Learning From Failure

Noah Wright

Table of contents

## Abstract

Incarcerated youth are an exceptionally vulnerable population, and body-worn cameras are an important tool of accountability for both incarcerated youth and the staff that supervise them. In 2018 the Texas Juvenile Justice Department deployed body-worn cameras for the first time, and this is a case study of how the agency developed a methodology for measuring the success of the camera rollout. This is also a case study of analysis failure, as it became clear that real-world implementation problems were corrupting the data and rendering the methodology unusable. However, the process of working through the causes of this failure helped the agency identify previously unrecognized problems and ultimately proved to be of great benefit. The purpose of this case study is to demonstrate how negative findings can still be incredibly useful in real-world settings.

## Background

### Why Body-worn Cameras?

Body-worn cameras became a standard tool of policing in the mid-2010s. By recording officer interactions with the public, law enforcement agencies could achieve a greater degree of accountability. Not only could credible claims of police abuse against civilians be easily verified, the argument went, but false accusations would decline as well, saving the agency time and resources that would otherwise be wasted on spurious allegations. Initial studies seemed to support this argument, and by the end of the decade it became common practice in the United States for police to have these devices.

The Texas Juvenile Justice Department faced similar issues, and body-worn cameras seemed like they could be a useful tool. Secure youth residential facilities in Texas all had overhead cameras, but these were very old (they still ran on tape) and captured no audio. This presented a number of problems when it came to deciphering contested incidents, not to mention that these cameras had clearly not prevented any of the agency’s prior scandals from taking place. TJJD received special funding from the legislature to roll out body-worn cameras system-wide, and all juvenile correctional officers were required to wear one.

The presence of body-worn cameras would, in theory, provide a strong behavioral incentive to both TJJD staff and youth alike, as well as provide evidence that could expedite any investigations of wrongdoing. But from the outset, the agency faced a major issue with implementation: In 2019, body worn cameras were mostly a tool for law enforcement. There was very little literature or best practice to draw from for their use in a correctional environment. Unlike police officers, juvenile correctional officers deal directly with their charges for virtually their entire shift. In an 8-hour shift, a police officer might record a few calls and traffic stops. A juvenile correctional officer, on the other hand, would record 8 consecutive hours. And because TJJD recorded round-the-clock for hundreds of employees at a time, this added up very quickly.

A typical dorm might have four JCOs assigned to it. Across a single week, these four JCO’s would be expected to record at least 160 hours of footage.

|  |
| --- |
| *4 JCOs x 40 hours per week =* ***160 hours of footage*** |

This was replicated across every dorm. Three dorms, for example, would produce nearly 500 hours of footage, as seen below.

|  |
| --- |
| *3 dorms x 4 JCOs x 40 hours per week =* ***480 hours of footage*** |

Finally, we had more than one facility. Four facilities with 3 dorms each would produce nearly 2,000 hours of footage every wekek.

|  |
| --- |
| 4 facilities x 3 dorms x 4 JCOs x 40 hours per week = **1,960 hours of footage** |

In actuality, we had a total of 5 facilities each with over a dozen dorms producing an anticipated **17,000 hours** of footage every week. To put this into perspective, a week only has 168 hours. That meant the facilities produced 100 hours of footage per hour of the day every week, an impossible amount to monitor manually.

As a result, footage review had to be done in a limited, reactive manner. If our monitoring team received an incident report they could easily zero in on the cameras of the officers involved and review the incident accordingly. But our executive team had hoped to be able to use the footage proactively, looking for ‘red flags’ in order to *prevent* potential abuses instead of only responding to allegations.

Because the agency had no way of automating the monitoring of footage, any proactive analysis had to be metadata-based. But what to look for in the metadata? Once again, the lack of best-practice literature left us in the lurch. So, we brainstormed ideas for ‘red flags’ and came up with the following that could be screened for using camera metadata:

1. **Minimal quantity of footage** – our camera policy required correctional officers to have their cameras on at all times in the presence of youth. No footage meant they weren’t using their cameras.
2. **Frequently turning the camera on and off** – a correctional officer working a dorm should have their cameras always on when around youth and not be turning them on and off repeatedly.
3. **Large gaps between clips** – It defeats the purpose of having cameras if they’re not turned on.

