# Communicating Uncertainty in Fertility Prognosis

**Hanna Schneider**
LMU Munich
hanna.schneider@ifi.lmu.de

**Julia Wayrauther**
LMU Munich
julia.wayrauther@gmail.com

**Mariam Hassib**
Bundeswehr University Munich
mariam.hassib@unibw.de

**Andreas Butz**
LMU Munich
andreas.butz@ifi.lmu.de

## ABSTRACT

Communicating uncertainty has been shown to provide positive effects on user understanding and decision-making. Surprisingly however, most personal health tracking applications fail to disclose the accuracy of their measurements and predictions. In the case of fertility tracking applications (FTAs), inaccurate predictions have already led to numerous unwanted pregnancies and law suits. However, integrating uncertainty into FTAs is challenging: Prediction accuracy is hard to understand and communicate, and its effect on users' trust and behavior is not well understood. We created a prototype for uncertainty visualizations for FTAs and evaluated it in a four-week field study with real users and their own data (N=9). Our results uncover far-reaching effects of communicating uncertainty: For example, users interpreted prediction accuracy as a proxy for their cycle health and as a security indicator for contraception. Displaying predicted and detected fertile phases next to each other helped users to understand uncertainty without negative emotional effects.

## CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**; Information visualization; • **Applied computing →** Health informatics;

## KEYWORDS

Fertility tracking applications; uncertainty visualization; menstrual cycle; personal informatics; women's health.


*CHI 2019, May 4–9, 2019, Glasgow, Scotland UK*

ACM ISBN 978-1-4503-5970-2/19/05…$15.00
https://doi.org/10.1145/3290605.3300391

**ACM Reference Format:**
Hanna Schneider, Julia Wayrauther, Mariam Hassib, and Andreas Butz. 2019. Communicating Uncertainty in Fertility Prognosis. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland UK*. ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3290605.3300391

## 1 INTRODUCTION

During the last years, many new FTAs have entered the market and gained significant popularity, for example Glow[1], Clue[2], and Natural Cycles[3] [10, 14]. At the same time, the use of exactly such applications has led to numerous unwanted pregnancies and law suits. By now, both researchers [3, 9, 34] and public press are voicing their concerns[4]. On the positive side, FTAs might be a more effective contraceptive than traditional fertility-awareness-based methods [2] allowing more women to prevent (or seek) pregnancies without taking medication (mainly to avoid side effects) and to learn about their body [10, 14]. Previous research in both digital health [3, 14, 34] and HCI [10] has highlighted the importance of communicating the (limited) accuracy of FTA predictions, which depends on various factors, such as the amount of data, cycle stability and consistency of lifestyle. However, none of the applications that we reviewed for this paper seems to communicate prediction accuracy beyond a general statement, e.g., on their websites. Communicating individually calculated accuracy poses several challenges: Prediction algorithms and accuracy can be hard to understand and communicate and – as the subject of female cycles and fertility is emotionally charged – they might trigger hopes and concern of varying nature. This paper explores these challenges.

The benefits of communicating uncertainty have been well demonstrated in HCI [16, 17, 22] and beyond (see, e.g., [1, 11, 15]). However, uncertainty visualizations in research were rarely tested in the wild with real user data (with the notable exception of work by Shaer et al. [35]). Our work therefore

[1]https://glowing.com
[2]https://helloclue.com
[3]https://www.naturalcycles.com
[4]https://www.theguardian.com/technology/2018/jan/17/birth-control-app-natural-cycle-pregnancies

expands research on uncertainty visualization with rich accounts of users' perceptions and emotions when interacting with personal uncertainty visualizations in their daily lives. The contributions of this paper are threefold:

(1) We present a prototype for displaying fertility data including uncertainty visualization
(2) We report findings from a four-week field study of a functional prototype with real users and their own data (N=9)
(3) We propose five design recommendations for visualizing uncertainty in personal (health) data

## 2 RELATED WORK

This research project builds on two bodies of work: (1) work on self-tracking technology, especially (the limited) work on fertility tracking; and (2) work on uncertainty visualization in HCI. Below, we will briefly summarize the two, highlighting how previous findings informed our approach.

### Self- and Fertility Tracking in HCI

Self-tracking technology has been a subject of increased public and research interest in the last years [20, 23]. However, despite its long tradition fertility tracking has only recently caught HCI researchers' attention. Fertility tracking or cycle tracking refers to observing the events and symptoms related to a woman's fertility cycle. While women have recorded their period and connected symptoms for many years, analogue notes have recently been replaced by smart-phone applications that not only record this information but use it to predict periods and symptoms in the future. Epstein et al. [10] investigated women's motivations to track their experiences with different tracking tools. They found that the accuracy of predictions was the most important criterion for women in menstruation tracking applications. Because this need is currently not met by applications on the market, Epstein et al. [10] suggest that "For both ovulation and period arrival, designers should consider and evaluate interfaces that present probabilities as an alternative to unreliable binary predictions." Moreover, in a recently published study, Gambier-Ross et al. [14] noted that women are interested "...in learning how the prediction methods work to see if they were personalized based on their own data or whether they were just based on a generic average cycle prediction". Their study also confirmed that accuracy and trust are among the most important qualities of FTAs. In a related study, Figueiredo et al. [6] analyzed publicly available posts in an online health community dedicated to fertility in the US. They report that for many women the topic is bound up with taboos, social pressure, frustration and the feeling of being inadequate. They stress that women require complex knowledge in order to understand the fertility cycle, fertility indicators and available tools and treatments. Moreover, Figueiredo et al. [6] highlight the inherent natural uncertainties of fertility care which include unknown health issues of reproductive organs, the vagueness of fertility indicators, and their subjective judgment. Building on this research, we investigate further how the uncertainty inherent in fertility tracking and prognosis can better be communicated to users while allowing them to make sense of the information in relation to their personal situation and goals. To the best of our knowledge, this is the first investigation of uncertainty communication in fertility tracking. While communicating probabilities seems especially relevant for fertility tracking, other self-tracking applications might benefit from it as well. Several researchers have noted that users tend to overtrust tracked data to the extent that they ascribe more importance to it than to their physical feeling [19, 25]. This can be harmful as sensors only track (often quite simple) selected metrics and then systems or users infer physiological or psychological states (e.g. health, fitness, mood, stress, fertility). As a system can also wrongly infer that a user is stressed, unhealthy, or infertile, some skepticism towards such systems can be helpful to avoid unnecessary worry or concerns and wrong decisions. The approach taken in this research project – communicating probabilities – might, hence, be interesting for other self-tracking applications, too.

