The importance of prolonged cough monitoring: Data-driven evidence for a 7-day standard

The Hyfe Team

Introduction

Progress in treating cough – both as a symptom and as a disease unto itself – requires a deep understanding of its underlying physiology and psychology. Rigorously assessing this progress relies, in turn, on a deep understanding of the underlying data generating process: what are the statistical properties of coughing, and how can we really tell if a person's cough has improved? For decades, cough statistics could not be fully investigated to answer these questions, for the simple reason that ample cough data was not available.

Now that continuous long-term cough monitoring is a reality, the tide is finally turning. We now have data on millions of coughs from thousands of people in real-world conditions, a cough database orders of magnitude larger than anything previously imaginable. Thanks to this abundance of data, we are constantly learning new things: how often people cough, how people experience cough, and how different objective and subjective cough metrics relate. As our understanding of cough deepens, investigators are realizing that long-term cough monitoring is essential in clinical trials and research studies – the shortcomings of merely monitoring subjects for 24 hours are now crystal clear. But knowing that one day is not enough is not enough of an answer! What people really need to know is just how long to monitor cough to gain reliable insights. Fortunately, our data coughs up a good general-purpose answer:

Monitor for 7 days.

Data-driven cough modeling

How long cough needs to be monitored is a statistical problem; to solve it, we need useful, realistic cough models. The natural first approach would be to model daily cough rates, since they have long been the primary endpoints of cough studies, but the longitudinal data that we have collected shows that this is problematic: daily cough rates do not generally follow nice, well-known distributions. To illustrate this finding, Figure 1 shows the distributions of daily cough rates for 30 individuals who monitored their cough for at least 20 hours per day for at least 60 days; we see that we cannot assume that individuals' daily cough rates are either normally or lognormally distributed. It follows that simulating daily cough rates to understand the role of monitoring time would be a complicated – and perhaps futile – undertaking.

