The Crockett-Rodeo Real-Time Health Monitoring Study: Summary of Findings
June 27, 2018

Cheryl Holzmeyer, Ph.D.
Postdoctoral Research Fellow
Center for Science, Technology & Society
Email: c.holzmeyer@gmail.com

Research Goals:
This study aimed to investigate how consumer-grade technologies, such as Fitbits and smart phone apps, might help residents of Crockett and Rodeo, living in proximity to the Phillips 66 refinery, to analyze how ambient air quality impacts their personal and community health. In particular, the study’s goal was to build on existing data platforms and technologies to investigate new ways for residents of fenceline communities to interact with personal and community environmental health data. Researchers’ hypothesis was that the use of such digital tools might surface new possibilities for public data exploration, interpretation, and advocacy. Study participants, meanwhile, tended to describe their goals in terms of health. They wanted to better understand the health impacts of air pollution on themselves and their neighbors:

“My interest continues to be the refinery impacts (picked up on air monitors and delivered by the breeze) on specific neighborhoods. Example: High methane blown over a neighborhood = increased heart rates of those who live in a 6 block radius.”
— Crockett-Rodeo resident & study participant

Research Methods:
To investigate how consumer health-tracking technologies might be used, alongside real-time air monitoring, to surface new possibilities for data exploration, we relied on the following methods. These methods included traditional research instruments, such as surveys, to gather data on participants’ baseline health status and other information. They also comprised new research instruments, such as Fitbits and apps, to collect real-time health indicator data over the two-month study period. These real-time health data were then correlated with real-time air quality data.

• Initial Survey: Participants completed an initial online survey, reporting on information such as the number of years lived in Crockett or Rodeo, basic demographic information, and any existing health issues.
• **Real-Time Health Indicator Tracking:** Between mid-March 2018 and mid-May 2018, participants wore Fitbits to measure their minute-by-minute heart rates, and reported their blood oxygen saturation (SPO2) in the morning and evening, using a pulse oximeter and associated app (Paco). Another app (Moves) tracked participants’ location data. These apps were all coordinated with a study e-mail address, unique to each participant.

• **Real-Time Air Quality & Health Indicator Correlations:** Each participant also had their own account on the data analysis platform Data Sense ([http://makesenseofdata.com/](http://makersenseofdata.com/)). Participants’ study data streams (from Fitbit, Paco, and Moves apps) were uploaded to Data Sense, and then analyzed for any correlations with real-time air quality data over the study timeframe.

• **Exit Survey (in-progress):** Participants will be invited to complete an exit survey, assessing technologies used in the study as well as their overall experience.

### Who Participated?

**Total participants** completing study data collection via apps: **31**
- **Crockett** participants: **20**
- **Rodeo** participants: **11**

![Map](https://via.placeholder.com/150)

**Blue:** Residences of Crockett-Rodeo participants in the Real-Time Health Monitoring Study

**Purple:** Refinery Fenceline Air Monitors

(Note: the Rodeo South monitor is closest to Rodeo, while the Rodeo North monitor is closest to Crockett)

**Years lived in Crockett or Rodeo:**
- Low: 6 months; High: 76 years
- Average: 24.24 years; Median: 11 years; Standard deviation: 24.76

**Participant ages:**
- Average: 61.68 years; Median: 64 years; Standard deviation: 10.19

**Participants’ Self-Reported Health Status** (overall assessment):
- Excellent: 7 (23%)
- Good: 23 (74%)
- Fair: 1 (3%)
- Poor: 0
Even so:
- 84% of participants (26 of 31) reported at least one health issue
- 55% (17 people) reported more than one health issue
- 16% (5 people) reported more than two health issues

The most prevalent health issues included:
- 29% (9 people) reported asthma as a health issue
- 29% (9 people) reported thyroid issues
- 13% (4 people) reported chronic headaches
- 10% (3 people) reported cancer as a health issue
- 10% (3 people) reported chronic skin rashes

What Was Crockett and Rodeo Air Quality Like During the Study?

The air quality monitors in closest proximity to study participants, which were used for study data collection, comprised two refinery fenceline air monitors (Rodeo North and Rodeo South), as well as a third community air monitor, Rodeo 2B (located south of Rodeo, in Hercules). Air quality data were imported into Data Sense (http://makesenseofdata.com/) for further analysis.