* In addition, we came up with a fourth red flag that could be screened for using camera metadata compared with shift-tracking metadata:

1. **Mismatch between clips recorded and shifts worked** ­– the agency had very recently rolled out a new shift tracking software. We should expect to see the hours logged by the body cameras roughly match the hours logged as shifts worked.

## Analysis Part 1 – QC and Footage Analysis

For this analysis, I gathered the most recent three weeks of body-worn camera data, which, at the time, covered April 1–21, 2019. I also pulled data from Shifthound (our shift management software) covering the same time period. Finally, I gathered HR data from CAPPS, the system that most of the State of Texas used at the time for personnel management and finance.[[1]](#footnote-32) I then performed some quality control work, summarized in the inset below.

### Initial Quality Control Steps

SkimR is a helpful package for exploratory analysis that gives summary statistics for every variable in a data frame, including missing values. After using the skim function on clip data, shift data, and HR data, I noticed that the Clip data had missing values for employee ID. This was an error (body-worn cameras do not record footage on their own) that indicated data entry mistakes.

From here I compared the employee ID field in the clip data to the employee ID field in the HR data. IDs existed in the clip data that did not correspond to any entries in the HR data, indicating more erroneous entries like the missing values. The HR data is the single source of truth for employee IDs as it is literally where they all come from. I checked the shift data for the same error - employee IDs that did not exist in the HR data - and found it. Data entry issues clearly existed for the shift data as well.

In addition to employee ID’s that did not exist in the HR data, I also looked for employee ID’s that existed in the HR data but for staff who were not actually employed between April 1 – 21, 2019. These also indicated clear errors: staff cannot use a body-worn camera or log a shift if they have yet to begin working or if they have been terminated (system permissions are revoked upon leaving employment).

I made a list of every erroneous ID to pass off to HR and monitoring staff before excluding them from the subsequent analysis. In total, 10.6% of clips representing 11.3% of total footage had to be excluded due to these initial data quality issues, foreshadowing the subsequent data quality issues the analysis would eventually uncover.

The full script [can be found on GitHub here.](https://t.ly/BUNRZ)

In order to operationalize our brainstorming session, I needed to see what exactly the cameras captured in their metadata. The variables most relevant to our purposes were:

* Clip start
* Clip end
* Camera Used
* Who was assigned to the camera at the time
* The role of the person assigned to the camera

Using these fields, I first created the following aggregations per employee ID:

|  |
| --- |
|  |

* **Number of clips** = Number of clips recorded.
* **Days with footage** = Number of discrete dates that appear in these clips.
* **Footage Hours** = Total duration of all shot footage.
* **Significant Gaps** = Number of clips where the previous clip’s end date was either greater than 15 minutes or less than 8 hours before this clip’s start date.

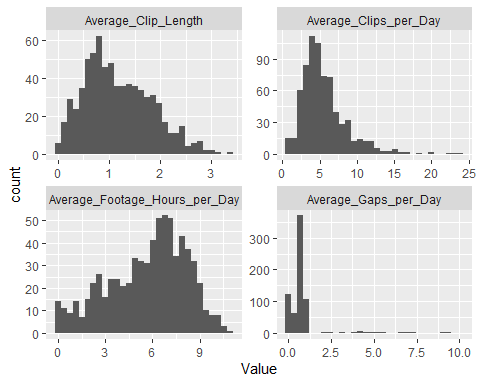
I used these to devise the following staff metrics:

|  |
| --- |
|  |

* **Clips per day** = Number of clips / Days with footage
* **Footage per day** = Footage hours / Days with footage
* **Average clip length** = Footage hours / Number of clips
* **Gaps per day** = Gaps / Days with footage

Once I established these metrics for each employee I looked at their respective distributions. Standard staff shift lengths at the time were 8 hours. If staff were using their cameras appropriately, we would expect to see distributions centered around clip lengths of about an hour, 8 or fewer clips per day, and 8-12 footage hours per day. We would also expect to see 0 large gaps.

library(tidyverse)  
  
Footage\_Metrics\_by\_Employee <- read\_csv("Output/Footage Metrics by Employee.csv")  
  
Footage\_Metrics\_by\_Employee %>%   
 select(-Clips, -Days\_With\_Footage, -Footage\_Hours, -Gaps) %>%   
 pivot\_longer(-Employee\_ID, names\_to = "Metric", values\_to = "Value") %>%   
 ggplot(aes(x = Value)) +  
 geom\_histogram() +  
 facet\_wrap(~Metric, scales = "free")



By eyeballing the distributions I could tell most staff were recording fewer than 10 clips per day, shooting about 0.5–2 hours for each clip, for a total of 2–10 hours of daily footage, with the majority of employees having less than one significant gap per day. Superficially, this appeared to provide evidence of widespread attempts at complying with the policy and no systemic rejection or resistance. If this were indeed the case, then we could turn our attention to individual outliers.