### Uncertainty in HCI

To design effective uncertainty visualizations, we reviewed work on uncertainty within and beyond HCI. As visualizing uncertainty has a long tradition in research, we focus this summary on findings that informed our approach. Importantly, definitions and taxonomies of uncertainty vary between and within different fields of research and often overlap with related concepts such as accuracy and reliability (e.g., [16, 21, 27, 30, 36, 37]). Within our project, we focused on prediction accuracy and conception probability, even though there are certainly more uncertainties involved in fertility tracking (see section 3).

Generally, the importance of communicating accuracy, completeness, consistency and certainty to facilitate proper understanding of data and informed decision-making is well-documented [1, 11, 15]. Related work on uncertainty visualization and probabilities, e.g., in cartography [1, 11, 26], meteorology [7, 28], health [35], computer vision [30], and risk management [24] evaluated a wide range of visual variables to communicate different levels of uncertainty: Results include that bar charts and pie charts were best suited to represent the magnitude of probabilities and proportions, while line charts are best for revealing trends [24, 36]. Moreover, fuzziness, location, color, size, arrangement, and transparency were found to perform well in visualizing uncertainty [27], with fading color and transparency being especially suitable for temporal uncertainty [18].

Recently, several researchers have investigated different approaches to communicate uncertainty in the field of HCI, for example of genomic data [35], bus arrival estimations [22], and weather forecasts [16]. For the purpose of developing a prototype to communicate uncertainty in bus arrivals, Kay et al. [22] presented several design requirements for uncertainty visualizations: uncertainty should be intrinsic to the representation; the visualization should allow users to apply situation-dependent risk tolerance; and it should preferably be framed in terms of discrete outcomes or frequencies.

A large body of work on Personal Informatics and self-tracking (e.g., sleep [4] or physical activity [32]) explored how to present personal and often uncertain data. However, uncertainty was rarely explored explicitly and systematically in this context. A notable exception in the personal health data visualization space is a tool to view personal genomic data presented by Shaer et al. [35]. These data share some characteristics with fertility data: they are complex and sensitive, involve multiple dimensions of uncertainty, and can have substantial implications for the individuals' well-being [6].

In our research project, we build on these findings and apply them to the domain of fertility tracking and forecasts. To study the effects of uncertainty information in this domain, we developed a prototype and tested it in a four-week field study. Our findings contribute new insights about the usefulness and potential risks and benefits of uncertainty visualizations that are specific to the sensitive context of personal health and family planning.

## 3 FIELD STUDY

To better understand the effects of communicating uncertainty in fertility tracking, we conducted a mixed-method field study with users' real data. Below, we first explain the uncertainty data communicated in our prototype (see figure 1), our visualization design and its limitations, and the study's method and results.

### Uncertainty Data

Fertility data and predictions inherit various uncertainties depending on the underlying data [12, 29]. Most commercial applications use manually entered data such as past periods, in some cases manually measured temperature, and additional symptoms. The analysis of the basal body temperature (measured in the morning) is one of the most popular fertility-awareness-based methods [29, 33]. Several technologies continuously measure temperature, freeing women of the burden to remember, measure, and record every day. For this study, we cooperated with one company that relies on this new approach. The company VivoSensMedical provides a temperature sensor called the OvulaRing[5] that is worn vaginally throughout the cycle where it measures the core body temperature every five minutes to detect a temperature shift that indicates ovulation. This technology is state-of-the-art with studies showing that the ovulation and thus the fertile window are identified with an accuracy of 99.11% retrospectively and 88.8% prospectively [31]. For this project, we used the data provided by their sensing technology and the predictions of their algorithm. Compared to relying on women recording their period and inferring ovulation based on population-generic statistics, this technology provides an accurate account of past ovulations and hence more accurate predictions. Specifically, we integrated the following data provided by VivoSensMedical in our visualizations:

*detected ovulation:* past ovulation days detected by the OvulaRing algorithm based on personal temperature pattern.

*predicted ovulation:* predicted ovulation days calculated by the OvulaRing algorithm based on past detected ovulations.

*prediction accuracy:* accuracy of the predicted ovulation day as probability ranging from 0 to 100%, which is calculated by the OvulaRing algorithm based on fluctuations within past ovulation patterns. In the data used for our research project, the prediction accuracy refers to the estimation of the ovulation day, which is in turn used to determine the fertile phase. The prediction accuracy does therefore not change within one cycle or fertile phase.

*conception probability:* probability of conceiving on a specific day given the day of ovulation. It ranges from 0 to 30% and is based on a statistical distribution inferred from scientific studies. This probability distribution accounts for biological uncertainty about the exact lifetime of egg cell and sperm and is commonly referred to as the fertile phase. Predicted ovulation (with respective fertile phase and prediction accuracy) and conception probability together present the uncertainty information that we integrated in our prototype. While there are uncertainties involved in fertility tracking that go beyond these metrics (e.g. *individual* lifespan of eggs and sperm), quantifying these is currently beyond the technologies' capabilities.

### Prototype Development

Below, we briefly introduce the interactive prototype visualizing personal cycle data, which we developed for our field study in several iterations of sketching, paper prototyping, and pilot testing. To provide realistic insights, the prototype

