Optimal Predictions in Illness Cognition

Talia Robbins
Pernille Hemmer
Rutgers University

Author Note
This data was previously presented at the annual meeting of the Cognitive Science Society, the annual meeting of the Society for Mathematical Psychology, and the Annual Meeting of the Psychonomic Society.

Correspondence concerning this article should be addressed to Talia Robbins, Department of Psychology, Rutgers University, Piscataway, NJ, 08854.

E-mail: talia.robbins@rutgers.edu
Abstract
People have been shown to make accurate predictions for many real world events. Accurate predictions might be particularly important in the domain of health, where illness knowledge directly influences patient outcomes. Therefore, we sought to investigate people’s ability to make predictions for illness durations. We evaluated predictions for both acute and chronic illnesses, as judgments for chronic illnesses have been shown to be influenced by people’s knowledge of acute illnesses. In two experiments we asked participants to estimate the duration of nine illnesses—five acute and four chronic—and we compared their judgments to the Bayesian optimal prediction determined from real-world distributions. For acute illnesses, people were able to estimate both the median duration and the shape of the distribution of illness durations. For chronic illnesses, people correctly estimated the shape of the distribution, but overestimated the median duration. We also tested a group of participants over the age of 40 and found their predictions did not differ from college aged participants. For both experiments, we evaluated the importance of experience in predictions judgments, and found that personal experience with an illness may be important for prediction accuracy. We found the same pattern of overestimation for all illnesses in both younger and older participants. We discuss the possible strategies people may be employing that lead to systematic overestimation.

Keywords: Bayesian; decision making; health; prediction
Significance Statement

Here we seek to apply the paradigm of Griffiths and Tenenbaum (2006) assessing optimal prediction for everyday events to the domain of health. Understanding how people make predictions for health is critically important because inaccurate judgments can lead to maladaptive decisions, as illustrated by poor medication adherence across many chronic illnesses. For instance, according to the American Heart Association, 24% of patients do not fill their prescription within 1 week of being discharged from the hospital, and this already low adherence continues to decrease with time (Ho, Bryson, & Rumsfeld, 2009). This suggests that people with chronic illnesses are miscalibrated for illness statistics, such as durations. The Common Sense Model (Leventhal, Diefenbach, & Leventhal, 1992), a well-established theoretical model in behavioral health, posits that low adherence may be explained by patients applying their expectations and strategies for managing acute illnesses to managing chronic illnesses. Acute illnesses generally follow a pattern whereby symptoms are diminished by treatment until the illness is cured. This pattern is not mirrored in chronic illnesses, which can often be asymptomatic, and cannot be cured. If people are applying their understanding of illness statistics, for acute illnesses (for which they have more experience) to chronic illnesses, they may have expectations that do not closely reflect real world statistics. Here we sought to evaluate whether people understand the duration of acute and chronic conditions, and can estimate the median duration as well as the shape of the duration distribution.
Optimal Predictions in Illness Cognition

Imagine that you have had a cold for a week, and need to decide if you will feel better in time for a trip beginning in two days. You are now faced with a question: How much longer will I be sick? Likewise, later in life, you could find yourself diagnosed with a chronic illness such as diabetes, and would then be faced with a new question: Given this diagnosis, what is my life expectancy? In both cases, you must make the best calculation possible to prepare for the future, and you will likely use your past experience with illnesses to make this estimation. The internal calculation must be based not only on your understanding of the average illness duration, but also on the shape of the duration distribution. In this way, as you progress further into an illness, you can adapt your duration estimate.

How people make judgments and predictions is dependent on their understanding of the statistical regularities of the world. People are well-calibrated to these regularities, and their predictions are, on average, quite accurate (e.g., Brady & Oliva, 2008; Griffiths & Tenenbaum, 2006; Hemmer & Persaud, 2014; Huttenlocher, Hedges, & Duncan, 1991; Huttenlocher, Hedges, & Vevea, 2000). For example, when people were asked to make predictions about events including movie grosses and life spans, they could accurately capture both the shape and median of the distributions for these events (Griffiths & Tenenbaum, 2006). For less ubiquitous areas, such as the reign of pharaohs, people were able to capture the shape of the distribution (i.e. approximately Erlang distributed) but overestimated the duration. This is likely because they were able to apply their understanding of lifespan, but did not account for the age of mortality in the era of pharaohs. As such, it appears that people are calibrated for many events, and can apply their expectations to those events for which they have less knowledge.