First, though, we thought we would attempt to validate this with another assumption. If each employee works on average 40 hours per week – a substantial underestimate given how common overtime was – we should expect, over a three-week period, to see about 120 hours of footage per employee in the dataset. This is *not* what we found.

Percent\_Display <- function(x) {  
 str\_c(100 \* round(x, 3), "%")  
}  
  
Total\_Footage <- sum(Footage\_Metrics\_by\_Employee$Footage\_Hours)  
Total\_Employees <- nrow(Footage\_Metrics\_by\_Employee)  
Average\_per\_Employee <- round((Total\_Footage/Total\_Employees),1)  
  
print(str\_c("Average Footage per Employee - ", Average\_per\_Employee, " hours"))

[1] "Average Footage per Employee - 70.2 hours"

print(str\_c(Average\_per\_Employee, " footage hours / 120 expected work hours = ", 100\*Average\_per\_Employee/120, "% of average shift recorded"))

[1] "70.2 footage hours / 120 expected work hours = 58.5% of average shift recorded"

Even with the unrealistic assumption of no overtime pushing the estimated total working hours down, less than 60% of these hours appear in the footage. A perfect match would be highly unlikely, but the fact that more than 40% of anticipated hours are unrecorded for unknown reasons warranted a follow-up.

Surely the shift data would clarify this.

## Analysis Part 2 – Footage and Shift Comparison

With the data on shifts worked from our timekeeping system, I could theoretically compare actual shifts worked to the amount of footage recorded. If there were patterns in where the gaps in footage fell, that might help to explain why.

In order to join the shift data to the camera data, I needed a common unit of analysis beyond “Employee ID”. Using only this value would produce a nonsensical table that joined up every clip of footage to every shift worked.

For example, let’s take employee #9001005 at Facility Epsilon between April 1–3. This employee has the following clips recorded during that time period:

| Employee\_ID | Clip\_ID | Clip\_Start | Clip\_End |
| --- | --- | --- | --- |
| 9001005 | 156421 | 2019-04-01 05:54:34 | 2019-04-01 08:34:34 |
| 9001005 | 155093 | 2019-04-01 08:40:59 | 2019-04-01 08:54:51 |
| 9001005 | 151419 | 2019-04-01 09:03:16 | 2019-04-01 11:00:30 |
| 9001005 | 153133 | 2019-04-01 11:10:09 | 2019-04-01 12:39:51 |
| 9001005 | 151088 | 2019-04-01 12:57:51 | 2019-04-01 14:06:44 |
| 9001005 | 150947 | 2019-04-02 05:56:34 | 2019-04-02 09:48:50 |
| 9001005 | 151699 | 2019-04-02 09:54:23 | 2019-04-02 12:17:15 |

We can join this to a similar table of shifts logged. This particular employee had the following shifts scheduled from April 1–3:

| Employee\_ID | Shift\_ID | Shift\_Start | Shift\_End |
| --- | --- | --- | --- |
| 9001005 | E050603 | 2019-04-01 06:00:00 | 2019-04-01 14:00:00 |
| 9001005 | E051303 | 2019-04-02 06:00:00 | 2019-04-02 14:00:00 |

Two 8-hour morning shifts from 6:00 AM to 2:00 PM. We can join the two tables together by ID on a messy many-to-many join, but that tells us nothing about how much they overlap (or fail to overlap) without extensive additional work.

We have a unique identifier for employee-clip and employee-shift, but what we need is a unique identifier that can be used to join the two. Fortunately, for this particular data we can *create* a unique identifier since both clips and shifts are fundamentally measures of *time*.

So, while Employee ID is not in itself unique (i.e., one employee can have multiple clips), Employee ID combined with time of day is unique. After all, a person can only be in one place at one time!

