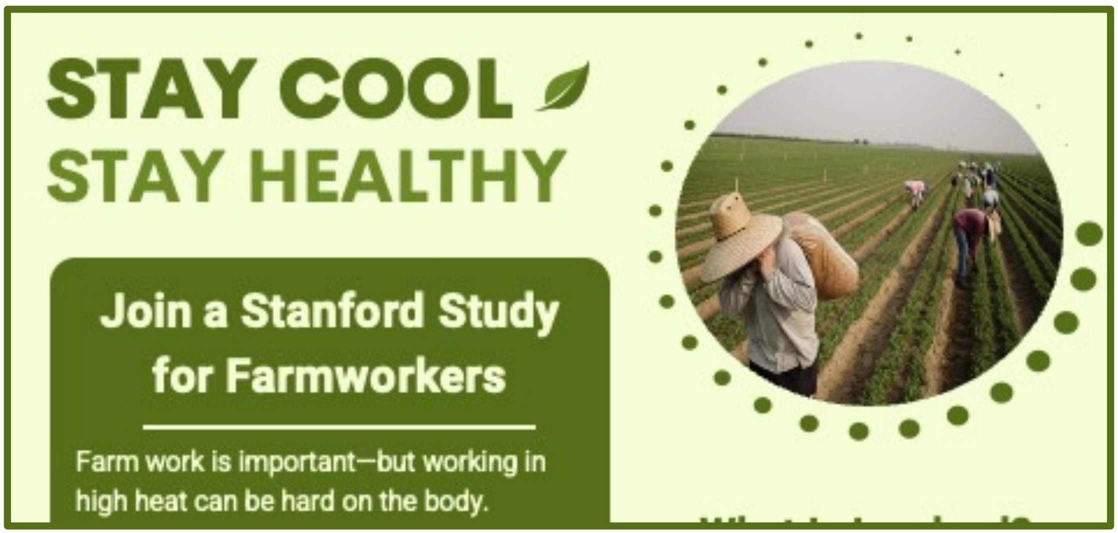
Modeling Longitudinal Core Temperature in a Crossover Trial of Farmworkers in California



Maria E. Montez Rath

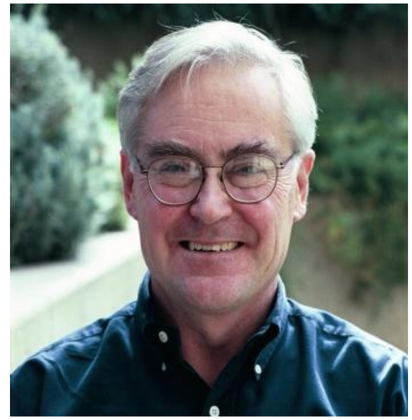
2026 Stata Virtual Forum: Biostatistics and Epidemiology

Stanford MEDICINE
School of Medicine
Biostatistics Core
Division of Nephrology

Our Team

Astrid Elliot

Catherine Ley



H. Craig Heller

Ernesto Orellana

Marimar Contreras Nieves

Julie Parsonnet

Maria Montez-Rath

Shuchi Anand

Study Context: Heat-Kidney Connection

- **Acute Risk:** Heat waves ⇔10–25% increase in hospitalizations for acute kidney injury
- **Workday Impact:** transient rise in serum creatinine (the primary marker of kidney function) across a single workday in heat-exposed workers
- **Global Epidemic:** Specific low-lying, high-heat regions (e.g., Sri Lanka, El Salvador, Nicaragua) are seeing an epidemic of CKDu (Chronic Kidney Disease of unknown etiology) among working-age populations
- **Compounding Effects:** Heat stress can significantly amplify the progression of pre-existing kidney disease

"Does repetitive heat stress from extended work in hot environments lead to kidney dysfunction?"

Existing Cal OSHA standards (enforceable)

🚰 **Water:** Free, Cool, & Close

⛱️ **Shade:** Required at 80F+; Universal access

⏱️ **Rest:** Encouraged & On-request

OSHA: US Occupational Safety and Health Administration

"Does repetitive heat stress from extended work in hot environments lead to kidney dysfunction?"

Existing Cal OSHA standards (enforceable)

🚰 **Water:** Free, Cool, & Close

⛱️ **Shade:** Required at 80F+ (27C+); Universal access

⏱️ **Rest:** Encouraged & On-request

Is this enough?

