, by the fourth author. This seemed to increase their understanding of the experiment, the tasks and other unfamiliar expressions. At the end of the introduction they were given the same questionnaire as Isaksen et al. used, capturing the preference of content types for tooltips.

### 4.4 The Booklets

For this experiment we created a booklet containing information about us, the experiment and 22 cases with tasks for each day. Diaries are used to collect data about user behavior and activities over a longer period of time, and may provide a contextual understanding of the usage of the system [3]. Thus, the booklets were inspired by a diary technique, where the task section would function as a diary. Here, the participants could write down when and where they entered the case, how they felt using the system, what data elements they used and thoughts on the cases. The goal of this was to make them reflect on their case, and to make it easier for them to discuss their thoughts and ideas during the post-interview. The participants were given the booklets after going through the example case.

The booklet also contained information about who we were, and what they were supposed to do. Email contact information was also given in the booklet, allowing for

the participants to contact us if they had any questions. In addition, they were also given a phone number to the fourth author, who functioned as a local contact, in case of urgent questions.

### 4.5 The Post-interviews

After approximately two weeks we asked the participants for a semi-structured interview, aiming to get a better understanding of their use of the tooltips and general thoughts of the entire experience. The questions focused on opinions on the information in the tooltips, and whether they opened the tooltips before or after data entry, and why they did so.

We also collected the booklet and had the participants do the aforementioned questionnaire again to see whether the opinion remained the same or changed. In addition, an online questionnaire was created capturing the participants user experience of the tooltips (hereby UX questionnaire). In this article, we are only using the responses from the 15 participants mentioned above, as well as the responses Isaksen et al. used in their study.

### 4.6 Analysis

The recordings were structured and analyzed in a google sheet document (Fig. 4). The participants were differentiated by having separate sheets, listing all data elements from the program. The first two authors registered whether the participants entered correct information, and if they opened any tooltips. The sheets were set up to calculate successful tooltips, if both data entry was correct and the tooltip was opened.

| | A | B | C | D | E | F | G | H | I | J | K | L | M |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Data element \| Midwife nurse | Case 1 | | | Case 2 | | | Case 3 | | | Case 4 | | |
| 2 | O=opened, C=Correct, S=successful tooltip. =1 if true | O | C | S | O | C | S | O | C | S | O | C | S |
| 3 | Gravidity | 1 | 1 | 1 | 1 | 1 | 1 | | 1 | 0 | 1 | 1 | 1 |
| 4 | Parity | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 5 | Live born | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | 0 |
| 6 | Antepartum stillbirth | 1 | | 0 | 1 | | 0 | 1 | 0 | 0 | | 1 | 0 |
| 7 | Intrapartum stillbirth | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | 0 | 1 | | 0 |
| 8 | Stillbirth of unknown timing | 1 | | 0 | 1 | | 0 | 1 | 1 | 1 | | 0 | 0 |
| 9 | Neonatal death | | | 0 | 1 | 1 | 1 | 1 | | 0 | 1 | | 0 |
| 10 | Abortion | 1 | | 0 | | | 0 | 1 | | 0 | 1 | 1 | 1 |
| 11 | SVD | 1 | | 0 | 1 | 1 | 1 | 1 | 1 | 1 | | | 0 |
| 12 | Assisted vaginal delivery | 1 | | 0 | | | 0 | | | 0 | | | 0 |
| 13 | C-section | | 1 | 0 | | | 0 | | | 0 | | 1 | 0 |
| 14 | Pre-eclampsia | 1 | 0 | 0 | 1 | 0 | 0 | 1 | | 0 | 1 | 0 | 0 |
| 15 | Eclampsia | 1 | 1 | 1 | | 1 | 0 | 1 | | 0 | | | 0 |
| 16 | Hypertension | | 0 | 0 | | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 |

**Fig. 4.** Screenshot of the spreadsheet used to register opened tooltips and correct data entry

### 4.7 Motivation

In order to motivate the participants to take part of the experiment, they were told, at the end of the introduction, that if they did all their tasks, the tablet would be theirs to keep. This is probably part of the reason why everybody entered all cases, and gave feedback to the tasks. In addition, being aware of that their usage of the systems was being monitored, may also have resulted in a higher willingness to finish the tasks given. We did not start with introducing the reward, as we wanted to recruit somebody that were somewhat interested in the project.