Our ability to use our understanding of the regularities of the environment to make
accurate judgments and predictions is particularly important within the domain of health. Everyone faces a problem of estimating illness statistics at some point, and these internal calculations can directly impact patient health (Peters, McCaul, Stefanek, & Nelson, 2006). For instance, people’s estimates of cancer risk directly influence their likelihood to receive cancer screenings (Peters et al., 2006); people who inaccurately estimate their risk are therefore unlikely to be screened regularly, increasing the chance of having a cancer go undetected.

While little is known about people’s understanding of illness statistics, prominent theoretical models of illness cognition make explicit assumptions about peoples’ understanding of illness statistics. For example, the Common Sense Model (CSM, Leventhal et al., 1992) of illness cognition, a well-established theoretical model of health behavior, makes specific claims about how patients make predictions and decisions concerning their health. The CSM asserts that people construct representations of an illness based on symptoms, and that these representations guide their decisions and behavior. The CSM also argues that patients tend to apply their over-learned model for managing acute illnesses (whereby symptoms are temporary and the illness is cured with some treatment) when attempting to manage chronic illnesses. This feedback loop, in which treatment reduces symptoms (often leading to a cure), does not exist for chronic illnesses, and is suggested as a factor in low adherence for chronic illnesses within the CSM framework. However, the CSM argues that people can still use strategies such as applying their understanding for the acute illness statistics when making predictions for chronic illnesses, and health decision making more generally.

In this paper, we sought to assess people’s understanding of illness statistics, namely the median and shape of illness duration distributions. The novelty of this paper is in its application of an established paradigm (Griffiths and Tenenbaum, 2006) for estimating participant
understanding of the environmental statistics to a domain with direct implications for real-world problems, such as patient health. Motivated by the common sense hypothesis that people have a well-learned model for acute illnesses, and less defined models for chronic illnesses, we asked participants to make estimations for both acute and chronic illnesses. As such, in cases where participants have very little knowledge of an illness, the best strategy may be to use their understanding of the distribution of other illnesses for which they have more experience. We expected that participants would perform better (defined as accurately capturing the median and shape of the illness duration distribution) for acute than chronic illnesses overall, with highest performance on illnesses for which they have the most experience. We used the median of the distribution as a good estimate for illness duration, as it is the point at which the illness duration is equally likely to be longer or shorter. Lastly, we hypothesized that participants might be able to accurately capture the shape of the distribution, even if they were not able to accurately estimate its median due to limited experience—that is, they could transfer their understanding of the shape of illness distributions from another illness with which they have experience.

In Experiment 1, we tested participants’ ability to capture the median and shape of both acute and chronic illness distributions by asking them to predict total illness duration based on current duration, and further, we examined the role of personal experience in these judgments. In Experiment 2, we repeated this task with older populations, to evaluate whether age played a role in these estimations, either due to increased experience, or an effect of using themselves as a reference point in estimations (e.g. assuming a later age of onset).

**Modeling Approach**

We followed the modeling approach of Griffiths and Tenenbaum (2006) to compare subjective performance to the optimal prediction from the clinical distributions using Bayes rule,
under the assumption that people’s prediction judgments follow optimal statistical principles. Bayes rule gives a principled account of how people should make predictions about the total duration of a particular illness given the duration of the illness thus far. In Equation 1, below, if $d_{\text{total}}$ indicates the amount of time the average person experiences an illness, and $d$ indicates the duration of the illness thus far, we can estimate $d_{\text{total}}$ given $d$ as follows:

$$p(d_{\text{total}} | d) \propto p(d|d_{\text{total}}) p(d_{\text{total}})$$

(1)

The posterior probability $p(d_{\text{total}} | d)$ is based on a combination of $p(d_{\text{total}})$, the prior probability of the total illness duration, and $p(d|d_{\text{total}})$, the likelihood of the current duration given the average total duration.