[5]http://ovularing.com/

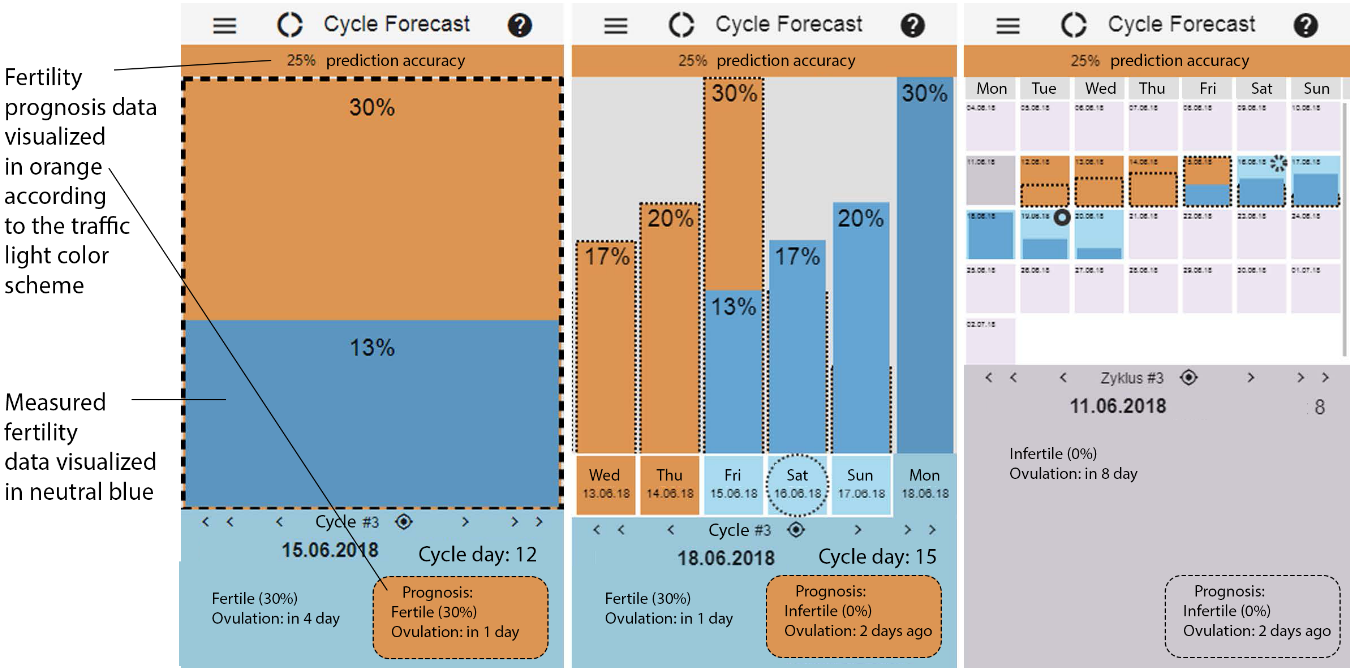


***Figure 1: Day, timeline, and calendar view showing prognosis data (orange) and detected fertile phase (blue)***

was tailored to visualize data collected with the OvulaRing sensing technology and was developed only in German (as the participant sample recruited by VivoSensMedical was German speaking). The final prototype was implemented as a web application with the MEAN[6] software stack and hosted privately. We received participants' data from VivoSensMedical and manually integrated it in the prototype. The data stored was completely anonymized to prevent any identification of the participants.

*Visualization and Interface.* The prototype featured three main views displaying users' personal fertility data (see figure 1): a day view, a timeline view, and a calendar view. We selected these views as they are common and widely known layouts for cycle visualizations[7] and because they performed best in our pilot testings. To integrate the uncertainty information in these views, we reviewed potential visual variables or marks to encode prediction accuracy and conception probability from related work. As a result, we considered the following options: varying color value, mapping to the traffic light color scale, or using bar charts, line charts, pie charts, and size. We used the above mentioned pilot testing to determine one visual mark for each uncertainty. Based on the results we decided to display conception probability as bar chart and to encode prediction accuracy using the traffic light color scale (see figure 1). Notably, we decided to use probabilities in our prototype, even though frequencies were found to be generally easier to grasp and interpret [21, 22] (see limitation section).

In each view, the uncertainties were displayed with visual elements and as exact probabilistic numbers: prediction accuracy in a separate view element below the header (colored according to the traffic light color of the prediction accuracy value) and the conception probability in the bars. Furthermore, we added a detail panel (lower part of each view) that provides detailed information about the currently selected day (highlighted with darker color): the conception probability, date and day of the week, the number of the current cycle and cycle day, and the fertility status (either "infertile" if the conception probability is 0% on the selected day or "fertile" otherwise). Instead of replacing predicted ovulation (and fertile phase) with detected ovulation (if ovulation has been detected), we decided to display both next to each other in past cycles as their co-existence illustrates the uncertainty of predictions as well. To distinguish predicted ovulation and detected ovulation visually, we used a dashed outline for the predicted ovulation and colored bars of detected ovulation in neutral blue. Arrows on top of the details panel allow users to navigate forward and backward in time (e.g., to the next day/cycle). To switch between the three visualizations users need to open a menu in the side panel by clicking on a button in the upper left corner. Finally, we included a help page that could be accessed by clicking on the question mark

[6]http://mean.io
[7]Similar views are, e.g., used by *OvulaRing*, *OvuView* or *Clue*.

icon on the very right of the header and addressed the differentiation between real and predicted data and explained how to interpret the prediction accuracy and the conception probability.

*Limitations.* The focus of our study was on understanding reactions to uncertainty in the domain of fertility tracking (rather than innovating uncertainty visualizations per se). While our prototype served our purposes well in the context of our study, we are aware of two shortcomings of our visualizations that need to be addressed by future work:

First, the visualization of predicted and detected ovulation data in the past (on top of each other) is not ideal: When predicted and detected fertile phases are exactly identical, the only visual cue of the prognosis is the dashed outline, a view which is clearly unfavorable. This is because, we obtained the permission to display all past ovulation data of a user after the prototyping phase. It is often that real-world case studies involving multiple stakeholders are exposed to unforeseeable circumstances and opportunities, especially in the health domain [13]. While we were aware that the visualization can certainly be improved, we saw an opportunity to learn about users' reactions by displaying some visualization of predicted data in the past. Nevertheless, based on the analysis of our interview data and users' journals, we are certain that participants understood the visualization.

Second, our visualizations might be improved by using a frequency framing, for example, implemented as icon arrays. In our project, we received uncertainty data specified as probabilities and displayed it as such. While it is possible to convert probabilities into frequencies, e.g. via draws from a probability distribution (see [22]), we decided against icon arrays because we considered probabilities easier to implement and sufficient for our study purpose. However, integrating a frequency framing in the different data views while preserving clarity (e.g. using icon arrays in the day view, low-density dotplots [22] in the timeline view, and icon arrays in the calendar view) might be a promising opportunity for future work to refine uncertainty visualizations of fertility data.