## 5 Results

On average, there were 14 cases recorded per user, in addition we lost all recordings from one user and had one user where we only received eight recordings. This was probably due to connectivity issues, as we, during the post-interviews, found all 22 cases on their tablets.

After analyzing the information we received from the booklets and the interviews, we learned that the participants, on average, spent 20–25 min on each case, and it took them about 3 days to get comfortable with the system. However, many of the participants also stated that they wished they had more training with using the application, as for some of them, this was their first time using a touch screen.

Several informants requested more detailed cases, in order to diagnose the patients properly. They also stated that instead of camouflaging the information we should have written it straight forward, indicating that they were not fully aware of the goal of the experiments. This makes the results more trustworthy.

### 5.1 Tooltips

Below is a graphical presentation of the number of opened tooltips throughout the 22 cases. Normal Value represent the opened tooltips of normal values, while Explanation represent the opened tooltips of explanations. The x-axis shows the cases, while the y-axis represent the total number of opened tooltips for all participants. A trendline was added to better see the development from the first to the last case (Fig. 5).

The graph below shows that both normal values and explanation have a decrease in number of opened tooltips, normal values being slightly lower. This corresponds with what we learned from the post-interview, that the participants used the tooltips a lot in the beginning and less during the last cases. There is no significant difference between the two.

Through the post-interviews, we found that most of the participants confirmed that they used the tooltips less throughout the cases, because they had learned them by heart. This also corresponds with several of our results from the UX questionnaire, where the participants gave a 4.5 out of 5, on both “The need for opening the tooltips were less as the days went by” and “The tooltips helped me learn medical terms by heart”. One of them even quoted the tooltip about eclampsia, proving that she really had learned the term. Another said that she “check with the information I got earlier”,

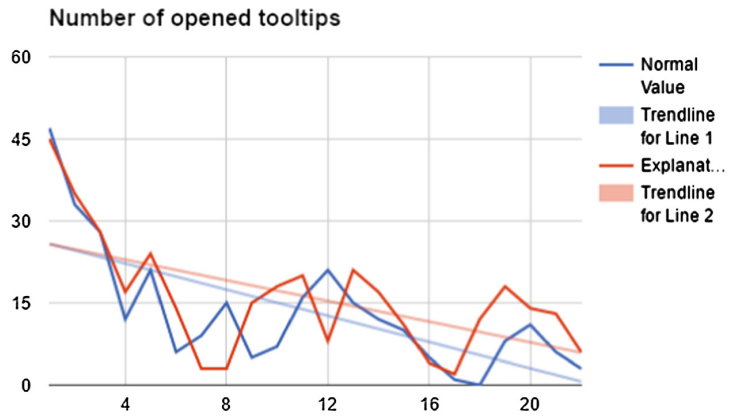


**Fig. 5.** Graph displaying opened tooltips throughout the cases

and further explained that she kept learning the terms when she opened the tooltips, and eventually she knew what to answer, without using them. One participant said she used the tooltips frequently in the first cases, but "Not frequently in the last cases because they helped us understand what it was.".

Another thing we noticed in the recordings, was that the tooltips were mostly used during the Previous Pregnancy stage, which may be because this is the first stage they enter information into. It may also be because pregnancies have different outcomes, and, therefore, it may be more difficult to differentiate between the different outcomes or delivery methods. Thus, it would require more of a need to consult with the tooltips. When we asked the participants during the interview what they found difficult in the system, the different stillbirths during previous pregnancies was mentioned several times. The difference between antepartum stillbirth, intrapartum stillbirth and stillbirth of unknown timing was confusing. Some also said that several of the terms used in the previous pregnancies stage, are terms that are more familiar to fully educated nurses and midwives, and might be difficult for people with less education to understand. Some also suggested that in order for non-medical personnel to understand what data to enter, signs and symptoms should be listed. This corresponds with the responses we received from the questionnaire regarding content types, that normal values is the most preferred content type.