Pilot Study: 3-arm Crossover Trial

	Week 1	Week 2	Week 3
Group 1	Control	Mitts	Bandana
Group 2	Mitts	Bandana	Control
Group 3	Bandana	Control	Mitts

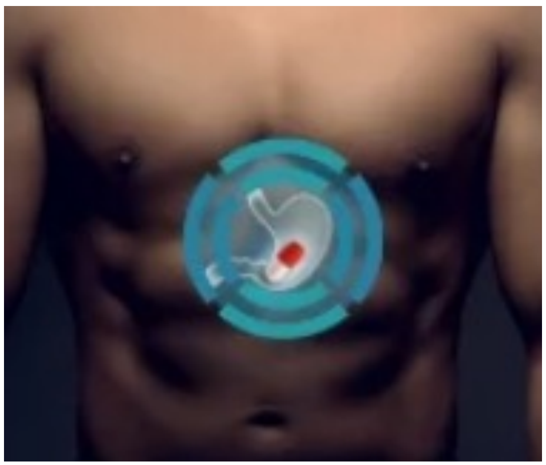
Control: Cal OSHA standards

Bandana

Mitts

Core Temperature

Measurement: Continuous core temperature monitoring via ingested telemetric pills



Raw Data:
~ 480 observations
per person/day
(8-hour shift)

Today’s talk

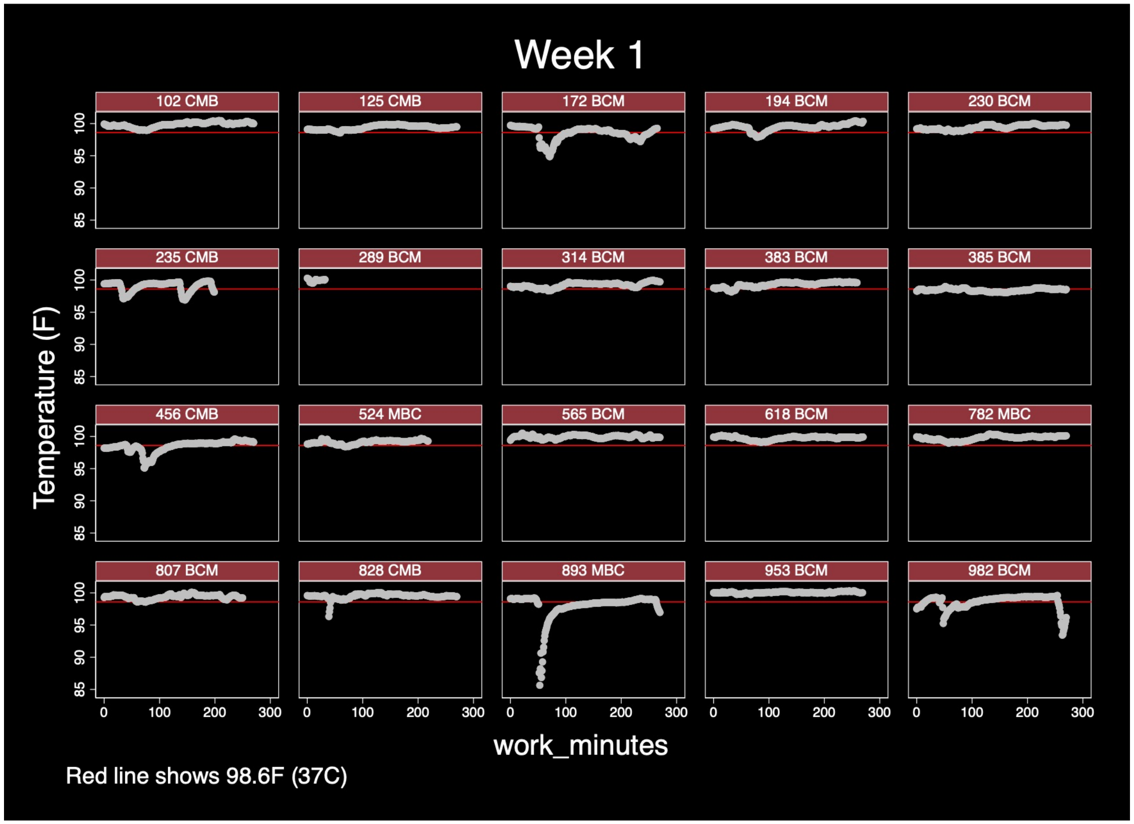
Population: Agricultural workers in California's Central Valley.

Problem: Occupational heat stress is often assessed via sporadic measurements, which miss cumulative "thermal load".

Objective: To evaluate the feasibility and comparative efficacy of cooling bandanas and mitts in a field setting.