### Method

The prototype was evaluated in a four-week field test. We collected data on participants' experience and behavior in pre- and post-study interviews as well as a diary study. To detect and counteract any problems, participants were contacted on a weekly basis to check if they had questions and to motivate them to use both the prototype and the diary. Below, we briefly explain pre- and post-study interviews as well as the diary study method:

*Pre-Study Interviews.* The semi-structured pre-study interviews took between 30 and 60 minutes (mean: 56 minutes and 36 seconds) and followed an interview guideline that included five parts: (1) an introduction of the researcher and the study, (2) a block of interview questions including participant's personal context and demographic information, cycle and cycle tracking experience; tracking goal and their perceptions of and trust in prediction, (3) an explanation of the study procedure, (4) a walk-through of the prototype website (making clear that it is an early stage prototype) including an explanation of prediction accuracy and conception probability and their visualization in the prototype; participants had the chance to ask questions right after and were encouraged to contact the researcher at any time, (5) a walk-through of the diary document. The interviews were conducted online via Skype or appear.in and audio-recorded, transcribed and coded applying the thematic analysis approach [5].

*Prototype Usage: Diary Study.* During the four-week study the participants could access their own fertility data via the prototype (in addition to accessing it via the OvulaRing website as usual). Moreover, they had the option to check real-time conception probability based on the temperature pattern of the current month – a new feature released by VivoSensMedical shortly before the study. We provided participants with an (optionally analog or digital) diary and instructed them to create an entry for each time they looked at their data (preferably on the prototype if possible). The input fields of the diary were beginnings of sentences that the participant could complete, (e.g., "Today, I used the prototype to..."). These sentences aimed to elicit the participant's reason for the visit, interesting or confusing insights gained from the presented data, perceptions of the prediction accuracy and conception probability, as well as perceptions of the current reliability of the prognosis. Additionally, there was space for any other comments. After four weeks the diaries were sent back to the study leader in order to inform the post-study interview questions.

*Post-Study Interviews.* Post-study interviews took on average one hour and covered six topics: (a) experiences made during the study, (b) knowledge gained, (c) prototype website usage (participants' diaries were used to refer to specific usage instances; vague/ unclear entries were clarified), (d) usefulness of available cycle information and risk perception, (e) reliability of cycle data and tracking technology, (f) usability of the prototype website. Again, the dialog was audio-recorded, transcribed and coded immediately after each interview. The results were then used to iteratively adjust the interview guidelines for subsequent interviews. Participants' quotes in this paper were translated from German.

*Ethics.* The study procedure complied with university ethics regulations. In addition, the recruitment and data handling was accompanied by VivoSensMedical and we recruited participants from their study volunteers pool. Participants had tracked their cycle for at least one and at most 50 cycles (see table 1). Prior to the study the OvulaRing system showed predicted ovulation data in the future as well as detected ovulation data in the past with no uncertainty information. Predicted data disappeared when ovulation was detected. Potential participants received a personal explanation of the study purpose, procedure and the uncertainty prototype. After they opted in and signed informed consent, they received personal access credentials that allowed them to view their OvulaRing data via the mobile study prototype website.

| | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 |
|---|---|---|---|---|---|---|---|---|---|
| Tracked Cycles | 7 | 18 | 1 | 2 | 17 | 13 | 22 | 4 | 50 |
| Months | 7 | 14 | 1 | 3 | 15 | 11 | 21 | 6 | 51 |

***Table 1: Prior experience with the OvulaRing sensing technology prior to the field test in number of tracked cycles and in months.***

| Goal | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 |
|---|---|---|---|---|---|---|---|---|---|
| Observing | ✓ | ✓ | | | (✓) | | | ✓ | ✓ |
| Contracep. | (✓) | (✓) | ✓ | | | ✓ | ✓ | | (✓) |
| Pregnancy | | | | ✓ | ✓ | | | (✓) | |

***Table 2: Cycle tracking goals by study participant; secondary goals are shown in brackets.***

## Participants

Nine women participated in the field study (aged between 22 and 43 years). Thereof, three women reported to use the sensor technology primarily for contraception, three were trying to become pregnant, and three primarily tracked to observe their body and secondarily to apply natural contraception (see Table 2).

## Results

Our study shed light on users' perceptions of and reactions to the uncertainty metrics and our visualizations. Below, we will describe results related to three aspects: (A) the performance of visual elements in communicating uncertainties, (B) participants' interpretations and emotions related to conception probability and prediction accuracy, (C) the impact of conception probability and prediction accuracy on participants' risk perception, trust, and behavior. Even though some of the presented results are not directly related to the designed visualizations, they provide important insights for the design of uncertainty visualizations.

*(A) Deviations between predicted and detected ovulation in past data were more effective in communicating uncertainty than prediction accuracy and traffic light color-coding.* To communicate uncertainty, we visualized the conception probability as bar charts and the prediction accuracy as traffic light color-coding (in addition to displaying the exact value). Moreover, we intentionally displayed both predicted and detected ovulation in past cycles. Below, we briefly summarize participants' perceptions of these visual elements:

*bar charts displaying conception probability.* Generally, the bar charts were well perceived and understood (in combination with explicit numbers). Most participants found the information interesting (P2, P4, P6, P7, P8, P9) and useful. For example, P8 who used the information to become pregnant said: "the better you understand it, the better you can target"(P8). Four participants were surprised by the general trend and the low maximum values of the conception probability (P2, P4, P7, P9). Nevertheless, some interpretation problems occurred: P3 and P8 were not aware that the conception probability was based on past studies. Lastly, P6 and P7 pointed out that the conception probability is not relevant for contraception, they only wanted to know if additional protection was needed or not.

*probabilities specifying prediction accuracy.* Two thirds of participants required an explanation in order to understand the current prediction accuracy and its changes over time. P1, P2, P5, and P6 wanted to know how the probability was calculated and why it changed: "Yes, that was really helpful to understand now that this is calculated based on past cycles"(P5). P6 and P8 confused the prediction accuracy with the conception probability: "Yes, I think at this point the two were still confused in my head: conception probability and prediction accuracy"(P3). Nevertheless, all participants found the prediction accuracy generally interesting and useful with P7 and P8 being slightly less interested. Lastly, P6 and P9 stressed the importance of the prediction accuracy for contraception purposes.