The graph below show the percentage of successful tooltips from first to last case. The percentage was found by dividing number of successful tooltips with all opened tooltips. Its representation is mostly the same as the graph above, except from the y-axis, which represent the percentage of successful tooltips (Fig. 6).

The graph show that the percentage of successful tooltips increase towards the last cases. Also, as seen, the tooltips containing explanations has both a higher percentage of successful tooltips, and a steeper increase through the cases, than normal values.

During the post-interviews we found out that eleven of the 15 participants claimed that they open the tooltips first, and then enter the information. The last four entered

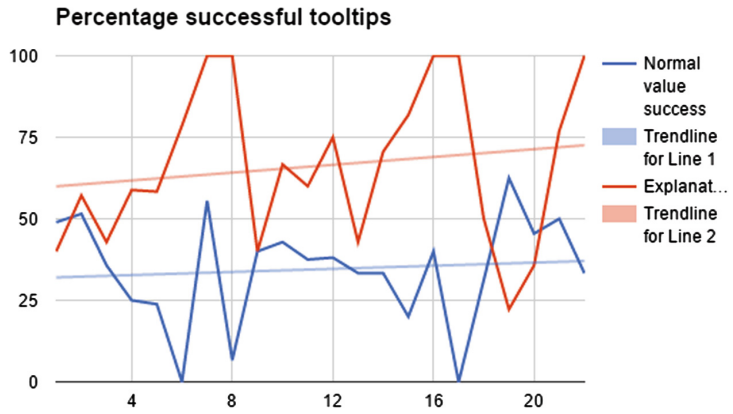


**Fig. 6.** Graph displaying percentage of successful tooltips throughout the cases

data first, and then used the tooltips to check the information they entered and to confirm their answer. We also found out that they had discussed with each other, and other colleagues, during the experiment, when solving the cases.

In addition to the interviews, we also used the booklet to find out what the participants thought. All of them wrote comments and thoughts for most of the cases, and also about the system and some of the tooltips they found useful. "I used the (i) to give me the meaning of the things or terms used" and similar comment are found in several of the booklets. A majority of the participants learned about gravidity and parity, and the different stillbirths. Especially did we notice that if the correct data entry was antepartum stillbirth, intrapartum stillbirth was quite often opened as well. "I learned the difference between antepartum and intrapartum stillbirth" one of the participants said. She often opened both tooltips to understand the difference between them. Also, we learned that ways of delivery contributed to learning. "The allow guided me on breech delivery" is a quote from one of the booklets, saying that the "allows", meaning the tooltips, taught her about breech delivery, something we also discussed during the interview.

Also, the tooltips for hypertension, pre-eclampsia and eclampsia were used more in the previous pregnancies stage. This was their first encounter with those tooltips during each case, and many of the participants found the terms confusing. We also found out that participants have different definitions of some terms, like for example pre-eclampsia. Some do not consider only protein in urine as a way of diagnosing pre-eclampsia, as it can indicate other diseases. Another interviewee said that "In our facility we don't have a lot of resources, so high BP means pre-eclampsia.", meaning they diagnose pre-eclampsia only based on high blood pressure. It is important to have formal definitions, however, it is absolutely vital to take into consideration the health facilities without the necessary resources for diagnosing certain conditions.

When analyzing the booklets and the post-interviews, several suggestion of improvement materialized. One participant suggested that we should add more vital signs to the data elements, another stated "Add more information to the i's. For example, can you have pre-eclampsia with only hypertension?". A third participant