Different monitors detected different pollutants during the study timeframe (March 8th – May 9th), as indicated in these Data Sense visualizations below:

- **Figure 1:** Pollutants detected during the study by the Rodeo South Fenceline Monitor, plus PM 2.5 from the Rodeo 2B Community Monitor.
- **Figure 2:** Pollutants detected during the study by the Rodeo North Fenceline Monitor.

**Summary of pollutants detected by each monitor:**

<table>
<thead>
<tr>
<th>Rodeo North Fenceline Monitor</th>
<th>Rodeo South Fenceline Monitor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ozone</td>
<td>Ozone</td>
</tr>
<tr>
<td>Sulfur Dioxide</td>
<td>Sulfur Dioxide</td>
</tr>
<tr>
<td>Methane</td>
<td>Ammonia</td>
</tr>
<tr>
<td>Nitrous Oxide</td>
<td>Nitrous Oxide</td>
</tr>
<tr>
<td>Carbon Monoxide</td>
<td>Carbon Monoxide</td>
</tr>
<tr>
<td>Total Hydrocarbons</td>
<td>Ethylene</td>
</tr>
<tr>
<td>Ethanol</td>
<td>Total Hydrocarbons</td>
</tr>
<tr>
<td><strong>Rodeo 2B Community Monitor</strong></td>
<td>Methane</td>
</tr>
<tr>
<td>PM 2.5</td>
<td></td>
</tr>
</tbody>
</table>

These visualizations indicate the differences in air quality data even across two fenceline monitors, which might be expected to be more consistent than a fenceline monitor and a community monitor, located further from the refinery. It is unclear what the reasons for all of these differences may be – both the different chemicals detected by each monitor and the different pollutant levels over time. Variable winds and variations in how different chemicals move through the air may be factors affecting the different pollutant levels. At the same time, monitors were supposed to detect a wider range of pollutants associated with petroleum refining, which they did not detect (as detailed at: [http://www.fenceline.org/rodeo/data.php](http://www.fenceline.org/rodeo/data.php)). So overall, it is important to recognize that these air quality data are incomplete and imperfect, limiting the possible correlations that can be found through them.
What Correlations Between Air Quality and Health Indicators Did We Find at the Community Level?

To investigate possible correlations between air pollutant levels and study participants’ health indicators, our initial data analysis plan was to rely on an R-based computer script, written specifically for this study, to make sense of its multiple, large datasets. This script would have gone beyond analyzing individuals’ correlations, as the Data Sense platform enables; it was instead designed to analyze patterns at the community level, and among participants in various sub-groups – for instance, participants situated in the same neighborhood, or who had lived in Crockett-Rodeo for decades. However, this data analysis script ended up more being cumbersome to implement and use than expected. Ultimately we were not able to fix the problems we encountered with the code within the timeframe of the study, or within any foreseeable timeframe.

Instead, we calculated the variability in participants’ individual correlations (i.e. between air quality and health indicators) in terms of their mean, median, and standard deviation. In doing so, we sought to assess whether there was any “signal” in these data. The details of these findings are in Appendix 1.

**Based on these data, we did not find consistent patterns, across participants, correlating air quality and health indicators during the study.** For example, we did not find that people living in the same city (Crockett or Rodeo), or in the same neighborhood vicinities, shared similar correlations between air quality and health indicators. Nor did we find that participants who had lived for multiple decades in Crockett or Rodeo shared correlation patterns, compared with those who had lived in the area for much shorter periods of time (e.g. less than 10 years).

It could be that no such patterns exist. But **there are multiple other possible reasons that we did not find them**, including:

- **Inadequate air quality data**, due to inadequate monitoring infrastructure. These inadequacies include placement of a small number of monitors only at the refinery fenceline, rather than a denser network of monitors located both at the fenceline and further out in the community. They include inadequate thresholds for monitors to detect pollutants in the first place, and therefore high rates of air monitor “non-detects.” Pollutants may still be present at levels below the monitors’ detection thresholds.
- **Apparent lack of significant pollution events**, such as chemical releases or major flaring, during the study. These releases and other air pollution fluctuations may have seasonal variations that would also not be picked up over the two-month duration of our study.
- **Inconsistent personal health indicators data.** For example, not all participants’ Fitbits functioned properly throughout the study to track their heart rates.
- **Confounding variables** – in other words, factors besides detected air quality, such as participants’ residential locations relative to pollution sources, or undetected chemicals, which interfered with identifying any consistent relationships between detected air pollution levels and health indicators.
- **No prevailing wind pattern** over the course of the study, which might have localized the effects of air pollution.
- **Insufficient numbers of participants in various sub-groups** (e.g. neighborhoods) to make comparisons.