*encoding of prediction accuracy according to the traffic light color scale.* The traffic light color-coding was only noticed by four women during the study - despite an explanation in the beginning (P1, P2, P4, P8). While they appreciated its visual appeal, the fact that more than half did not recognize it suggests that the color-coding is not suitable to communicate the prediction accuracy on its own. P3 missed the option of a part-to-whole comparison since the whole color scale was only visible on the help page (due to limited screen space).

*deviation between predicted and detected ovulation.* Displaying both predicted and detected ovulation in past cycles showed to be a more effective way to increase users' awareness of potential uncertainties than the prediction accuracy. Seven participants appreciated the option to compare real and predicted data in past cycles (P1, P3, P4, P5, P6, P7, P8. P1 had noticed before the study that her cycle deviated from the prognosis and used the visualizations to examine these deviations further. For women who were not aware of the existence or the extent of possible deviations before (P3, P7), seeing the differences was surprising and helped to clarify and illustrate the meaning of prediction accuracy. P3, for example, commented on a deviation: "that would have made it really clear for the third cycle, this is really... it just does not coincide. And that's how the prediction accuracy comes about" (P3). In general, the timeline view visualized deviations better than the calendar view. In some cases, the relation or difference between predicted and detected ovulation was unclear – especially when predicted and actual fertile phases overlapped completely or not at all. In several cases, the combination of prediction accuracy and deviation between predicted and detected ovulation led to confusion: P5, for example, was confused by a large deviation of a fertile phase predicted with 90% prediction accuracy while a completely correct prognosis showed 47% prediction accuracy. The explanation that the prediction accuracy was calculated based on previous cycles, helped her to understand.

*(B) Prediction accuracy and conception probability affected participants' self-confidence and emotions.* Five participants drew conclusions on their health and fertility based on the prediction accuracy. Although these results are not directly related to uncertainty visualization design, they can help to inform it: On the positive side, P2 and P9 interpreted a high prediction accuracy of 90% and 79% as signs of high regularity and fertility leaving them with a positive feeling. Similarly, P4 (using the technology to become pregnant) also related the increase of prediction accuracy from 20% to 30% to her cycle becoming more healthy: "That gave me a bit of hope again, I thought it was great." She also felt that her chances for conception improved as the rise in prediction accuracy reflected for her "this being healthy". While a further increase in prediction accuracy would increase her self-confidence, she would experience a decrease as "a slap in the face" (P4). Similarly, P2 feared that a low prediction accuracy could cause pressure and the feeling of something being wrong. For example, she had experienced one cycle without ovulation as a shock. Thus, she imagined that women with a very irregular cycle might feel bad about seeing it reflected in a low prediction accuracy. In contrast, P8 stated that she had accepted her irregular and hard-to-predict cycle a long time ago and thus the low prediction accuracy of 20% did not affect her emotionally. Two women (P2, P4) stated that changes in color (based on the traffic light color scale) intensify the emotional impact of the prediction accuracy indicating that design elements can amplify or mitigate such feelings.

Prediction accuracy changes can also influence feelings of safety, especially of contraception users. P6 described that it made a difference for her whether the prognosis was 92% accurate or 75%. She explained that even though a low prediction accuracy was reason for concerns, "having the information at the same time decreases my worries again". P9 was shocked when viewing low prediction accuracy levels of for example 47% or 33% in past cycles: "At that time, I was using it for contraception, that's really crazy actually, that nothing happened but ok..." Further, she stated that she would only rely on a prognosis with at least 80% prediction accuracy in the future.

Conception probability affected participants' emotions to some extent as well: P5, for example, found it relaxing to know that the conception probability was at most 30%. She concluded that trying to become pregnant without success did not necessarily mean that something was wrong with her. However, P4 was discouraged by the same values as they decreased her hope of becoming pregnant in the near future. Conversely, the low conception probability of 17% raised P3's hope when she was fearing an unintentional pregnancy.

*(C) Prediction accuracy and deviations between predicted and detected ovulation affected participants' risk awareness and behavior.* The prediction accuracy and deviations between predicted and detected ovulation affected the risk awareness and behavior of some participants but not of others (P2, P5). In the latter group, P2 found the information nice to know, but it had no impact on her daily life. Similarly, P5 used the technology "more as a back-up" and to improve her body knowledge and P1, P2, and P5 stated that they did not rely solely on a prognosis before the study either. P2 stated: "Trust is good, control is better. [...] I rather trust the accompanying symptoms" (P2).

On the contrary, P7 stated to rely on the prognosis despite the prediction accuracy being lower than expected (30% instead of 95%) while using the technology for contraception. She explained: "that's my trust in the matter; I rely on it and fare very well with it for two years." (P7) However, seeing deviations between predicted and detected ovulation in past data did increase her risk awareness: "Yes, yes, it is dangerous then [laughs]" (P7). Similarly, observing deviations decreased trust in the prognosis for P1, P4, and P6 and impacted their behavior. As a consequence, P1 decided to investigate the cause of the deviation further and P4 and P6 checked the real-time conception probability (P4 to become pregnant and P6 for contraception). On this occasion, P4 found a positive conception probability (8 days before predicted ovulation and one day before detected ovulation) and intentionally got pregnant. While P6 felt safe with a prediction accuracy of 92%, a decrease to 70% caused her to "add more safety margin around the predicted fertile phase" (P6). Moreover, two participants stated that bigger deviation between predicted and detected ovulation or lower prediction accuracy might lead to a loss of interest in the prognosis (P4, P5).

## 4 DISCUSSION

Deciding if and how to integrate uncertainty information in health technologies is highly complex. Although, our study results cannot answer these questions conclusively, they shed light on several benefits and drawbacks of displaying uncertainty. Moreover, we propose several design recommendations for uncertainty visualizations in health technologies and beyond.