To model prediction for illness duration we obtained the clinical data for the duration of 9 illnesses (see Table 1 for sources of clinical data). Illness durations have been found to be well modeled by a type of distribution known as a survival function, which includes Gamma, Exponential, and Weibull. The Erlang distribution is a special case of the Gamma distribution, where $\alpha$ must be an integer, that is often used to model illness duration and illness stages in transmission models of infectious disease, and to infer parameters from clinical data (Krylova & Earn, 2013). Optimal predictions from an Erlang distribution follow an approximately linear function (see Figure 1 bottom panels) with a slope of 1 and a non-zero intercept.

[Insert Figure 1 here]

As a concrete example, the common cold is approximately Erlang distributed with median of $M=4.5$ days. Using Equation 1, the participant’s task is to calculate $p(d_{\text{total}} | d)$ for every possible cold duration for someone met on day $d$ of their cold. If someone has had a cold for 3 days, you would expect their total duration to be around 4.5 days. Likewise, if someone has had a cold for 6 days (which is longer than the median), you might estimate approximately 9.5 days.
In this way, prediction would include both an understanding of the median duration and the Erlang distribution of duration.

**Experiment 1**

**Methods**

**Participants.** One-hundred and eighty-eight Rutgers students participated in exchange for course credit.

**Materials.** We selected nine illnesses—five acute and four chronic (see Table 1)—intended to span a range of durations and familiarity. An acute illness is defined as one which can be cured with treatment, while a chronic illness is defined as one that can be managed but not cured. Familiarity was determined based on prevalence statistics for the number of people diagnosed with that illness each year (see Table 1). Table 1 also includes the source of the clinical data used for the illness duration distributions.

<table>
<thead>
<tr>
<th>Illness</th>
<th>Source of Clinical Data</th>
<th>Prevalence (per 10,000)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Acute (in order of prevalence)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacterial Meningitis</td>
<td>Kilpi &amp; Anttila (1991)</td>
<td>.14</td>
</tr>
<tr>
<td></td>
<td>Goveia et al. (2011)</td>
<td></td>
</tr>
<tr>
<td>Mononucleosis</td>
<td>Cameron et al. (2006)</td>
<td>5</td>
</tr>
<tr>
<td>Appendicitis</td>
<td>Atema et al. (2015)</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Rehman &amp; Hussain (1985)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Singh et al. (2014)</td>
<td></td>
</tr>
<tr>
<td>Seasonal Flu</td>
<td>Michael et al.</td>
<td>1250</td>
</tr>
<tr>
<td>Common Cold</td>
<td>Michael et al.</td>
<td>2360</td>
</tr>
<tr>
<td><strong>Chronic (in order of prevalence)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COPD*</td>
<td>Shavelle (2009)</td>
<td>4.5</td>
</tr>
<tr>
<td>Asthma</td>
<td>American Lung Association (2012)</td>
<td>800</td>
</tr>
<tr>
<td>Type II Diabetes</td>
<td><a href="http://www.cdc.gov/diabetes">http://www.cdc.gov/diabetes</a></td>
<td>860</td>
</tr>
<tr>
<td></td>
<td>/statistics/duration/fig1.htm</td>
<td></td>
</tr>
<tr>
<td>Chronic Heart Disease</td>
<td><a href="http://www.cdc.gov/mmwr/preview/mmwrhtml/">http://www.cdc.gov/mmwr/preview/mmwrhtml/</a></td>
<td>1130</td>
</tr>
<tr>
<td></td>
<td>su6203a26.htm?_s_cid=su6203a26_w#Tab</td>
<td></td>
</tr>
</tbody>
</table>

*COPD refers to chronic obstructive pulmonary disease*
For each illness we sampled five data points from the distribution over duration to be used as probes in the experimental task (see Table 2 for illnesses, median durations and duration probes, listed in order of median duration). Following the procedure of Griffiths and Tenenbaum (2006), samples were obtained by fitting Erlang distributions to each of the nine distributions. See Figures 2 and 3, top row, for the Erlang fits to the empirical illness distributions, and Table 2 for parameter values.

**Procedure.** Each participant completed one duration prediction trial for each of the nine illnesses, given the duration probe. On each trial participants were asked: “*Given that you meet someone who has had illness X for time period Y, what do you think will be the total duration of their illness?*” They responded by entering a number, and choosing a unit of time from a dropdown menu (options included hours, days, weeks, months, and years). The duration probe was randomly selected from the fixed set of 5 possible probes and illness presentation order was randomized.