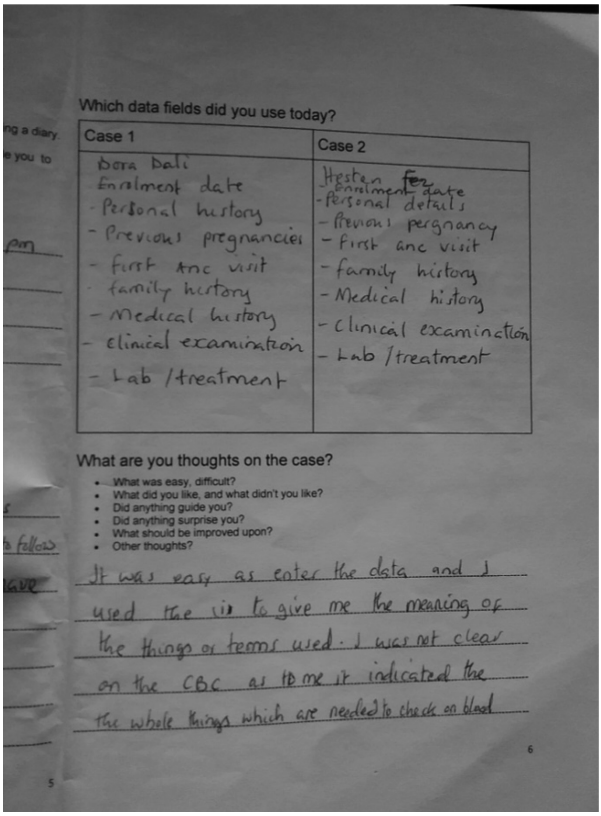


**Fig. 7.** An example from the tasks in the booklet

suggested that we should "for instance giving the normal ranges for BP". A fourth participant suggested signs and symptoms instead of formal definitions. She justified the statement by saying that non-medical personnel would not know what a condition is, based on the explanations. What is interesting is that all these participants had been using the testing program containing explanations as their content type for tooltips. These finding are also cohesive with the response from the UX questionnaire, where the following statements, "..should have provided more information.." and "..should have provided different information" received scores of 3.2 and 2.9 out of 5, indicating that they partly agree with the statements.

The chart below shows a scatter plot of the number of opened tooltips per user (x-axis) and % correct data (y-axis). Each dot represents a participant (Fig. 8).

There seems to be two users never or seldom opening tooltips who nevertheless enter data of with a high percentage of correctness (upper left). One of these was a nurse, who was sufficiently educated and outside the target group for the tooltips. Two other nurses participated.

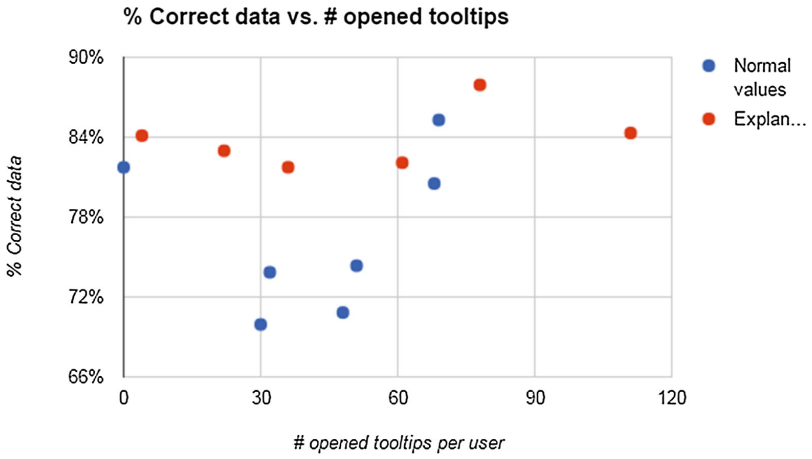


**Fig. 8.** A scatter plot of number of opened tooltips per user and % correct data

The other participants were scattered more linearly. A weak correlation between the number of opened tooltips and correct data entry was found (Pearson, r = 0.26). For successful tooltips correlated with correct data entry, r = 0.35, hence a moderate correlation.

## 5.2 Normal Values Versus Explanations

The graph below represents the correctness of data entry in all the cases (Fig. 9). The x-axis is the same as in the graph under "Tooltips", the cases, while y-axis is the correctness, measured in percent per case. Also here, a trendline was added in order to get a better view of the development from the first to the last case.