### Positive and Negative Impact of Accuracy Indicators

*Increasing Complexity.* The field of cycle tracking is very complex due to the biological processes involved [6]. Introducing uncertainty indicators adds to this complexity and poses new challenges for users. While some participants naturally started calculating the aggregated probability of conception for multiple cycles, others confused the meaning of the different probabilities. For example, in one case a user assumed that a prediction accuracy of 10% in combination with a conception probability of 17% indicated a low risk of pregnancy. Later she realized that a low prediction accuracy actually indicated a higher risk due to the increased uncertainty. Moreover, the lack of knowledge about the origin of the uncertainty information caused interpretation problems. One participant thought that the conception probability was based on her cycle data and not retrieved in medical studies. Other participants did not know that the prediction accuracy was based on the variance in the data of their completed cycles. Thus, they were not able to make sense of prediction accuracy changes and trends. However, we also found that most participants were interested in the additional uncertainty information and preferred to use the system with this information. Hence, making the meaning of uncertainty indicators understandable, tangible, and concrete is a challenge for future research and interaction designers.

*Knowledge Gain.* Most participants found it interesting to learn about the conception probability regardless of their cycle tracking goal, whether just out of curiosity or to optimize their attempts to become pregnant. Learning about the comparably high chances of a pregnancy on six days caused one women to stop focusing solely on the day of ovulation and to extend her conception attempts to the given time range. Similarly, users reported that the prediction accuracy and deviations between predicted and detected ovulation helped them to understand how the ovulation detection and the prognosis calculation actually worked. Hence, uncertainty indicators may help users to reflect on and better understand the technology they use, which may help them to develop appropriate expectations.

*Emotional Impact.* In the case of cycle tracking, displaying uncertainty indicators had a huge emotional impact on participants. The prediction accuracy was often interpreted as an indicator of cycle health – as a more regular cycle will lead to a higher accuracy. However, cycle health is associated with a strong emotional burden for many women, especially when it comes to the desire to have children [6]. Our study results showed that a decreasing accuracy can have a strong negative and an increasing accuracy a strong positive emotional impact on these women. Enhancing the prediction accuracy with a traffic light color coding amplified this effect. Based on these observations, we recommend to refrain from displaying metrics that are likely interpreted as "health index" or as judgmental. A less emotionally straining way to indicate uncertainty is to display deviations between predicted and detected ovulation in past data.

*Safety Indicator.* Some of the women using cycle tracking for contraception purposes interpreted the prediction accuracy as a safety indicator. For example, they mentioned personal prediction accuracy thresholds which had to be met for them to rely on the prognosis. Even though low accuracy caused feelings of insecurity, the study results suggest that women prefer having this information as it allows them to apply safety precautions. On the contrary, such indicators might also lead to contraception users relying fully on the predicted fertile phase when facing high accuracy, which might be reason for concern: After all the prognosis is a statistical measure and does not consider other sources of uncertainty such as real-time influences on the cycle including, e.g., stress or illness. For these reasons we would recommend to refrain from displaying accuracy in this form. Nevertheless, its ability to raise awareness of uncertainty, to create realistic expectations and inform responsible behaviour is promising. Future work needs to investigate how the usefulness of accuracy indicators can be maximized while limiting their negative effects.

*Users' Trust in Prognosis and Provider.* One reason to refrain from displaying uncertainty data for companies is that it might decrease trust in their company/brand. In contrast, we found in our study that uncertainty data decreased trust in the prognosis (a desirable result when avoiding unwanted pregnancies is the goal) but it increased trust in the provider company. When competitors feature technology that is seemingly infallible, the challenge will be to communicate the advantages of transparency and uncertainty data.

### Design Recommendations

We conclude this paper with several design recommendations for uncertainty visualizations of personal data such as fertility and health.

*Show Past or Future Errancy.* The displayed deviations between real and predicted data in completed cycles was easier for participants to grasp than the prediction accuracy and gave them a better idea of potential risks. In addition to that, a negative emotional impact and reduction of the perception of risk could not be observed. Furthermore, it helped them understand how the prognosis works. Hence, we recommend to display deviations or algorithmic errors in past data and to explore visualization techniques that show potential errors in the future.

*Use Visuals.* The use of additional visual representations of the uncertainties was appreciated by the study participants. According to them, the visuals used made the meaning of the probabilities clearer and even had an eye-opening effect for some. Especially the women who disliked working with numbers were fond of the visualizations. However, it is important to choose appropriate visual representations and here we see definite room for improvement by future work: While the conception probability bars in our study performed well, the color mapping of the prediction accuracy was barely recognized especially if the changes between cycles were small. Furthermore, as mentioned before, future work might focus on framing uncertainties in fertility tracking as frequencies (for example using icon arrays or low-density dotplots [22]) – as this technique has yielded promising results in other domains as well.

*Provide Goal-based Interpretation Aids.* Since the prototype targeted women with different cycle tracking goals generic interpretation aids were not possible. For example, contraception users should be made aware of low prediction accuracy levels by all means (e.g. red color or exclamation marks) whereas for women who try to conceive additional emphasis on low prediction accuracy levels would cause unnecessary stress. Hence, symbolic or verbal interpretation assistance need to be adapted to the user's tracking purpose. One participant would have preferred a message like "Attention! No reliable prediction possible" accompanying a low prediction accuracy. This could on the one hand reduce the cognitive load and prevent misinterpretation, but on the other hand the user is tempted to follow predefined recommendations instead of making her own informed decisions based on the data. Consequently, the interpretation aids should be subtle and rather assist in understanding the data than in making decisions based on it.

*Explain on the Spot.* The required knowledge for understanding the cycle data presented needs to be accessible for the user where it is required. Most of the participants did not access the help section of the prototype app even though they required additional information on the uncertainties. In addition to that, some who visited it could not find the information they required. Hence, finding out which knowledge is needed by the participants in order to interpret their cycle data has to be the first step. Thereafter, methods for effective communication can be elaborated. For desktop solutions such explanations on the spot could be integrated as informative overlays, explanatory figures or little animations triggered by hovering the concerning information. On mobile applications, separate buttons next to the referenced information might be necessary (like for example provided by Clue). How uncertainty information can be explained on demand is an interesting and important topic for future work in cycle tracking.

*Personalize.* Cycle tracking is highly personal as cycle characteristics and the occurrence of symptoms do not only vary between women but also from cycle to cycle [8, 14]. In our study, we found unsurprisingly that explanation and feedback were much more effective when they were based on their personal data. Ideally, explanation will also take users personal needs into account (e.g., emotionally involved, lack of knowledge...). Presumably, personalizing FTAs appropriately is not as simple as classifying the users by cycle tracking goal. FTAs with different modes tailored to the information and interpretation needs of specific user groups could be a potential solution.