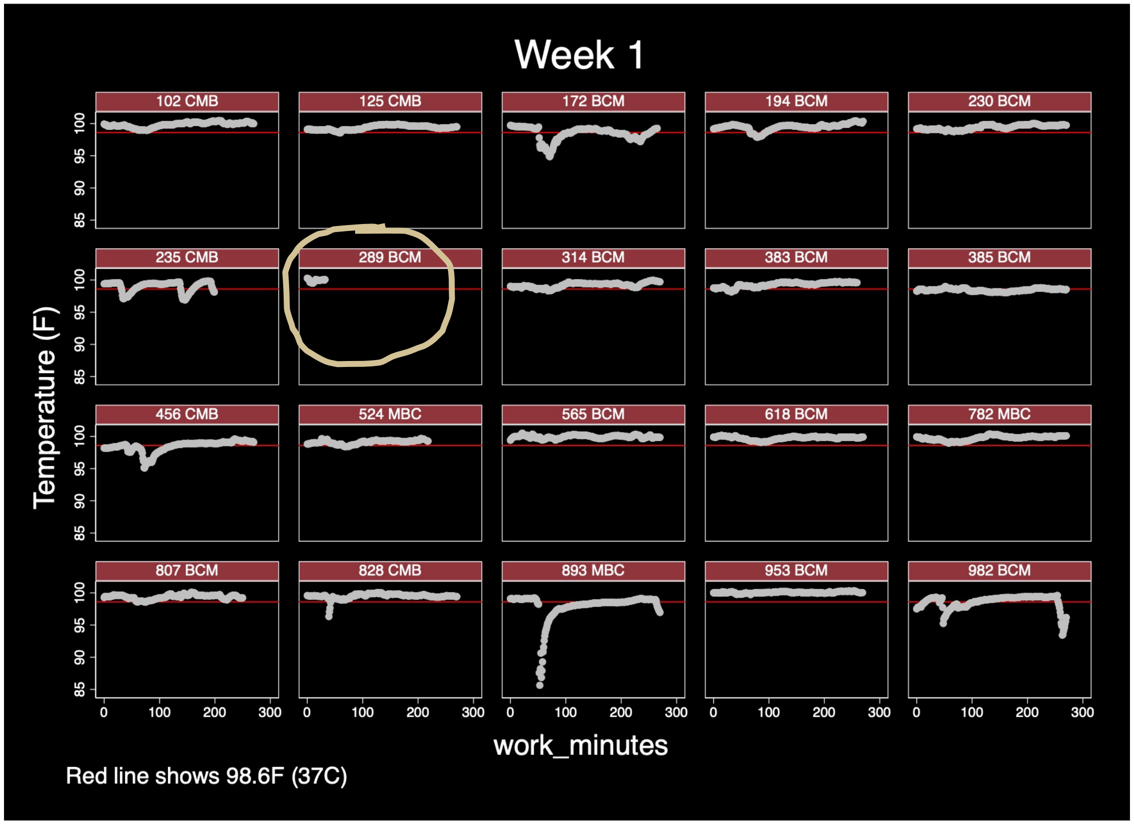
Focus: How to handle minute-by-minute core temperature data using Stata.