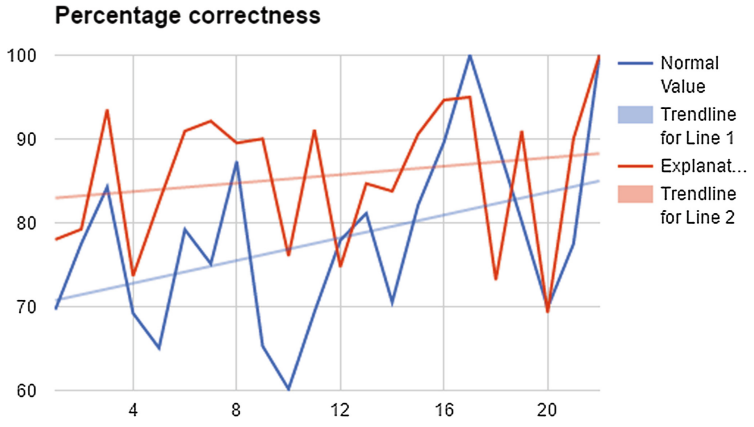


**Fig. 9.** A graph displaying the percentage of correctness (Color figure online)

The graph shows that explanations (red line) clearly start out with a higher percentage of correctness compared to normal values. However, if we look at the normal values (blue line), we can see that it increases faster than explanations. This may indicate that if the experiment had lasted over a longer period of time, normal values would have approached 100% correctness before explanations.

The scatter plot above shows that the users who received Normal value tooltips, performed less well than the Explanation group (means 77% vs 85%). Although the number of users is small, their individual scores are averaged over 22 cases. We therefore used the T-test (two-sided, two-sample), and it came out with a significant difference ($p = 0.01$) between the two groups.

To check possible statistically significant change of performance over time, the 22 cases were divided into three portions; the first seven, the eight middle and the last seven cases. Then the number of correct data entries in the first seven were averaged per participant and also for the last seven cases. The table below shows the mean values of correct data entry (Table 2).

**Table 2.** Results on correct data entry from logging use

| | Average % correct first 7 | Average % correct last 7 |
|---|---|---|
| Normal values (n=7) | 76 | 87 |
| Explanations (n=6) | 83 | 85 |
| All participants | 79 | 86 |

The difference in correctness between normal value tooltips (76%) and explanations (83%) is significant for the first seven cases (T-test, two-sample, equal variance) (yellow). Since the improvement for Normal values is stronger than for Explanations, the study cannot conclude about the long term effect.

The T-test (two sided, paired) shows a significant ($p = 0.04$) difference between the first and the last seven for the normal values (grey). Thus, the normal value group had fewer correct data entries in the beginning, but in the end of the 22 cases, they were at an insignificantly higher level than the Explanation group. This may be because normal values started out with less correct answers than explanations, and may therefore have "more room to grow".

There is also a significant difference between the first and last seven cases for the total group ($p = 0.03$). Normally, people improve their performance through repetitions. Our study was not designed with a placebo to differentiate effects of tooltips vs. no tooltips. Therefore, we cannot state that a particular percentage of the improvements followed tooltip use.

However, the interviews indicate that some of these improvements are due to tooltips, which is also cohesive with the UX questionnaire. "The tooltips helped answer correctly to the tasks given" received a total of 4.7 out of 5, meaning that they strongly

agree with the statement. Also, there was a low correlation between opening of tooltips and correct responses (Pearson r = 0.26). The difference in performance between the Normal value and Explanation tooltips groups shows that the tooltips had effects. We therefore conclude that tooltips caused improvement in correct data entry.

Our usage of similar terms both in the cases and in tooltips containing explanations may have influenced the results of the experiment. This may be part of the reasons why the participants using the tooltips containing explanation had a higher correctness and higher percentage of successful tooltips, as they more easy could recognize the phrases used.

## 6 Conclusion and Further Research

The goal of this research was to find out whether tooltips helped users entering correct data and whether specific contents for tooltips were better than other. The study comprises an experiment with 30 users, where all their use of the software was logged and the participants were interviewed after completion. At the time of final paper submission, only 15 of the participants had completed the experiment, thus only the results for these 15 have been included in the paper. The results may therefore change after all participants have completed, and the final results will be presented during the conference.

Isaksen et al. [5] identified normal data values as the most preferred content type for tooltips for data fields. Formal explanations was the second most preferred type. Previous studies of tooltips [1, 7] have also come up with preferences and have not tested effects of long term use.