While it is impossible to know precisely how to evaluate and rank each of these reasons above, there is a case to be made that the inadequacies of the Phillips 66 air monitoring system are particularly responsible for the lack of data available for correlation analysis over the study timeframe. Unlike the Fitbit malfunctions, for instance, inadequate air quality data affected possibilities for data analysis across all participants.
What Correlations Between Air Quality and Personal Health Indicators Were Available to Individual Participants?

Even so, individual participants in the Real-Time Health Monitoring Study could explore correlations between air quality and their personal health indicators from the study – namely, resting heart rate and blood oxygen saturation (SPO2) – through the Data Sense platform (http://makesenseofdata.com/).

This platform enabled participants to examine both positive and negative correlations between levels of air pollutants and personal health indicators, though correlation grids like the one below. These grids highlighted the wide range of correlations in the data of different participants. The scatterplots and line graphs, on the grid’s right side, should correspond to the numerical correlations on the left side that are located at the same intersections (e.g. of heart rate + methane).

Interpreting Data Sense grids: The above grid shows a series of correlations between this participant’s health data – in column #1 (resting heart rate) and column #2 (blood oxygen saturation, or SPO2) – and air pollution data over the course of the study.

So, scanning down the first column here (resting heart rate, in blue), one sees the magnitude and direction (positive or negative) of correlation between this participant’s resting heart rate and the following air pollutants:

- row 3 air pollutant (PM 2.5) = +0.1  \(\Rightarrow\) as PM 2.5 levels rose, resting heart rate rose slightly
- row 4 air pollutant (Ozone) = -0  \(\Rightarrow\) no registered correlation between Ozone levels and resting heart rate
- row 5 air pollutant (Sulfur Dioxide) = -0.1  \(\Rightarrow\) as Sulfur Dioxide levels rose, resting heart rate declined slightly
- row 6 air pollutant (Carbon Monoxide) = -0.2  \(\Rightarrow\) as Carbon Monoxide levels rose, resting heart rate declined slightly
- row 7 air pollutant (Total Hydrocarbons) = +0.3  \(\Rightarrow\) as Total Hydrocarbon levels rose, resting heart rate rose somewhat
- row 8 air pollutant (Methane) = +0.3  \(\Rightarrow\) as Total Hydrocarbon levels rose, resting heart rate rose somewhat
To zoom in further to analyze the data, Data Sense users can click on any rectangle of this grid, which will lead that portion of the grid to open in full-screen mode. Clicking in the upper righthand corner, for instance, at the intersection of “Heart” and “Methane,” will lead the following graph to open in full-screen mode.

![Graph showing positive correlation between heart rate and methane levels](image)

Above: this line graph shows a positive correlation (+0.3) between this participant’s resting heart rate and levels of methane in the air. So as methane rises, so does this person’s heart rate – though any correlation could be due to multiple factors.

**Do Positive Correlations = Positive Health Impacts? (No, it depends.)**

In general, a positive correlation between pollutants and resting heart rate would indicate a negative health impact, rather than a positive one – i.e. increased pollution = higher heart rate = worse health.

Yet for blood oxygen saturation (SPO2), the situation is reversed: a positive correlation between pollutant levels and SPO2 would indicate more oxygenated blood, or a positive health impact.

**How Did the Technologies Work?**

Multiple technological challenges arose in the course of the study. Participants and researchers encountered difficulties with both the data collection technologies (i.e. the apps and Fitbits) and the data analysis technologies (i.e. the Data Sense platform, as well as a computer code written in the programming language known as “R,” to analyze participants’ collective data).

- While intriguing to some participants, the **Fitbits and apps** were also challenging and burdensome to many – especially because the study required orientation to multiple apps and devices at the same time in order to begin data collection.
- The **Data Sense platform** was also challenging to navigate for many participants – including to add data sources that required multiple steps to connect, or whose connections (Application Programming Interfaces, or APIs) might expire and need to be reconnected. For researchers, Data Sense did not offer adequate tools for data aggregation and analysis, across either participants or data streams.
- In addition, researchers ran into unexpected difficulties running the **R computer code** that was written specifically to analyze the study’s datasets in a coherent way.
• Last but not least, the **refinery fenceline air monitoring systems** are inadequate to detect all the pollution in the air, which may be below the monitors’ detection thresholds, or otherwise not captured (e.g. during refinery flaring events that send pollution above fenceline air monitors, while a network of community air monitors are not in place to detect pollution wafted further afield).