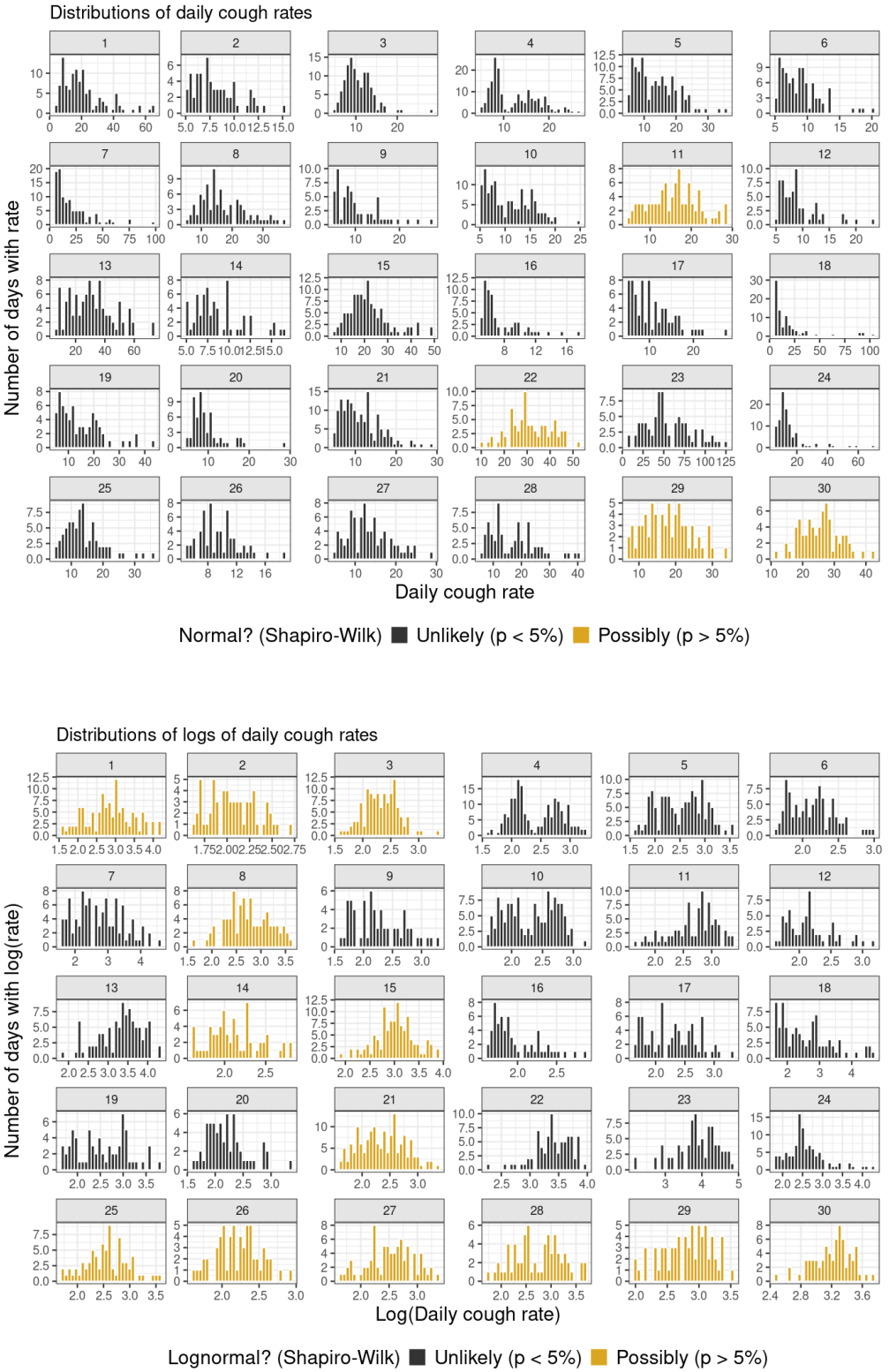

Fig 1. Distributions of daily cough rates and their logarithms for 30 individuals who monitored their cough for at least 20 hours per day over at least 60 days. As indicated, normality and lognormality were assessed with the Shapiro-Wilk test.

All hope is far from lost, however! It turns out that hourly cough counts are much better behaved, statistically speaking, than daily cough rates. Using the hourly data from the same 30 people from Figure 1, Figure 2 shows at a glance that most hourly cough count distributions look similar, and for good reason: hourly cough counts generally follow zero-inflated negative binomial (ZINB) distributions quite well.

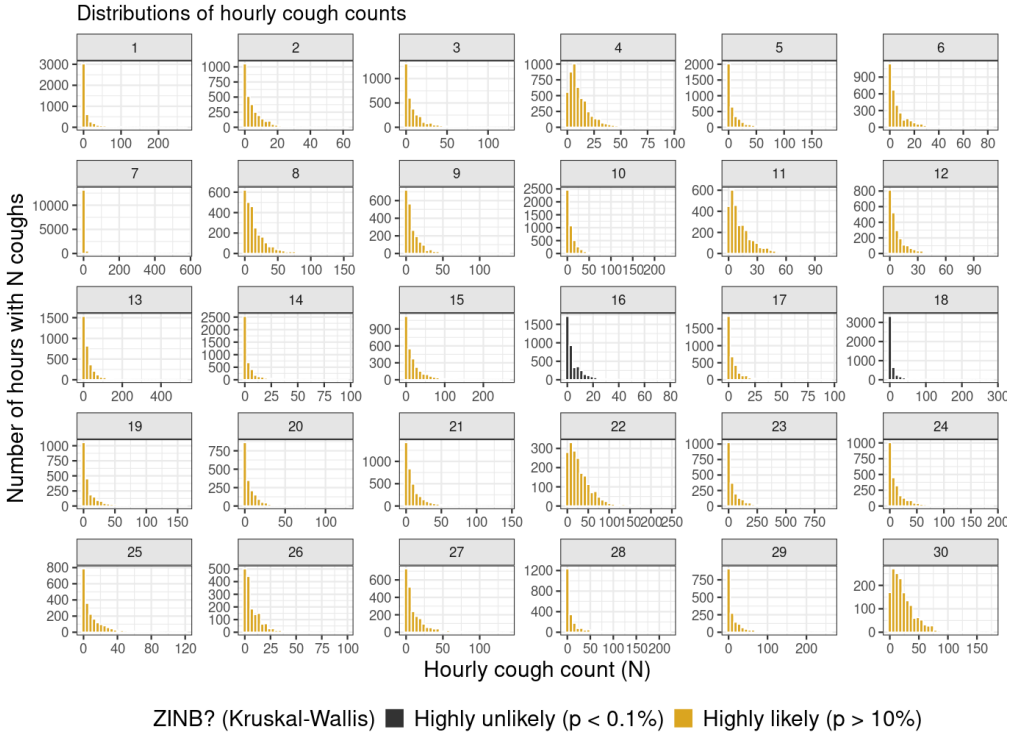

Fig 2. The distributions of hourly cough counts for the same 30 people from Figure 1 are all structurally similar. According to the Kruskal-Wallis test, all but two of these distributions (93.3%) agree well with the corresponding ZINB distributions.