Participants were instructed that they were being asked to predict *total* duration, not remaining duration (see Appendix A for the experimental instructions) and given a sample

<table>
<thead>
<tr>
<th>Illness</th>
<th>Median Duration</th>
<th>Time Unit</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>T5</th>
<th>α</th>
<th>β</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appendicitis</td>
<td>34.5</td>
<td>Hours</td>
<td>15</td>
<td>20</td>
<td>28</td>
<td>41</td>
<td>67</td>
<td>3</td>
<td>13</td>
</tr>
<tr>
<td>Seasonal Flu</td>
<td>3.5</td>
<td>Days</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>7</td>
<td>3</td>
<td>1.3</td>
</tr>
<tr>
<td>Common Cold</td>
<td>4.5</td>
<td>Days</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>8</td>
<td>20</td>
<td>3</td>
<td>1.7</td>
</tr>
<tr>
<td>Bacterial Meningitis</td>
<td>5.1</td>
<td>Days</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>6</td>
<td>9</td>
<td>5</td>
<td>1.1</td>
</tr>
<tr>
<td>Mononucleosis</td>
<td>8.4</td>
<td>Weeks</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>13</td>
<td>22</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Chronic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COPD</td>
<td>4.1</td>
<td>Years</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>11</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Chronic Heart Disease</td>
<td>8.9</td>
<td>Years</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>10</td>
<td>16</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>Type II Diabetes</td>
<td>10.1</td>
<td>Years</td>
<td>1</td>
<td>5</td>
<td>9</td>
<td>14</td>
<td>32</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Asthma</td>
<td>10.3</td>
<td>Years</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>8</td>
<td>24</td>
<td>1</td>
<td>15</td>
</tr>
</tbody>
</table>

*COPD refers to chronic obstructive pulmonary disease
question to evaluate whether they understood the instructions (see Appendix B for experimenter instructions for administering sample questions). If participants answered the test question incorrectly, the task was explained again, followed by a second sample question. If they answered the second question incorrectly, they would be excused from participating in the experiment. No participants failed the second sample question.

One hundred and thirteen participants received the experimental procedure as described above. Seventy-five participants received three additional questions: “Have you ever experienced illness X?”, “Has a friend or family member ever experienced illness X?”, and “How familiar are you with illness X (on a scale from 1 to 10)?” with 1 being no familiarity and 10 being high familiarity. We asked questions about their familiarity and experience with the illnesses since these are possible factors which could have influenced performance on the duration task. The three questions were presented together for each of the nine illnesses. The three questions were always presented together on the computer screen for each illness. The questions were always in the same order, but the illnesses were presented in the same random order as in the illness duration questions.

All the research presented in this paper was conducted in accordance with the Declaration of Helsinki, and was approved by the Rutgers University Institutional Review Board under protocol # 13-618M.

**Results**

The following data was excluded from analysis: data points smaller than the presented duration \( d_{\text{total}} < d \), unreasonably large responses (defined as those 3 standard deviations greater than the median response for a given illness duration probe), participants who responded using negative numbers, and participants who had more than two data points excluded based on the above
criteria. The responses analyzed were 162 for appendicitis, 171 for the seasonal flu, 170 for the common cold, 164 for bacterial meningitis, 162 for mononucleosis, 170 for COPD, 173 for chronic heart disease, 170 for type II diabetes, and 173 for asthma.

**Prediction.** To evaluate whether people’s predictions for illness durations captured the median and shape of the real-world distribution, we first calculated optimal predictions from the Erlang prior: $d^* = d + \beta \log 2$, where $d^*$ is the predicted value of $d_{total}$ and $d$ is the duration probe (see Griffiths and Tenenbaum, 2006 appendix for the derivation of the prediction equation). Figures 2 and 3, second row, show participant predictions for total durations given the duration probe with optimal predictions calculated from the Erlang distributions, as well as best-fitting Erlang predictions to participant data.

[Insert Figure 2 here]

[Insert Figure 3 here]

A qualitative comparison for the five acute illnesses suggested that the best fitting predictions to the participant data was relatively close to the Bayesian optimal prediction for the clinical distributions. This can be seen by the closeness of the red line to the grey line, as well as the closeness of the respective medians (see Figure 2 row 2). In this way, participants’ predictions for acute conditions were consistent both with the median and shape (following a linear trend with a slope of 1) of the assumed Erlang distribution of the empirical data. For chronic conditions (see Figure 3 row 2) participants overestimated the median, while they still accurately captured the shape of the distributions. It is also important to note that for all illnesses where participants did not accurately estimate the median, the pattern was to systematically overestimate the duration, a result elaborated on in the conclusions.