To reshape the data for joining, I created a function that takes any data frame with a start and end column and unfolds it into discrete units of time. Using the code below to create the “Interval\_Convert” function, the shift data above for employee 9001005 converts into one entry per hour of the day per shift. As a result, two 8-hour shifts get turned into 16 Employee-Hours.

library(sqldf)  
library(lubridate)  
  
Interval\_Convert <- function(DF, Start\_Col, End\_Col, Int\_Unit, Int\_Length = 1) {  
  
   
 Start\_Col2 <- enquo(Start\_Col)  
 End\_Col2 <- enquo(End\_Col)  
   
 Start\_End <- Sample\_Shifts %>%  
 ungroup() %>%  
 summarize(Min\_Start = min(!!Start\_Col2),  
 Max\_End = max(!!End\_Col2)) %>%  
 mutate(Start = floor\_date(Min\_Start, Int\_Unit),  
 End = ceiling\_date(Max\_End, Int\_Unit))  
   
 Sample\_Shifts <- Sample\_Shifts %>%  
 mutate(Single = !!Start\_Col2 == !!End\_Col2)  
   
 Interval\_Table <- data.frame(Interval\_Start = seq.POSIXt(Start\_End$Start[1], Start\_End$End[1], by = str\_c(Int\_Length, " ", Int\_Unit))) %>%  
 mutate(Interval\_End = lead(Interval\_Start)) %>%  
 filter(!is.na(Interval\_End))  
   
 by <- join\_by(Interval\_Start <= !!End\_Col2, Interval\_End >= !!Start\_Col2)   
   
 Interval\_Data\_Table <- Interval\_Table %>%   
 left\_join(Sample\_Shifts, by) %>%   
 mutate(Seconds\_Duration\_Within\_Interval = if\_else(!!End\_Col2 > Interval\_End, Interval\_End, !!End\_Col2) -  
 if\_else(!!Start\_Col2 < Interval\_Start, Interval\_Start, !!Start\_Col2)) %>%  
 filter(!(Single & Interval\_End == !!Start\_Col2),  
 as.numeric(Seconds\_Duration\_Within\_Interval) > 0)  
   
 return(Interval\_Data\_Table)  
}

| Interval\_Start | Interval\_End | Employee\_ID | Shift\_ID | Shift\_Start | Shift\_End | Seconds\_Duration\_Within\_Interval |
| --- | --- | --- | --- | --- | --- | --- |
| 2019-04-01 06:00:00 | 2019-04-01 07:00:00 | 9001005 | E050603 | 2019-04-01 06:00:00 | 2019-04-01 14:00:00 | 3600 secs |
| 2019-04-01 07:00:00 | 2019-04-01 08:00:00 | 9001005 | E050603 | 2019-04-01 06:00:00 | 2019-04-01 14:00:00 | 3600 secs |
| 2019-04-01 08:00:00 | 2019-04-01 09:00:00 | 9001005 | E050603 | 2019-04-01 06:00:00 | 2019-04-01 14:00:00 | 3600 secs |
| 2019-04-01 09:00:00 | 2019-04-01 10:00:00 | 9001005 | E050603 | 2019-04-01 06:00:00 | 2019-04-01 14:00:00 | 3600 secs |
| 2019-04-01 10:00:00 | 2019-04-01 11:00:00 | 9001005 | E050603 | 2019-04-01 06:00:00 | 2019-04-01 14:00:00 | 3600 secs |
| 2019-04-01 11:00:00 | 2019-04-01 12:00:00 | 9001005 | E050603 | 2019-04-01 06:00:00 | 2019-04-01 14:00:00 | 3600 secs |
| 2019-04-01 12:00:00 | 2019-04-01 13:00:00 | 9001005 | E050603 | 2019-04-01 06:00:00 | 2019-04-01 14:00:00 | 3600 secs |
| 2019-04-01 13:00:00 | 2019-04-01 14:00:00 | 9001005 | E050603 | 2019-04-01 06:00:00 | 2019-04-01 14:00:00 | 3600 secs |
| 2019-04-02 06:00:00 | 2019-04-02 07:00:00 | 9001005 | E051303 | 2019-04-02 06:00:00 | 2019-04-02 14:00:00 | 3600 secs |
| 2019-04-02 07:00:00 | 2019-04-02 08:00:00 | 9001005 | E051303 | 2019-04-02 06:00:00 | 2019-04-02 14:00:00 | 3600 secs |
| 2019-04-02 08:00:00 | 2019-04-02 09:00:00 | 9001005 | E051303 | 2019-04-02 06:00:00 | 2019-04-02 14:00:00 | 3600 secs |
| 2019-04-02 09:00:00 | 2019-04-02 10:00:00 | 9001005 | E051303 | 2019-04-02 06:00:00 | 2019-04-02 14:00:00 | 3600 secs |
| 2019-04-02 10:00:00 | 2019-04-02 11:00:00 | 9001005 | E051303 | 2019-04-02 06:00:00 | 2019-04-02 14:00:00 | 3600 secs |
| 2019-04-02 11:00:00 | 2019-04-02 12:00:00 | 9001005 | E051303 | 2019-04-02 06:00:00 | 2019-04-02 14:00:00 | 3600 secs |
| 2019-04-02 12:00:00 | 2019-04-02 13:00:00 | 9001005 | E051303 | 2019-04-02 06:00:00 | 2019-04-02 14:00:00 | 3600 secs |
| 2019-04-02 13:00:00 | 2019-04-02 14:00:00 | 9001005 | E051303 | 2019-04-02 06:00:00 | 2019-04-02 14:00:00 | 3600 secs |

