

Some Suggestions for Sleep Apnea Self-Experimentation

Bo Adler

May 6, 2010

Abstract

Tracking parts of your life is easy, if somewhat time-consuming. What is harder is doing a meaningful analysis with valid conclusions. This article documents some of the lessons I have learned from trying to experiment on myself.

1 Introduction

Several people have expressed interest in learning more details about my self-experimentation, wondering how they could do their own experimentation. The basic principles are the same as those taught in high school science class:

1. develop a testable hypothesis
2. run your experiment and collect data, keeping careful notes
3. analyze your data
4. consider potential sources of error

Initially, the hardest step is the first; you need a clear idea of what you are trying to test, and that might involve learning about how your body or illness works. Once you have learned enough to create a hypothesis, later experiments will probably be variations on a theme outlined by this first theory. My suggestion is to explore a website like CureTogether¹ to see how other people with similar symptoms are trying to address their issues. For example, CureTogether has put together a nice graph showing the different treatments that depression sufferers have tried, along with the perceived effectiveness. [5] A graph like this can help you narrow down the list of treatments that you'd like to test to a manageable few.

After spending so much time learning and thinking about your particular case, the second step then tests your patience because it is so boring and time consuming. Even with gadgets to help you measure things, you have to upload the data to your computer or manually enter it into a spreadsheet; during the worst period, I was spending more than 30 minutes a day fiddling with my computer trying to organize my collection. My experience is that the devices and software are still pretty primitive, so don't try to get too ambitious. There are several websites (e.g., CureTogether, MedHelp²) and smartphone applications (e.g., The Carrot, Polka, Lose It!) that can help with this. (See [14, 15] for lists of resources available.)

The analysis step will again require some diligent thought, if you want to be thorough. It isn't enough to simply count something and then compare two different scenarios; you need to do a statistical analysis (testing the *null hypothesis* to determine the *p-value*) [3], but it is easy to confuse what the results actually mean (see [10, 9] for nice overviews of the problems with interpretation, and [6, 7, 8] for detailed discussion). Thorp [11] walks through a nice example of how to use the Processing language to help discover unexpected relationships, and other tools include R³ or even just a spreadsheet with graphing capabilities. The web has several resources introducing the statistical concepts (for example, see [12, 13, 2]); textbooks for biostatistics classes also contain the relevant material.

¹<http://www.curetogether.com>

²<http://www.medhelp.org>

³<http://www.r-project.org>

2 My Failed Tracking

When my sleep apnea first became serious, the doctors didn't realize that I had sleep apnea and suggested that I just exercise more to get a proper night's sleep. The exercise didn't seem to help, so I decided to try to figure out if anything was related to my good nights or bad nights. I made a spreadsheet where I listed a bunch of factors related to my sleep:

- how many hours I had sleep
- how many times I woke up in the night
- how much time I had spent at the gym
- how many sleeping pills I took
- how many pills of nyquil I took
- whether I ate healthy or poorly the day before
- whether my girlfriend spent the night or not
- how sleepy I felt when I first woke up the next day
- how sleepy overall I had felt during the whole next day

This turned out to be too much to collect (and if you're just starting with tracking, you should be wary of this issue [4]). I gave up after about 20 days, because even though I had lots of information I wasn't sure how to approach analyzing it. I had a vague idea that I felt better some days and not others, and so I suppose I had some hypothesis that I slept better because of the nyquil or sleeping pills.

The last two items suggest that "how I felt during the day" was going to be my measurement for testing my hypothesis, but there are two big problems with this:

- the measurement is subjective. I wrestled with this problem a lot while I was tracking the data, but couldn't see a way past it.
- accumulated sleep debt means that even if you "slept well" you can still feel lousy the next day. The only way around this is to test the same combination of parameters for a week or more, which means that coming to a conclusion was going to take a really long time.

The final nail in the coffin for this round of tracking was when the doctors realized that I had a severely deviated septum, and that was certainly impacting my sleep. My sleep deprivation was so bad that I would fall asleep while driving and couldn't really focus long enough to have coherent thoughts, so there was no question about having the surgery to adjust my septum.

3 A Second Chance

As it happens, adjusting my septum improved my sleep enough that I can now drive safely and have coherent thoughts again. I still have sleep apnea, but not as severe as before. This gave me a second opportunity to enter the world of self-tracking, this time armed with a pulse oximeter so that I could actually see apnea events on a chart (see [1] for slides containing more information).