In order to quantitatively evaluate the difference between participant predictions and the
Bayesian optimal prediction, we performed a bootstrap analysis. For each illness, we drew a random sample with replacement from participant responses for each duration probe. We then used these samples to fit the optimal Erlang prediction (as we did for the red dashed line in Figures 2 and 3), recording the median. This procedure was repeated 1000 times, resulting in 1000 medians for each illness. Finally, we computed the bootstrap 95 percentile confidence interval. If the accurate median fell within that confidence interval, the bootstrapped samples were all practically equivalent to the true median. For confidence intervals see Table 3. Figure 4 shows the lines from the optimal Erlang fit to the bootstrapped samples that fell within the 95 percentile confidence interval, as well as the optimal Erlang fit to the true illness distribution.

![Insert Figure 4 here](image)

Table 3: Bootstrap 95 percentile confidence intervals for means fit to participant data as compared to means for the real world distribution

<table>
<thead>
<tr>
<th>Illness</th>
<th>Experiment 1 Confidence Interval</th>
<th>Experiment 2 Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>True Median Lower Upper</td>
<td>Lower Upper</td>
</tr>
<tr>
<td><strong>Acute</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appendicitis</td>
<td>34.5 29.7 654.7</td>
<td>16.2 324.0</td>
</tr>
<tr>
<td>Seasonal Flu</td>
<td>3.5 4.0 13.9</td>
<td>3.8 14.4</td>
</tr>
<tr>
<td>Common Cold</td>
<td>4.5 2.6 8.9</td>
<td>2.9 9.8</td>
</tr>
<tr>
<td>Bacterial Meningitis</td>
<td>5.1 5.8 29.4</td>
<td>9.2 1063.8</td>
</tr>
<tr>
<td>Mononucleosis</td>
<td>8.4 2.9 87.4</td>
<td>3.2 756.0</td>
</tr>
<tr>
<td><strong>Chronic</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COPD</td>
<td>4.1 4.6 47.9</td>
<td>7.8 49.3</td>
</tr>
<tr>
<td>Chronic Heart Disease</td>
<td>8.9 9.5 57.1</td>
<td>11.3 54.5</td>
</tr>
<tr>
<td>Type II Diabetes</td>
<td>10.1 18.5 73.0</td>
<td>21.4 67.5</td>
</tr>
<tr>
<td>Asthma</td>
<td>10.3 17.3 81.4</td>
<td>24.3 83.1</td>
</tr>
</tbody>
</table>

* COPD refers to chronic obstructive pulmonary disease

The bootstrap analysis found that the true median fell within the bootstrap confidence interval for 3 out of 5 acute illnesses (i.e. appendicitis, the common cold, and mononucleosis),
and none of the 4 chronic illnesses. For the chronic illnesses, it is important to note that the confidence intervals were very large, suggesting low agreement within subjects. This pattern was also the case for less prevalent acute illnesses, such as bacterial meningitis. This illustrates that participants were able to make accurate predictions for the common acute illnesses overall, but were less accurate when making predictions for the chronic illnesses. While the true median for the seasonal flu fell outside the confidence interval for participant responses, participants may have been using a prior expectation for the duration of the common cold when making these estimations, causing them to overestimate the duration.

**Experience.** In the experiment, we also assessed participants personal experience with the illnesses. Analyzing participant responses separately with and without experience could reveal whether participants who had experience were more accurate. However, only one illness, namely the seasonal flu, had enough responses for both personal experience and lack of experience at each duration probe (e.g. the common cold had a high incidence of experience, but not enough responses for lack of experience; conversely, asthma had some responses for experience, but this did not include responses at all duration probes). Thus, for the seasonal flu we separated participant responses into two groups, those with personal experience and those without. We performed the same bootstrap analysis from above, and found that the true median fell within the bootstrap confidence interval when participants have personal experience with the seasonal flu (true median=3.5, 95 percentile CI[3.2 12.6]), and the true median did not fall within the confidence interval when participants did not have personal experience (true median=3.5, 95 percentile CI[5.5 12.1]). This suggests that having personal experience may improve prediction accuracy for the median of the illness distribution.
Experiment 2