These empirical insights mean that we can simulate coughers realistically to answer all sorts of questions, including the one that prompted this article.

7 days should do it

To determine how long we need to monitor to understand a person's cough, we used ZINB distributions to simulate 504 hours of coughing – 3 uninterrupted weeks – for more than 20 million subjects with different cough parameters. For each combination of hourly average and hourly standard deviation, we simulated 2000 subjects and calculated how many hours of observation were needed for the middle 80% of sample means (10th to 90th percentiles) to be within 20% of their true values.

The results are shown in Figure 3. The red dashed lines delimit the region of main interest, coughers whose hourly cough counts have coefficients of variation (CVs) between 0.8 and 1.65; 25 of the 30 individuals above (83.3%) belong to this zone. As shown and as might be expected, coughers with less variability – CVs around or less than 1 – only need a few days of monitoring to obtain good estimates of their hourly average and hourly standard deviation, while coughers with greater variability – CVs closer to 1.65 – need 7 or 8 days. Since most coughers have CVs less than 1.65, it seems reasonable to expect good results from 7 days of monitoring for most subjects.

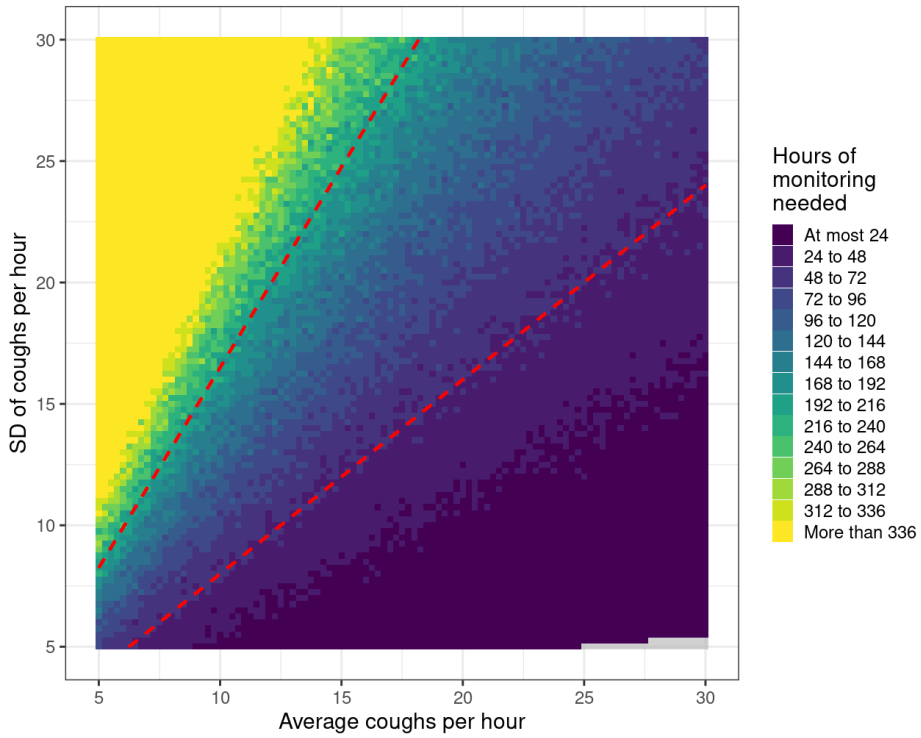

Fig 3. 7 days of monitoring (168 hours) yield good estimates of hourly averages and hourly standard deviations for most coughers.

Conclusions

In this paper, we have made the case that 7 days should be the default monitoring period for cough studies. This period is significantly longer than the current default (24 hours), but advances in cough monitoring tools have made monitoring 7 days continuously fairly easy to do.

That said, our proposed default monitoring period of one week is not necessarily the right monitoring period for all conditions and all studies. The data presented here come from adults who chose to monitor their cough; other demographics with other etiologies may have different underlying distributions of hourly cough counts, which may require different monitoring periods.

What is clear is that the phenomenon of coughing is highly variable, not just between individuals, but also for each individual. An individual's cough rate is like a city's weather: observing it for just one day is not enough to provide reliable information from which to generalize. When it comes to what the "right" observation period is, "the longer the better" is the correct scientific answer, but may not be the correct operational, financial, or logistical answer. For these reasons, we propose 7 days as the default amount of time to monitor in most cases.