The hypothesis I had developed was that my sleep apnea was being caused by chronic congestion and inflammation in my sinus region. I had various bits of anecdotal evidence leading me to this hypothesis, including a brief period right after the surgery where my apnea (and congestion) seemed to go away. And unlike previously, taking decongestants like Sinex and Nyquil seemed to definitely make me feel better the next day.

To test this hypothesis, I selected several treatments (Nyquil, Waldryl, Sudafed, and a CPAP machine) and pseudo-randomly chose a medication (or none at all) each night.

This time, I chose to do measurement by using a CMS-50E⁴ pulse oximeter. This device will record your heart rate and blood oxygen saturation at one second intervals, which you can upload to a Windows program which will generate charts for you and do a rudimentary analysis.

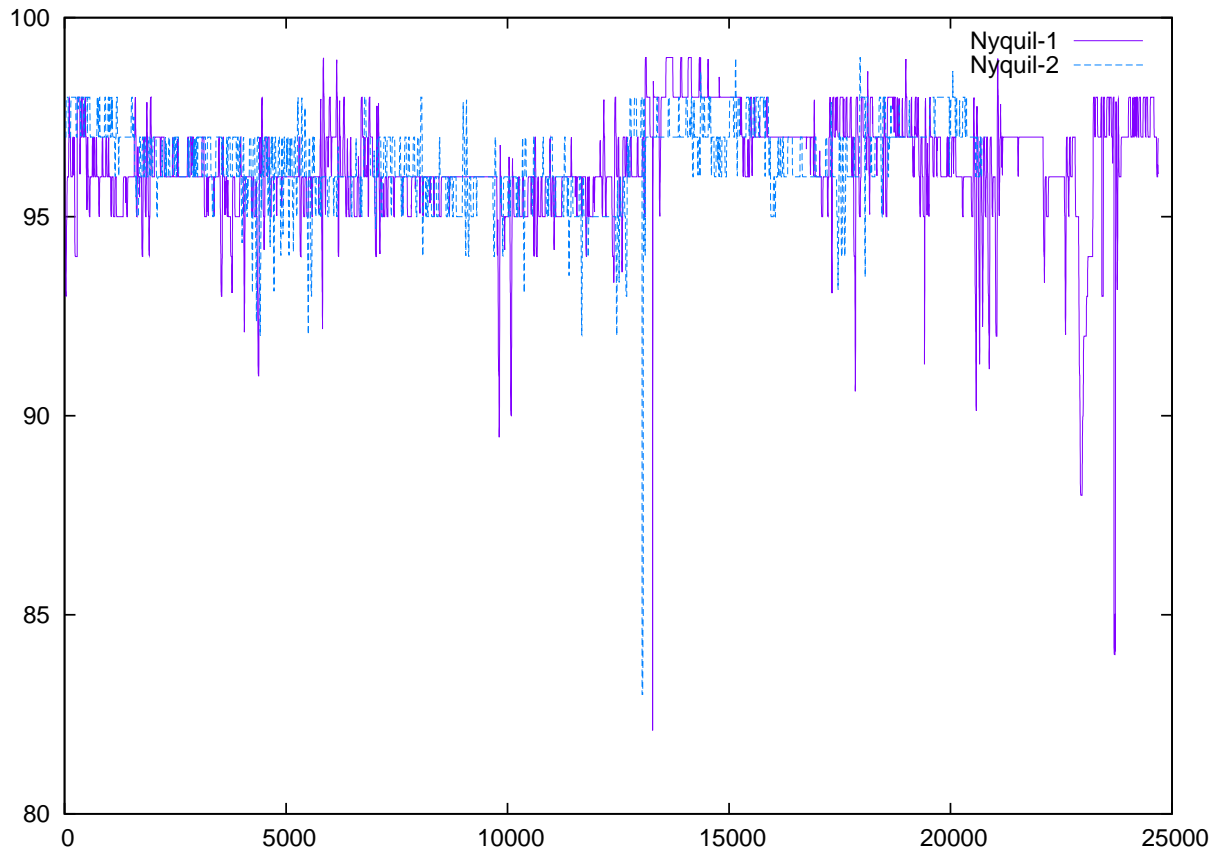


Figure 1: Blood oxygen levels for two different nights, after taking Nyquil. The y -axis shows the percentage of oxygen measured. Values below 88% are typically considered to be apnea events.