In Experiment 1 we found that participants were generally able to capture the shape of illness distributions, and more closely captured the median for acute than chronic conditions. Very few participants had personal experience or familiarity with the chronic illnesses which may be a result of the sample being drawn from college students ranging in age from 18-24. An older population might have more experience, and thus, their predictions might be closer to the optimal prediction for chronic conditions. It is also possible that an older population would assume a later age of onset than younger participants, who may be using themselves as a reference point. Estimating a later age of onset might lead to lower estimations of total duration, making them closer to the true duration. Therefore, in Experiment 2 we sought to examine prediction from an older participant sample. In this experiment we used the same experimental paradigm with participants on Mechanical Turk who were aged 40 or older.

Methods

Participants. One hundred and thirty-five Mechanical Turk workers aged 40 or older from the United States were paid $1 for their participation.

Procedure. Both the materials and procedure were identical to that of Experiment 1, and included the questions regarding experience and familiarity.

Results

Data in this experiment was analyzed using the same exclusion criteria from Experiment 1. The responses analyzed were 112 for appendicitis, 120 for the seasonal flu, 114 for the common cold, 119 for bacterial meningitis, 111 for mononucleosis, 119 for COPD, 117 for chronic heart disease, 116 for type II diabetes, and 120 for asthma.

Prediction. In the same manner as Experiment 1, we calculated optimal predictions as
well as the best fitting Erlang prediction to the observed participant data (see Figure 5). The most striking result is illustrated by comparing results to those in Experiment 1. A qualitative comparison of the best-fitting predictions to the data relative to the Bayesian optimal prediction revealed that participant performance in this task closely paralleled that of Experiment 1, as reflected in the similarity of the median estimations, and the shape of the predictions fitting the Erlang prediction function.

[Insert Figure 5 here]

[Insert Figure 6 here]

For the quantitative assessment, we replicated the bootstrap procedure from Experiment 1, and calculated the bootstrap 95 percentile confidence intervals for the responses of the older adults (see Table 3). When comparing the results to the true distributions, we found that the true median fell within the confidence interval for 3 out of 5 acute illnesses (i.e. appendicitis, the common cold, and mononucleosis), and none of the 4 chronic illnesses.

[Insert Figure 7 here]

Using the bootstrap samples, we also compared the medians between experiments. We found that the confidence intervals for Experiments 1 and 2 overlapped for all 9 illnesses, illustrating that the two groups responses were practically equivalent to one another. These results indicate that overall, older participants did not perform differently than college aged participants.

**Experience.** As in Experiment 1, we separated participant responses by whether or not participants had experience with an illness. Only two illnesses, the seasonal flu and chronic heart disease contained adequate data for this analysis. We did not find an effect of personal experience for the seasonal flu (true median=3.5, no personal experience: 95 percentile CI[4.3
10.7, personal experience: 95 percentile CI[3.8 15]) or chronic heart disease (true median=8.9, no personal experience: 95 percentile CI[10.1 53.1], personal experience: 95 percentile CI[17.0 58.9]). While participants in this experiment did have significantly more experience with 3 of the 4 chronic illnesses than those in Experiment 1 (type II diabetes $\chi^2(1, N=109)=6.4, p=.01$, chronic heart disease, $\chi^2(1, N=105)=13.9, p<.01$, and asthma $\chi^2(1, N=110)=7.8, p<.01$), their experience level was still low, with personal experience not reaching more 20% for any of the chronic illnesses. Although this is higher than the prevalence in the population, (e.g. for chronic heart disease 20% of participants reported experience which is higher than the population mean of 11%), this proportion does not appear to have been enough to drive performance in the sample, as the majority of respondents still had no experience.

**Conclusions**

In this paper, we applied the paradigm of Griffiths and Tenenbaum (2006) assessing optimal prediction for everyday events to the domain of health. We measured how people make predictions about illness durations and compared performance for acute and chronic conditions. The data show that participant responses closely matched the optimal predictions for both the shape and median of the illness distributions for acute conditions, with near perfect performance for the common cold. While previous research has suggested that health decisions operate differently from other decision processes (Levy et al., 2014, we show that for acute illnesses for which people have experience, participants follow optimal statistical principles and have understanding of the regularities of illness distributions.