The footage can be converted in a similar manner, and this way I could break down both the shift data and the clip data into an hour-by-hour view and compare them to one another. Using this new format, I joined together the full tables of footage and shifts to determine how much footage was recorded with no corresponding shift in the timekeeping system, and vice-versa.

| HR\_Location | Footage\_Hours\_No\_Shift | Employee\_IDs\_With\_Missing\_Shift |
| --- | --- | --- |
| Alpha | 1805 | 122 |
| Beta | 3749 | 114 |
| Delta | 1208 | 133 |
| Epsilon | 2899 | 157 |
| Gamma | 4153 | 170 |

To summarize, almost every employee has footage hours that do not match with logged shifts, totaling nearly 14,000 hours. But what about the opposite? How many hours of shifts got logged with no corresponding footage?

| HR\_Location | Shift\_Hours\_No\_Footage | Employee\_IDs\_With\_Missing\_Footage |
| --- | --- | --- |
| Alpha | 7338 | 127 |
| Beta | 6014 | 118 |
| Delta | 12830 | 141 |
| Epsilon | 9000 | 168 |
| Gamma | 11960 | 183 |

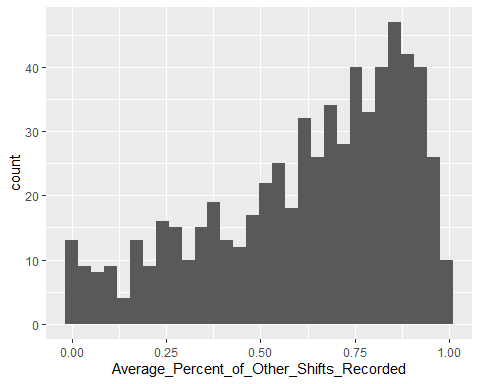
Oh dear.

That is also almost every employee, but this time totaling about 47,000 hours. To put it another way, that’s an entire work week per employee not showing up in camera footage.

At this point, we could probably rule out deliberate noncompliance. The clip data already implied that most employees were following the policy, and our facility leadership would have noticed a mass refusal large enough to show up this clearly in the data.

One way to check for deliberate noncompliance would be to first exclude shifts that contain zero footage whatsoever. This would rule out total mismatches, where – for whatever reason – the logged shifts had totally failed to overlap with recorded clips. For those remaining shifts that *do* contain footage, we could look at the proportion of the shift covered by footage. So, if an 8-hour shift had 4 hours of recorded footage associated with it, then we could say that 50% of the shift had been recorded. The following histogram is a distribution of the number of employees organized by the percent of their shift-hours they recorded (but only shifts that had a nonzero amount of footage).

Warning: Removed 156 rows containing non-finite values (`stat\_bin()`).



As it turned out, most employees recorded the majority of their matching shifts, a finding that roughly aligns with the initial clip analysis.