The method of my analysis was to compare the charts from the different nights. The apnea events are readily apparent in Figure 1, and Figure 2 shows a distinct lack of apnea events for the night shown.

3.1 Sources of Error

In hindsight, there are many errors in the design of this experiment. The most significant is that an analysis of “comparing charts” is subjective, depending on the human eye to gauge the variability and the severity of events. I concluded by eye that Nyquil wasn’t so helpful for my apnea, but Sudafed was. One friend observed that my apnea gets more severe as the night progresses, and that perhaps the length of sleep in each chart was not enough to truly assess the impact of the drugs. (I had shorted each chart to be the same length, since I slept different amounts each night.) Clearly, the subjectivity introduced made drawing conclusions a matter of seeing what I wanted to see.

⁴http://www.contecmed.com/paypal/product_show.asp?id=23

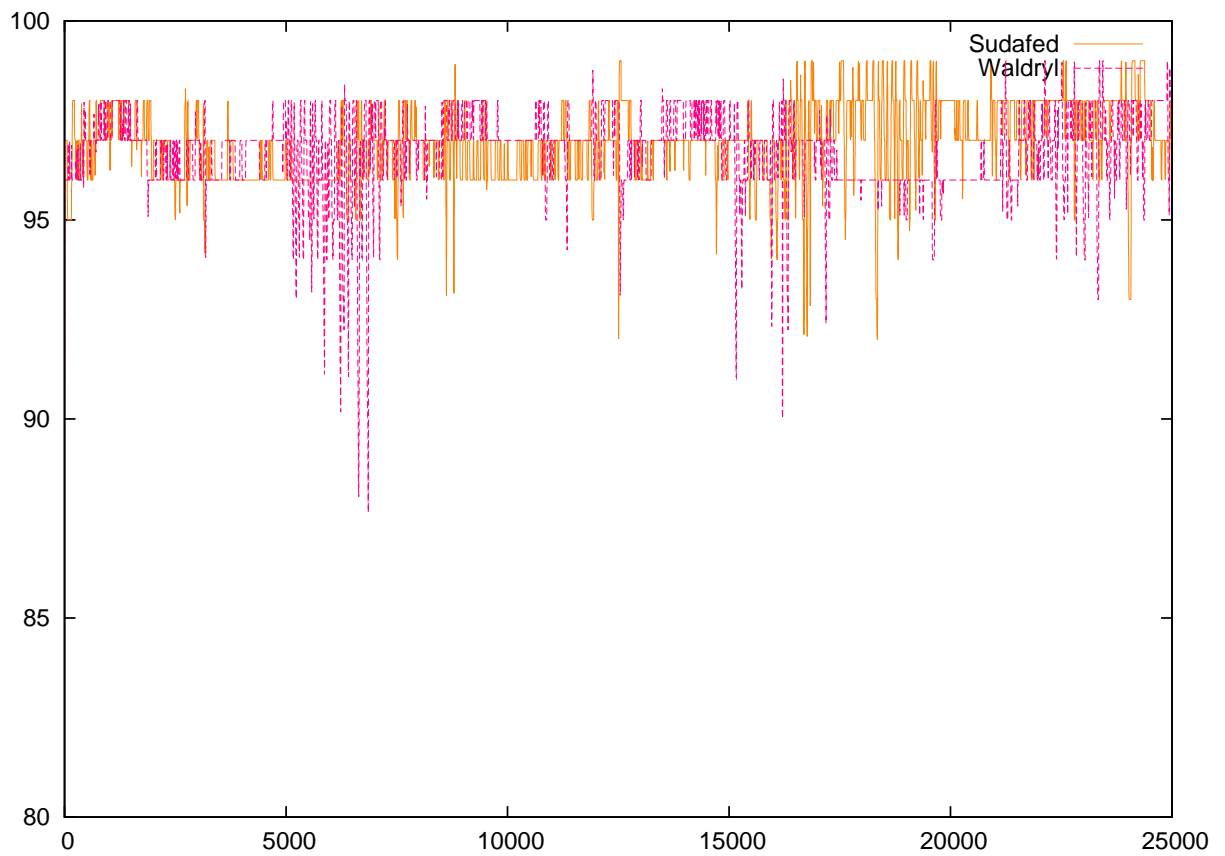


Figure 2: Blood oxygen levels for two different nights, after taking Sudafed or Waldryl. The y -axis shows the percentage of oxygen measured. Values below 88% are typically considered to be apnea events.