Furthermore, for chronic conditions the data show that while responses follow the shape of the distribution, the median durations are systematically overestimated. This is in line with the common sense model (Leventhal et al., 1992) which suggests that people should be able to apply
their understanding of acute illnesses to judgments about chronic illnesses, but that a lack of 
experience with chronic illnesses might also lead to misalignment when applying the acute 
model (e.g., overestimating duration).

A strategy of overestimation might be adaptive in terms of planning for the future 
(whether that be short or long term). Recall the opening scenario where you were asked to 
imagine that you had a cold for a week, and needed to predict if you would be feeling better in 
time for a trip beginning in two days. For other illnesses, where you might be unsure of the 
duration, how would you make an estimation of when you are likely to recover? You might take 
an illness you understood better, such as a cold, and adjust upward to ensure yourself an 
adequate recovery time. The same may be true for chronic illnesses. When planning for the 
future (e.g. retirement savings), it may be safer to assume a longer duration. Indeed, when 
planning for the future, it may be safer to overestimate the duration of an illness rather than risk 
underestimating the duration.

In addition, optimism about lifespan for chronic illnesses may be important for positive 
health behaviors. People who report higher levels of optimism about their condition report being 
less bothered by symptoms (Scheier & Carver, 1985) and show faster recovery from surgery 
(Scheier, Owens, Magovern, Leferebve, Abbot, & Carver, 1989). For this reason, it might be 
advantageous to overestimate the duration of chronic illnesses. This could signal optimism, 
which might, in turn, help patients to engage in behaviors that are good for their health, and 
remain healthier longer.

One surprising find of our study was that the pattern of accuracy for acute illnesses and 
overestimation for chronic illnesses was true for both older and younger populations. This might 
be due to the older participants being a healthy online sample with relatively little personal
experience. The data on the impact of personal experience showed an inconsistent influence on participant accuracy. However, since many of the selected illnesses have a low prevalence in the population, there was a similarly low prevalence in our sample, which made it difficult to assess whether those with personal experience would have performed better than those without. When the sample was large enough, in the case of the seasonal flu, we found an improvement in performance for those with personal experience only in the younger adults. Although this appears counterintuitive, the flu often has a longer duration and greater severity in older adults than in younger ones (Rothberg, Haessler, & Brown, 2008). Therefore, even the older adults who reported experienced with seasonal flu might estimate a longer duration, as that may be an accurate reflection of what they experienced.

A critical feature of chronic illnesses that might make prediction for total duration more complex is that by definition persons with chronic conditions have not yet experienced the duration of that illness in its entirety, and therefore do not have knowledge of the total duration. There is evidence that successful predictions require not just some experience in a domain, but a relevant amount of experience. For instance, when asked to estimate the duration of bus routes, participants systematically underestimated the durations (Stephens, Dunn, Rao, & Li, 2015). The authors posited that this was because bus riders rarely complete a journey through an entire bus route, and rather only know the length of their typical journey. This may explain why participant performance seemed to improve for younger participants who had personally experienced the seasonal flu, but not for those who had experienced chronic heart disease. Participants who have experienced the flu have experienced it in its entirety, and therefore have some firsthand knowledge of its duration.

Understanding illness duration information has important implications for health decision
making (McAndrew et al., 2008). People’s understanding of illness duration is directly linked to their health decisions, and ultimately to their health care seeking behavior. For example, if you attribute your symptoms to the common cold, but still find yourself sick after three weeks, you may re-evaluate your illness assignation. Furthermore, accurate understanding of illness statistics impacts patient doctor communication. Doctors often have misaligned expectations of their patients’ illness knowledge (Street & Haidet, 2011), incorrectly believing that their patients have knowledge more closely matching their own. This causes poor communication about illnesses and treatment, and ultimately affects patient health decision making leading to low adherence to treatment regimens.

The significance of the work presented here is both in its novelty—to our knowledge this is the first investigation assessing people’s judgments for illness statistics—and in its importance in understanding people’s ability to make optimal statistical judgments. The findings extend our knowledge of how people make judgment about everyday events to health-based decisions. As such, provides an important step in understanding how people reason about illnesses and illness outcomes, and it provides a foundation for future investigations into patient judgments and decisions.