## 5 CONCLUSION

We contribute an exploration of the design possibilities and consequences of communicating uncertainty in FTAs. Despite the increased complexity, most women appreciated receiving information on uncertainty. However, we also found that prediction accuracy was interpreted as representation of cycle health and caused women trying to conceive high levels of emotional stress.

Moreover, women using the technology for contraception might rely fully on the predicted fertile phase when faced with high prediction accuracy. Showing deviations between prognosis and detected ovulation in past cycle data helped users to understand uncertainties in predictions without causing negative emotional effects and reducing users' risk perception. Beyond showing algorithmic errancy in past data, we recommend to use visual representations for uncertainty, to integrate goal-based interpretation aids and uncertainty explanations on the spot that refer to personal cycle characteristics and are personalized to users' needs.

## 6 ACKNOWLEDGEMENTS

We would like to thank the company VivoSensMedical (especially Bettina Brammer) for supporting this research and everyone involved in the review process for their insightful comments and their help in improving this manuscript.

## REFERENCES


[1] Jeroen C. J. H. Aerts, Keith C. Clarke, and Alex D. Keuper. 2003. Testing Popular Visualization Techniques for Representing Model Uncertainty. *Cartography and Geographic Information Science* 30, 3 (2003), 249–261. https://doi.org/10.1559/152304003100011180 arXiv:https://doi.org/10.1559/152304003100011180

[2] E. Berglund Scherwitzl, O. Lundberg, H. Kopp Kallner, K. Gemzell Danielsson, J. Trussell, and R. Scherwitzl. 2017. Perfect-use and typical-use Pearl Index of a contraceptive mobile app. *Contraception* 96, 6 (2018/09/17 2017), 420–425. https://doi.org/10.1016/j.contraception.2017.08.014

[3] Simon Brown, Leonard F Blackwell, and Delwyn G Cooke. 2017. Online fertility monitoring: some of the issues. *International Journal of Open Information Technologies* 5, 4 (2017), 85–91.

[4] Eun Kyoung Choe, Bongshin Lee, Matthew Kay, Wanda Pratt, and Julie A. Kientz. 2015. SleepTight: Low-burden, Self-monitoring Technology for Capturing and Reflecting on Sleep Behaviors. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. ACM, New York, NY, USA, 121–132. https://doi.org/10.1145/2750858.2804266

[5] Victoria Clarke and Virginia Braun. 2013. Teaching thematic analysis: Overcoming challenges and developing strategies for effective learning. *The psychologist* 26, 2 (2013), 120–123.

[6] Mayara Costa Figueiredo, Clara Caldeira, Tera L. Reynolds, Sean Victory, Kai Zheng, and Yunan Chen. 2017. Self-Tracking for Fertility Care: Collaborative Support for a Highly Personalized Problem. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 36 (Dec. 2017), 21 pages. https://doi.org/10.1145/3134671

[7] National Research Council. 2006. *Completing the Forecast: Characterizing and Communicating Uncertainty for Better Decisions Using Weather and Climate Forecasts*. The National Academies Press, Washington, DC. https://doi.org/10.17226/11699

[8] Mitchell D. Creinin, Sharon Keverline, and Leslie A. Meyn. 2004. How regular is regular? An analysis of menstrual cycle regularity. *Contraception* 70, 4 (2018/09/17 2004), 289–292. https://doi.org/10.1016/j.contraception.2004.04.012

[9] Marguerite Duane, Alison Contreras, Elizabeth T. Jensen, and Amina White. 2016. The Performance of Fertility Awareness-based Method Apps Marketed to Avoid Pregnancy. *The Journal of the American Board of Family Medicine* 29, 4 (2016), 508–511. https://doi.org/10.3122/jabfm.2016.04.160022 arXiv:http://www.jabfm.org/content/29/4/508.full.pdf+html

[10] Daniel A. Epstein, Nicole B. Lee, Jennifer H. Kang, Elena Agapie, Jessica Schroeder, Laura R. Pina, James Fogarty, Julie A. Kientz, and Sean Munson. 2017. Examining Menstrual Tracking to Inform the Design of Personal Informatics Tools. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 6876–6888. https://doi.org/10.1145/3025453.3025635

[11] Beverley J. Evans. 1997. Dynamic display of spatial data-reliability: Does it benefit the map user? *Computers & Geosciences* 23, 4 (1997), 409 – 422. https://doi.org/10.1016/S0098-3004(97)00011-3 Exploratory Cartograpic Visualisation.

[12] Richard J. Fehring, Mary Schneider, and Kathleen Raviele. 2006. Variability in the Phases of the Menstrual Cycle. *Journal of Obstetric, Gynecologic & Neonatal Nursing* 35, 3 (2006), 376 – 384. https://doi.org/10.1111/j.1552-6909.2006.00051.x

[13] Dominic Furniss, Rebecca Randell, Aisling Ann O'Kane, Svetlena Taneva, Helena Mentis, and Ann Blandford. 2014. Fieldwork for Healthcare: Guidance for Investigating Human Factors in Computing Systems. *Synthesis Lectures on Assistive, Rehabilitative, and Health-Preserving Technologies* 2, 1 (2014), 1–146. https://doi.org/10.2200/S00606ED1V02Y201410ARH007 arXiv:https://doi.org/10.2200/S00606ED1V02Y201410ARH007

[14] Katie Gambier-Ross, David J. McLernon, and Heather M. Morgan. 2018. A mixed methods exploratory study of women's relationships with and uses of fertility tracking apps. *DIGITAL HEALTH* 4 (2018), 2055207618785077. https://doi.org/10.1177/2055207618785077 arXiv:https://doi.org/10.1177/2055207618785077

[15] Nahum Gershon. 1998. Visualization of an imperfect world. *IEEE Computer Graphics and Applications* 18, 4 (July 1998), 43–45. https://doi.org/10.1109/38.689662

[16] Miriam Greis, Passant El. Agroudy, Hendrik Schuff, Tonja Machulla, and Albrecht Schmidt. 2016. Decision-Making Under Uncertainty: How the Amount of Presented Uncertainty Influences User Behavior. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction (NordiCHI '16)*. ACM, New York, NY, USA, Article 52, 4 pages. https://doi.org/10.1145/2971485.2971535