Another error is that I didn't use any sort of statistical analysis to make the comparison. Since I used such a limited dataset in my presentation, the results could have been a fluke due to some other circumstance. Ideally, I would have tested the data against the hypothesis that each drug was better than no drug at all; notice that with the subjective evaluation I chose, it's not even possible to calculate anything as simple as a mean or standard deviation.

Other issues with this experiment include:

- Noisy data from the sensor when my hand moves. When hooked directly to a computer, the sensor is able to capture data at 60Hz and movement seems very significant to the pulse detection. I don't know how sensitive the oxygen detection is to movement, but the potential for error is there.
- Trying to estimate when sleep started to synchronize the charts. I often take half an hour or longer to fall asleep, so I had to estimate when I fell asleep so that I could align the charts and make them "comparable." Again, I used a subjective measurement which might create coincidences that aren't legitimate. What I should have done is chosen some particular heart rate below my waking BPM, and made that the event which signifies sleep onset.

3.2 Improvements for Next Time

It's not unusual to realize so many problems until after the data collection has happened. The experience will allow you to refine your process and create better experiments which have more validity. In my case, moving towards more quantification (and removing the subjectivity) would allow me to do the statistical analysis necessary for drawing a good conclusion.

I previously chose to evaluate by eye because there were many factors I was trying to balance: how many apnea events, how severe they were, and how long I slept for. Now that I have been reflecting on it, I think there's a very simple solution suggested by professional sleep studies: find the maximum number of apnea events within a 60 minute period. To make it more rigorous, I define an apnea event as any sequence of measurements which fall below 90% but don't fall before 50%; the 50% restriction serves to eliminate noisy measurements which typically drop to zero.

The change to strict quantification then allows me to do a proper statistical analysis and obtain a confidence level for the results.

4 My Zeo Experience

Another device I experimented with is the Zeo⁵. It's a pretty interesting device for measuring the quality of your sleep, and the website has trackers for various factors that can influence that quality (like light, noise, caffeine). It also has some simple graphing abilities that lets you judge correlations by eye. As noted earlier, judging things by eye is pretty subjective, but it can give you some ideas for where to focus your attention. A nice change feature is that Zeo responded positively to criticism from the Quantified Self crowd, and it is now possible to export your data from the website so that you examine the data beyond what their website enables.

Figures 3 and 4 show some example graphs from my own data. The chart comparing quality to day (Figure 4) is interesting because you can see that weekends aren't much different from weekdays. This matches my expectation, because I generally work seven days a week (so I don't behave very differently on weekends). There is some deviation on Thursdays and Fridays, which I can speculate might be related to my group meeting on Thursdays.

The genuine benefit to tracking is when you can learn something new about yourself. I didn't expect the various trackers on the Zeo website to reveal anything interesting, since I had already spent a lot of time trying to improve my sleep environment because of the sleep apnea. I was surprised to see Figure 5 reveal a distinct correlation between how much soda I drank after 3pm and the quality of my sleep the next night.