**Participant Expectations**

Again, while researchers articulated the goals of the Real-Time Health Monitoring Study in terms of new technological possibilities for making sense of environmental and health data, study participants were most oriented toward its implications for personal and community health. In the initial survey, most participants described their interest in terms of health, e.g.: “I am interested in the impact living near a refinery and freeway is having on my health and the health of my community. I hope that by being part of this study I will gain a clearer picture of what impact the air pollution is having, if any.”

Only one participant articulated what they hoped to learn in terms of technology, saying, “I understand that we are testing these technologies including the Fitbit and oximeter and the apps to see if they could be used to monitor health of the Crockett population in case of some kind of toxic release from the Rodeo Phillips 66 refinery.” Researchers’ and participants’ overlapping, yet somewhat distinct, sets of goals mean that these research results may not fulfill everyone’s expectations equally.

**Key Take-Aways**

• The Real-Time Health Monitoring Study primarily surfaced research findings on technological contingencies and limitations, rather than environmental or community health. The inadequacies of air monitoring technologies, consumer-grade wearable devices, and data analysis tools were among the multiple reasons that the study did not find possible correlations between air pollution and participants’ health indicators.

• In the case of refinery fenceline air monitors, this study’s “non-finding” could point to the need for improved monitoring infrastructures and data access – including improved enforcement of Bay Area Air Quality Management District (BAAQMD) regulations, such as Rule 12-15. Currently regulators are more than a year behind in releasing refinery emissions inventories, as mandated under this rule. Regulators have also not addressed the high rates of fenceline monitor non-detects, in which pollutants may still be present at levels below the monitors’ detection thresholds. Or, pollutants may not be detected due to the locations and sparse placement of fenceline monitors, compared with the higher rates of pollutant detection enabled by monitors installed throughout the community.

• In addition, the large datasets generated for each participant in the course of this study may be based on quite different frequencies of data collection for each data stream (e.g. minute-by-minute heart rate tracking, compared with twice daily SPO2 readings, or hourly air pollutant averages). These differences complicate the aggregation and correlation analysis of disparate data streams in a consistent way – for instance, at the same timescale. Such discrepancies also complicate any singular notion of “real-time” data, or the idea of a singular cloud of “big data” waiting to be mined for new insights.

• Hence, it is crucial to consider how new types of sensors, wearable health trackers, apps, and other technologies may seem to offer new bridges between people’s embodied experiences of health and their surrounding environments. But, demonstrating coherent and rigorous relationships between environmental conditions and people’s health through new data streams is no simple matter.
New technologies are not panaceas to fill in gaps in data, nor to intervene in community and environmental health. Existing forms of personal and environmental health documentation (e.g. activist reports, oral histories, journalism, photography), bodies of public health and social science research, local knowledge, political understanding, an ethic of prevention of harm, and principles of environmental justice remain crucial guides to action.

Finally, it is crucial to recognize how decision-makers may try to substitute monitoring for meaningful policy change to reduce and prevent pollution in the first place – or to delay such actions. Not only new monitoring infrastructures, but new policies to cap refinery emissions (e.g. BAAQMD Rule 12-16), and to reduce and prevent pollution on the basis of existing data, are needed.

References & Data Sources

Air Watch Bay Area (http://www.airwatchbayarea.org) air monitoring data have two primary sources: 1) http://fenceline.org/, and 2) the Environmental Sensor Data Repository (ESDR). Yet, unlike the real-time monitoring data available through http://fenceline.org/, Air Watch Bay Area has data from as far back as May 2015. Meanwhile ESDR includes both fenceline air monitoring data and data from monitors set up further away, by community residents – such as Hound and PurpleAir monitors. More information on: BAAQMD Rule 12-15 (“Petroleum Refinery Emissions Tracking”) and BAAQMD Rule 12-16 (“Petroleum Refining Greenhouse Gas Emissions Limits”) is at: http://www.baaqmd.gov

More information on Data Sense (a project of Intel Labs) is at: https://makesenseofdata.com/faq#!/faq

Acknowledgments

Thank you to all of the participants in the Real-Time Health Monitoring Study, for the time and attention they devoted to this research. Thanks as well to Dawn Nafus, Gwen Ottinger, and Nancy Rieser – along with her fenceline community colleagues – for all of their work in designing this study and bringing it to fruition, from its inception in their 2016 Real-Time Health Monitoring Pilot Study. This research was funded by National Science Foundation (NSF) Award #1352143.