[17] Miriam Greis, Aditi Joshi, Ken Singer, Albrecht Schmidt, and Tonja Machulla. 2018. Uncertainty Visualization Influences How Humans Aggregate Discrepant Information. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 505, 12 pages. https://doi.org/10.1145/3173574.3174079

[18] Theresia Gschwandtner, Markus Bögl, Paolo Federico, and Silvia Miksch. 2016. Visual Encodings of Temporal Uncertainty: A Comparative User Study. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (Jan 2016), 539–548. https://doi.org/10.1109/TVCG.2015.2467752

[19] Noura Howell, Laura Devendorf, Tomás Alfonso Vega Gálvez, Rundong Tian, and Kimiko Ryokai. 2018. Tensions of Data-Driven Reflection: A Case Study of Real-Time Emotional Biosensing. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 431, 13 pages. https://doi.org/10.1145/3173574.3174005

[20] Dandan Huang, Melanie Tory, Bon Adriel Aseniero, Lyn Bartram, Scott Bateman, Sheelagh Carpendale, Anthony Tang, and Robert Woodbury. 2015. Personal Visualization and Personal Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics* 21, 3 (March 2015), 420–433. https://doi.org/10.1109/TVCG.2014.2359887

[21] Jessica Hullman. 2016. Why Evaluating Uncertainty Visualization is Error Prone. In *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization (BELIV '16)*. ACM, New York, NY, USA, 143–151. https://doi.org/10.1145/2993901.2993919

[22] Matthew Kay, Tara Kola, Jessica R. Hullman, and Sean A. Munson. 2016. When (Ish) is My Bus?: User-centered Visualizations of Uncertainty in Everyday, Mobile Predictive Systems. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 5092–5103. https://doi.org/10.1145/2858036.2858558

[23] Ian Li, Anind Dey, and Jodi Forlizzi. 2010. A Stage-based Model of Personal Informatics Systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 557–566. https://doi.org/10.1145/1753326.1753409

[24] Isaac M. Lipkus and J. G. Hollands. 1999. The Visual Communication of Risk. *JNCI Monographs* 1999, 25 (1999), 149–163. https://doi.org/10.1093/oxfordjournals.jncimonographs.a024191

[25] Deborah Lupton. 2015. Quantified sex: a critical analysis of sexual and reproductive self-tracking using apps. *Culture, Health & Sexuality* 17, 4 (2015), 440–453. https://doi.org/10.1080/13691058.2014.920528 arXiv:https://doi.org/10.1080/13691058.2014.920528 PMID: 24917459.

[26] Alan M. MacEachren, Anthony Robinson, Susan Hopper, Steven Gardner, Robert Murray, Mark Gahegan, and Elisabeth Hetzler. 2005. Visualizing Geospatial Information Uncertainty: What We Know and

What We Need to Know. *Cartography and Geographic Information Science* 32, 3 (2005), 139–160. https://doi.org/10.1559/1523040054738936 arXiv:https://doi.org/10.1559/1523040054738936

[27] Alan M. MacEachren, Robert E. Roth, James O'Brien, Bonan Li, Derek Swingley, and Mark Gahegan. 2012. Visual Semiotics amp; Uncertainty Visualization: An Empirical Study. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (Dec 2012), 2496–2505. https://doi.org/10.1109/TVCG.2012.279

[28] M. Granger Morgan. 2009. *Best practice approaches for characterizing, communicating and incorporating scientific uncertainty in climate decision making*. US Climate Change Science Program, https://www.globalchange.gov.

[29] Stephen R. Pallone and George R. Bergus. 2009. Fertility awareness-based methods: another option for family planning. *The Journal of the American Board of Family Medicine* 22, 2 (2009), 147–157.

[30] Alex T. Pang, Craig M. Wittenbrink, and Suresh K. Lodha. 1997. Approaches to uncertainty visualization. *The Visual Computer* 13, 8 (01 Nov 1997), 370–390. https://doi.org/10.1007/s003710050111

[31] Pedro-Antonio Regidor, Marta Kaczmarczyk, Esther Schiweck, Maren Goeckenjan-Festag, and Henry Alexander. 2018. Identification and prediction of the fertile window with a new web-based medical device using a vaginal biosensor for measuring the circadian and circamensual core body temperature. *Gynecological Endocrinology* 34, 3 (2018), 256–260. https://doi.org/10.1080/09513590.2017.1390737 arXiv:https://doi.org/10.1080/09513590.2017.1390737 PMID: 29082805.

[32] John Rooksby, Mattias Rost, Alistair Morrison, and Matthew Chalmers. 2014. Personal Tracking As Lived Informatics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 1163–1172. https://doi.org/10.1145/2556288.2557039

[33] J. P. Royston. 1982. Basal Body Temperature, Ovulation and the Risk of Conception, with Special Reference to the Lifetimes of Sperm and Egg. *Biometrics* 38, 2 (1982), 397–406. http://www.jstor.org/stable/2530453

[34] Robert Setton, Christina Tierney, and Tony Tsai. 2016. The accuracy of web sites and cellular phone applications in predicting the fertile window. *Obstetrics & Gynecology* 128, 1 (2016), 58–63.

[35] Orit Shaer, Oded Nov, Johanna Okerlund, Martina Balestra, Elizabeth Stowell, Lauren Westendorf, Christina Pollalis, Jasmine Davis, Liliana Westort, and Madeleine Ball. 2016. GenomiX: A Novel Interaction Tool for Self-Exploration of Personal Genomic Data. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 661–672. https://doi.org/10.1145/2858036.2858397

[36] David Spiegelhalter, Mike Pearson, and Ian Short. 2011. Visualizing Uncertainty About the Future. *Science* 333, 6048 (2011), 1393–1400. https://doi.org/10.1126/science.1191181 arXiv:http://science.sciencemag.org/content/333/6048/1393.full.pdf

[37] Barry N. Taylor and Chris E. Kuyatt. 1994. *Guidelines for evaluating and expressing the uncertainty of NIST measurement results (NIST Technical Note 1297)*. Technical Report. National Institute of Standards and Technology, U.S. Government Printing Office.