⁵<http://www.myzeo.com>

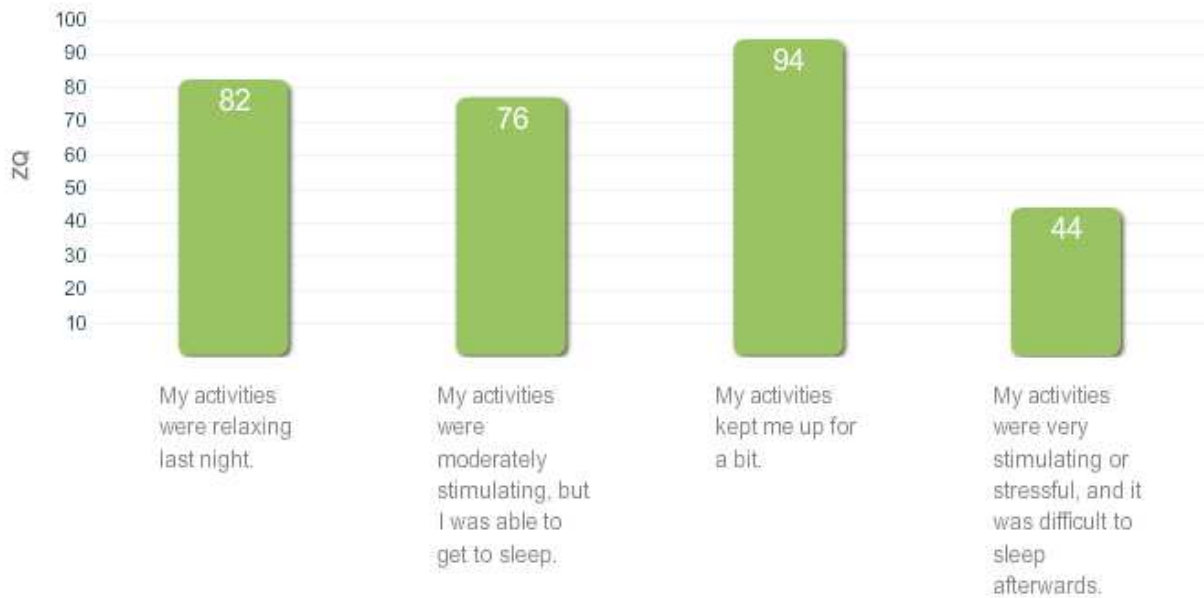


Figure 3: Average sleep quality compared to the night time activity. The graph suggests that even if I am pretty active just before going to sleep, it doesn't effect my sleep in predictable way. The last column suggests that stressful activities before bedtime have a dramatic effect on my sleep quality. What the graph doesn't reveal is how many data points are in each column; the *stress* column represents data from a single night.

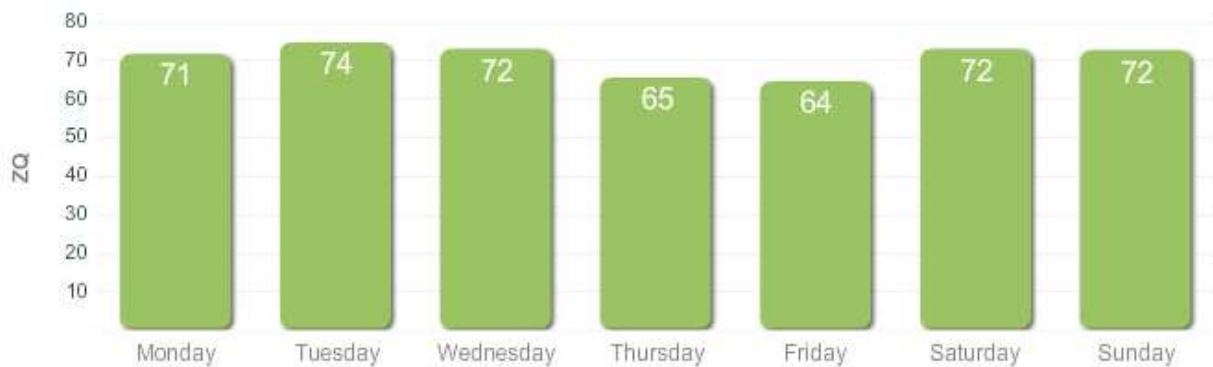


Figure 4: Average sleep quality for each night of the week.



Figure 5: Average sleep quality compared to amount of caffeine consumed. As with Figure 3, the amount of supporting data isn't revealed; there is only a single day where I drank two cans of soda after 3pm.

What made this especially surprising was that during this set of tests, I would stop drinking soda by 6pm and then go to bed at midnight...so the caffeine was significantly affecting my sleep more than six hours later.

Before I finish jumping to that conclusion, let me point out a problem: there is no indication of how many nights of data are being averaged in each column. After exporting the data, I was able to discover that there were only four days where I had one soda, and one day where I had two sodas. Clearly, any trend we can see in the graph is still speculation because there isn't enough data to support it, but it does suggest a direction for future experimentation.