## Causes of Failure

Here, I believed, we had reached the end of what I could do with data alone. I presented these findings (or lack thereof) to executive leadership, and the failure to gather reliable data from linking the clip data to the shift data prompted follow-ups into what exactly was going wrong. As it turned out, *many* things were going wrong.

First, a number of technical problems plagued the early roll out of the cameras.

* All of our facilities suffered from high turnover, and camera ownership was not consistently updated. Employees who no longer worked at the agency could therefore appear in the clip data – somebody else had taken over their camera but had not put their name and ID on it.
* We had no way of telling if a camera was not recording due to being docked and recharging or not recording due to being switched off.
* In the early days of the roll out, footage got assigned to an owner based on the owner of the *dock*, not the camera. In other words, if Employee A had recorded their shift but uploaded the footage using a dock assigned to Employee B then the footage would show up in the system as belonging to Employee B.

The shift data was, unsurprisingly, even worse, and it was here we came across our most important finding. While the evidence showed that there wasn’t any widespread non-compliance with the use of the cameras, there *was* widespread non-compliance with the use of our shift management software. Details are included in the inset below.

### Quality Issues in Shift Tracking Data

Our HR system, CAPPS, had a feature that tracked hours worked in order to calculate leave and overtime pay. However, CAPPS was a statewide application designed for 9–5 office workers, and could not capture the irregular working hours of our staff (much less aid in planning future shifts).We had obtained the shift management software to fill these gaps, but not realized how inconsistently it was being used. All facilities were required to have their employees log their shifts, but some followed through on this better than others. And even for the ones that did make a good-faith effort at follow-through, quality control was nonexistent.

In CAPPS, time entry determined pay, so strong incentives existed to ensure accurate entry. But for our shift management software, no incentives existed at all for making sure that entries were correct. For example, a correctional officer could have a 40-hour work week scheduled in the shift software but miss the entire week due to an injury, and the software would still show them as having worked 40 hours that week. Nobody bothered to go back and correct these types of errors because there was no reason to.

The software was intended to be used proactively for planning purposes, not after-the-fact for logging and tracking purposes. Thus, the produced data that was totally inconsistent with actual hours worked, which became apparent when compared to data (like body-worn cameara footage) that tracked actual hours on the floor.

In the end, we had to rethink a number of aspects of the shift software’s implementation. In the process of these fixes, leadership also came to make explicit that the software’s primary purpose was to help facilities schedule future shifts, not audit hours worked after the fact (which CAPPS already did, just on a day-by-day basis as opposed to an hour-by-hour basis). This analysis was the only time we attempted to use the shift data in this manner.

## What we learned from failure

Whatever means we used to monitor compliance with the camera policy, we learned that it couldn’t be fully automated. The agency followed up this analysis with a random sampling approach, in which monitors would randomly select times of day they knew a given staff member would have to have their cameras turned on and actually watch the associated clips. This review process confirmed the first impressions from the statistical review above: most employees *were* making good faith attempts at complying with the policy despite technical glitches, short-staffing, and administrative confusion. It also confirmed that proactive monitoring of correctional officers was a human process which had to come from supervisors and staff.

The one piece of the analysis we did use going forward was the clip analysis (converted into a Power BI dashboard, included in the [GitHub repository](https://github.com/enndubbs/Body-Worn-Camera-Monitoring) for this article), but only as a supplement for already-launched investigations, not a prompt for one. Body-worn camera footage remained immensely useful for investigations after-the-fact, but inconsistencies in clip data in and of themselves were not particularly noteworthy ‘red flags’. At the end of the day, analytics can contextualize and enhance human judgment, but it cannot replace it.

In academia, the bias in favor of positive findings is well-documented. The failure to find something, or a lack of statistical significance, does not lend itself to publication in the same way that a novel discovery does. But, in an applied setting, where results matter more than publication criteria, negative findings can be highly insightful. They can falsify erroneous assumptions, bring unknown problems to light, and prompt the creation of new processes and tools. In this context, a failure is only truly a failure if nothing is learned from it.

1. The underlying data for the analysis as presented in this article was requested through the Texas Public Information Act and went through TJJD’s approval process for ensuring anonymity of records. It is available on GitHub along with the rest of the code used to write this article. [↑](#footnote-ref-32)