This study therefore compared the two types of tooltips during a two weeks experiment.

The explanations group had a higher percentage of successful tooltips, meaning instances of opening a tooltip and entering a correct value, possibly in the opposite sequence. The explanations group also had a steeper increase than normal values. We also found that, in terms of the correctness in data, explanations have a higher percentage. However, correctness for normal values increase faster, and after two weeks, the normal value group was slightly ahead of the explanations on correctness. When comparing the first seven cases with the last seven, we found that tooltips containing normal values has a significant increase in correctness. The difference in correctness between explanations and normal values for the last seven cases is insignificant, as is the increase in the explanations group.

Thus, we see no correlation between user preference and the usefulness of the different content types. In addition, the UX questionnaire revealed that the participants found the tooltips both helpful and understandable.

Both normal values and explanation has a decrease in number of opened tooltips from the first to the last case. The difference between them is not significant. This is also consistent with what we learned through our post-interviews, as participants told us that they did not need the tooltips at the end of the experiment, as the information was learned by heart. This is consistent in the increase in the percentage of successful tooltips from first to last case.

An unexpected finding was that users also opened tooltips after they had entered the data. During post-interviews, they said that this was in order to check that they had entered data correctly. This way of learning from tooltips has not been mentioned in previous user studies of tooltips [1, 7].

We also learned that they used tooltips more during the previous pregnancy stage, which was probably due to it being the first encounter with the terms, difficulties in differentiating the pregnancy outcomes, or because the terms are more used by nurses and midwives.

In order to increase the validity of the experiment, we could have included a control group of participants. Here, the aim would have been to compare the effects of a system with tooltips and a system without tooltips. This is similar to research on medication, where one group is given real medicine, while the other is given placebo medication. However, the comparison between the two groups would not have been symmetric, as one group would have been introduced to tooltips and the other group not. An alternative way could be create a testing program with some meaningless tooltips. This would have made the groups more symmetric, giving one group actual tooltips and the other group "placebo-tooltips".

**Acknowledgment.** This research has been supported by QU Horizon 2020 "mHealth4Afrika - Community-based ICT for Maternal Healthcare in Africa" (project 668015, topic ICT-39-2015), Norwegian Centre for International Cooperation in Education "Scholarly Health Informatics Learning" (UTF-2016-longterm/10032) and Norwegian Agency for Development Cooperation "Support to the Health Informations Systsem Project - HISP" (QZA-14/0337).

## References

1. Dai, Y., Karalis, G., Kawas, S., Olsen, C.: Tipper: contextual tooltips that provide seniors with clear, reliable help for web tasks. In: CHI 2015 Extended Abstracts, pp. 1773–1778 (2015)
2. Duh, H.B.L., Tan, G.B.C., Chen, V.H.H.: Usability evaluation for mobile device: a comparison of laboratory and field tests. In: Proceedings of the 8th Conference on Human-Computer Interaction with Mobile Devices and Services, Helsinki, Finland, pp. 181–186 (2006)
3. Flaherty, K.: Diary Studies: Understanding Long-Term User Behaviour and Experiences. https://www.nngroup.com/articles/diary-studies/
4. Grossman, T., Fitzmaurice, G.: ToolClips: an investigation of contextual video assistance for functionality understanding. In: ACM Conference on Human Factors in Computing Systems 10, Atlanta, Georgia, USA, pp. 1515–1524. ACM (2010)
5. Isaksen, H., Iversen, M., Kaasbøll, J., Kanjo, C.: Design of tooltips for health data. Submitted for Publication
6. Michelsen, C.D., Dominick, W.D., Urban, J.E.: A methodology for the objective evaluation of the user/system interfaces of the MADAM system using software engineering principles. In: ACM Southeast Regional Conference, pp. 103–109 (1980)
7. Petrie, H., Fisher, W., Weimann, K., Weber, G.: Augmenting icons for deaf computer users. In: CHI 2004 Extended Abstracts on Human Factors in Computing Systems, Vienna, Austria, pp. 1131–1134 (2004)