**Author Contributions**

The study was conceived and designed by both Talia Robbins and Pernille Hemmer. Talia Robbins was responsible for data collection and both authors collaborated on data analysis. Talia Robbins drafted the manuscript, Pernille Hemmer edited the manuscript, and both authors approved it for submission.

**Competing Interests**

The authors have no financial or non-financial competing interests.
Acknowledgments

This research was supported by the National Science foundation grant #1453276.
References


Appendix A

Participant Instructions:

In this experiment, you will be asked to make predictions based on a single piece of information. Please read each question carefully. We are interested in your intuition so please do not make complicated calculations, just tell us what you think.

Specifically, you will be asked to estimate the total duration of different illnesses, based on how long someone has already had the illness. To give you an example of how to think about this question, imagine that you meet a man that is 50 years old and you are asked to estimate the total duration of his life. You might guess that his lifespan is likely to be 79 years of age (because this is the national average).

Importantly, you are NOT being asked how much longer he is likely to live, but rather the total age that he would reach.

PLEASE CALL OVER THE EXPERIMENTER BEFORE CLICKING TO CONTINUE.
Appendix B

Experimenter Instructions:

After participants read the instructions, give them the following test question to ensure that they understand the task:

*Given that you meet someone who has had food poisoning for 2 days, what do you expect the total duration of this illness will be?*

If they answer with any value LESS than 2 days, explain the task to them again and then ask this follow up question:

*Given that you have had a headache for 1 hour, what do you expect the total duration of this illness will be?*

If participants answer with any value LESS than 1 hour, they should be excluded from the experiment.
Figures

**Figure 1**: Top row shows probability density functions, bottom row shows simulated optimal prediction for Erlang distributions. Column 1 shows the Erlang distribution with $\alpha=2$, which is a special case of the Gamma distribution. Column 2 shows the Erlang distribution with $\alpha=1$, this reduces to an exponential distribution.

**Figure 2**: The top row shows real world distributions for the durations of the five acute illnesses and corresponding Erlang distribution fits. The second row shows participant predictions for illness duration in Experiment 1. Red circles show the median predicted duration as a function of presented duration, with error bars indicating the 68% confidence interval (estimated by a 1000 sample bootstrap). The red dashed line is the fits from the Erlang prior to participant responses and the gray line shows the Bayesian optimal prediction, and the black dotted line illustrates an uninformative prior.
Figure 3: The top row shows real world distributions for the durations of the four chronic illnesses and corresponding Erlang distribution fits. The second row shows participant predictions for illness duration in Experiment 1. Red circles show the median predicted duration as a function of presented duration, with error bars indicating the 68% confidence interval (estimated by a 1000 sample bootstrap). The red dashed line is the fits from the Erlang prior to participant responses, the gray line shows the Bayesian optimal prediction, and the black dotted line illustrates an uninformative prior.

Figure 4: The grey lines in the top row show lines fit to the Erlang prior for the bootstrapped medians within the bootstrap 95 percentile confidence interval for the acute illnesses. The green lines illustrate the Bayesian optimal prediction. The bottom row shows the same for the chronic illnesses.
**Figure 5:** The top row shows real world distributions for the durations of the five acute illnesses and corresponding Erlang distribution fits. The second row shows participant predictions for illness duration in Experiment 1. Red circles show the median predicted duration as a function of presented duration, with error bars indicating the 68% confidence interval (estimated by a 1000 sample bootstrap). The red dashed line is the fits from the Erlang prior to participant responses, the gray line shows the Bayesian optimal prediction, and the black dotted line illustrates an uninformative prior.

**Figure 6:** The top row shows real world distributions for the durations of the four chronic illnesses and corresponding Erlang distribution fits. The second row shows participant predictions for illness duration in Experiment 1. Red circles show the median predicted duration as a function of presented duration, with error bars indicating the 68% confidence interval (estimated by a 1000 sample bootstrap). The red dashed line is the fits from the Erlang prior to participant responses, the gray line shows the Bayesian optimal prediction, and the black dotted line illustrates an uninformative prior.
Figure 7: The grey lines in the top row show lines fit to the Erlang prior for the bootstrapped medians within the bootstrap 95 percentile confidence interval for the acute illnesses. The green lines illustrate the Bayesian optimal prediction. The bottom row shows the same for the chronic illnesses.