Appendix 1

Below are tables that provide a summary of average (mean) and median correlations between study participants’ local air monitoring data and their personal health indicator data. None of these tables show significant patterns correlating air quality and health indicators over the course of the study, as discussed further in the body of this research brief.

Given the different pollution patterns detected by the Rodeo North monitor versus the Rodeo South monitor, we examined participants’ health data along with only the pollutant data from their nearest monitor, for all pollutants except PM 2.5. The PM 2.5 data came from a third monitor (Rodeo 2B) located south of Rodeo, in Hercules, rather than a fenceline monitor. These PM 2.5 data were added to all participants’ correlation grids in Data Sense – though they may be most relevant to Rodeo participants, considering their proximity.

<table>
<thead>
<tr>
<th>Rodeo participants, correlations between air pollutants (Rodeo South Monitor) &amp; resting heart rate</th>
<th>Ozone</th>
<th>Sulfur Dioxide</th>
<th>Carbon Monoxide</th>
<th>Total Hydrocarbons</th>
<th>Methane</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resting heart rate correlations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>-.218</td>
<td>-.05</td>
<td>.191</td>
<td>-.045</td>
<td>.009</td>
</tr>
<tr>
<td>Median</td>
<td>-.3</td>
<td>0</td>
<td>.1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>.279</td>
<td>.184</td>
<td>.413</td>
<td>.515</td>
<td>.311</td>
</tr>
</tbody>
</table>
### Rodeo participants, correlations between air pollutants (Rodeo South Monitor) & SPO2

<table>
<thead>
<tr>
<th>SPO2 correlations</th>
<th>Ozone</th>
<th>Sulfur Dioxide</th>
<th>Carbon Monoxide</th>
<th>Total Hydrocarbons</th>
<th>Methane</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>.045</td>
<td>0</td>
<td>-.055</td>
<td>.036</td>
<td>.009</td>
</tr>
<tr>
<td>Median</td>
<td>.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>.225</td>
<td>.110</td>
<td>.242</td>
<td>.238</td>
<td>.170</td>
</tr>
</tbody>
</table>

### Crockett participants, correlations between air pollutants (Rodeo North Monitor) & resting heart rate

<table>
<thead>
<tr>
<th>Resting heart rate correlations</th>
<th>Ozone</th>
<th>Sulfur Dioxide</th>
<th>Carbon Monoxide</th>
<th>Total Hydrocarbons</th>
<th>Methane</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>-.025</td>
<td>.05</td>
<td>.013</td>
<td>.031</td>
<td>-.031</td>
</tr>
<tr>
<td>Median</td>
<td>-.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>.286</td>
<td>.197</td>
<td>.354</td>
<td>.236</td>
<td>.346</td>
</tr>
</tbody>
</table>

### Crockett participants, correlations between air pollutants (Rodeo North Monitor) & SPO2

<table>
<thead>
<tr>
<th>SPO2 correlations</th>
<th>Ozone</th>
<th>Sulfur Dioxide</th>
<th>Carbon Monoxide</th>
<th>Total Hydrocarbons</th>
<th>Methane</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>.037</td>
<td>-.021</td>
<td>.005</td>
<td>.068</td>
<td>.063</td>
</tr>
<tr>
<td>Median</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.1</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>.239</td>
<td>.202</td>
<td>.168</td>
<td>.226</td>
<td>.229</td>
</tr>
</tbody>
</table>

### All participants, correlations between PM 2.5 (Rodeo 2B monitor, in Hercules) & resting heart rate

<table>
<thead>
<tr>
<th>PM 2.5</th>
<th>Resting heart rate correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0</td>
</tr>
<tr>
<td>Median</td>
<td>0</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>.198</td>
</tr>
</tbody>
</table>

### All participants, correlations between PM 2.5 (Rodeo 2B monitor, in Hercules) & SPO2

<table>
<thead>
<tr>
<th>PM 2.5</th>
<th>SPO2 correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>-.027</td>
</tr>
<tr>
<td>Median</td>
<td>-.1</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>.178</td>
</tr>
</tbody>
</table>