Another question this episode brings up for me is, why is it possible to see a correlation between soda and sleep when I thought sleep debt would be a source of error? I believe that the answer is in the definition of what is being measured. In Section 2, what I was measuring was "how I felt" the day after sleep, which is a subjective measure which can be affected by sleep debt. The Zeo is instead measuring a more objective notion of your sleep quality, based on how long you were in each phase of sleep. (This probably isn't completely independent from accumulated sleep debt, but you can see how it's likely to introduce a smaller error and be consistent in the size of the error.)

5 Conclusion

The best way for us to advance the state of the art is to share our experiences, both good and bad. If you run your own experiments, please consider sharing what you did, especially if you discover or build new tools to make collection or analysis easier. Good luck, and don't give up!

References

- [1] B. Thomas Adler. QS show & tell: What I did on summer break, September 2009. Available from World Wide Web: <http://users.soe.ucsc.edu/~thumper/proj-ddhc/20090914-QuantifiedSelf.pdf>. [Online; accessed 1-May-2010].
- [2] Rob Britton. Statistics for programmers VI: Omitted variable bias, April 2010. Available from World Wide Web: <http://lovehateubuntu.blogspot.com/2010/04/statistics-for-programmers-vi-omitted.html>. [Online; accessed 1-May-2010].

- [3] Alexandra Carmichael. How to measure small effects in your data, December 2009. Available from World Wide Web: <http://www.kk.org/quantifiedself/2009/12/how-to-measure-small-effects-i.php>. [Online; accessed 1-May-2010].
- [4] Alexandra Carmichael. What tools should I use to make tracking easier?, December 2009. Available from World Wide Web: <http://www.kk.org/quantifiedself/2009/12/what-tools-should-i-use-to-mak.php>. [Online; accessed 1-May-2010].
- [5] Alexandra Carmichael. 6 surprisingly effective treatments for depression, April 2010. Available from World Wide Web: <http://curetogether.com/blog/2010/04/22/6-surprising-depression-treatments/>. [Online; accessed 1-May-2010].
- [6] Frank Davidoff. Standing statistics right side up. *Annals of Internal Medicine*, 13(12):1019–1021, June 1999. Available from World Wide Web: <http://www.annals.org/content/130/12/1019.full.pdf+html>.
- [7] Steven N. Goodman. Toward evidence-based medical statistics. 1: The p-value fallacy. *Annals of Internal Medicine*, 13(12):995–1004, June 1999. Available from World Wide Web: <http://www.annals.org/content/130/12/995.full.pdf+html>.
- [8] Steven N. Goodman. Toward evidence-based medical statistics. 2: The Bayes factor. *Annals of Internal Medicine*, 13(12):1005–1013, June 1999. Available from World Wide Web: <http://www.annals.org/content/130/12/1005.full.pdf+html>.
- [9] Jonah Lehrer. Self-tracking, May 2010. Available from World Wide Web: <http://scienceblogs.com/cortex/2010/05/self-tracking.php>. [Online; accessed 3-May-2010].
- [10] Tom Siegfried. Odds are, it’s wrong. *Science News*, 177(7):26, March 2010. Available from World Wide Web: http://www.sciencenews.org/view/feature/id/57091/title/Odds_Are,_Its_Wrong.
- [11] Jer Thorp. Your random numbers — getting started with Processing and data visualization, April 2010. Available from World Wide Web: <http://blog.blprnt.com/blog/blprnt/your-random-numbers-getting-started-with-processing-and-data-visualization>. [Online; accessed 1-May-2010].
- [12] Wikipedia. P-value — Wikipedia, the free encyclopedia, 2010. Available from World Wide Web: <http://en.wikipedia.org/wiki/P-value>. [Online; accessed 1-May-2010].
- [13] Larry Winner. Introduction to biostatistics, July 2004. Available from World Wide Web: http://www.stat.ufl.edu/~winner/sta6934/st4170_int.pdf. [Online; accessed 1-May-2010].
- [14] Gary Wolf. QS resource links, 2009. Available from World Wide Web: <http://www.kk.org/quantifiedself/qs-resource-links.php>. [Online; accessed 1-May-2010].
- [15] Gary Wolf. QS resources on delicious, January 2010. Available from World Wide Web: <http://www.kk.org/quantifiedself/2010/01/qs-resources-on-delicious.php>. [Online; accessed 1-May-2010].