Some issues we will need to handle



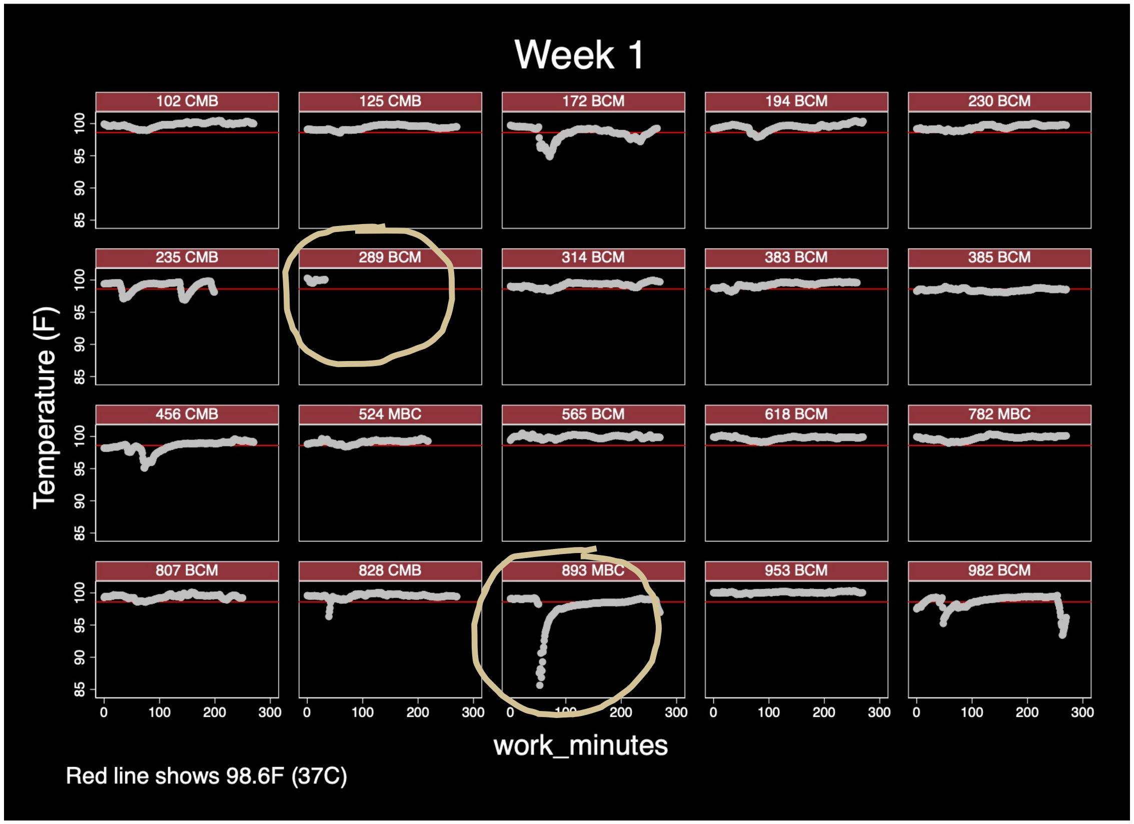
Some issues we will need to handle

Missing Data



Some issues we will need to handle

Missing Data



Sudden temperature drops (“artifact”)

Stanford MEDICINE
School of Medicine
Biostatistics Core
Division of Nephrology

Common protocols include

1. Change temperature artifacts to missing
2. Correction of “islands in the data”
3. 20% exclusion rule (drop days with insufficient data)
4. Missing data interpolation
5. Smoothing

The data

trial_week: 1, 2, 3

trial_arm: 0 (Control), 1 (Bandana), 2 (Mitts)

work_time: 9.5-14 corresponding to time 9:30 to 14:00

temperature – observed value

N individuals = 20

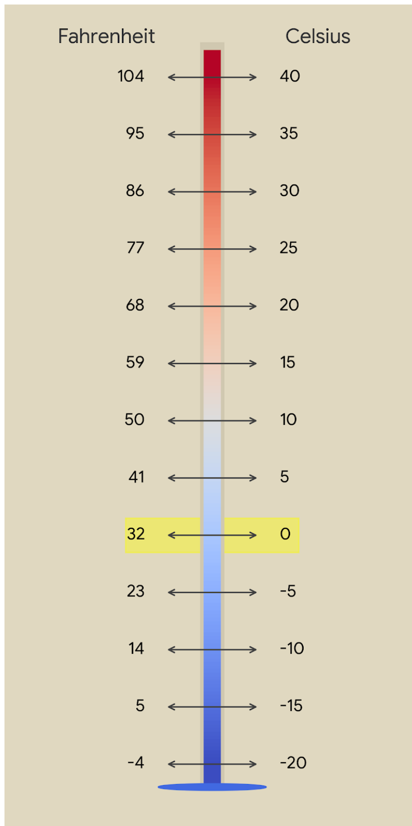
Note:

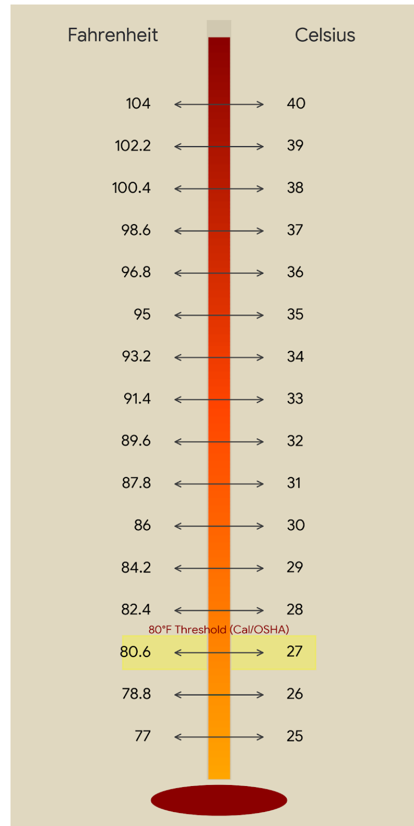
breaks at 8:30 and 13:00 for 15 min

lunch at 10:00 for 45 min

Stanford MEDICINE
School of Medicine
Biostatistics Core
Division of Nephrology

Temperature Conversion Chart





Getting data ready

```
*** Create a unique ID for every worker-week combination (60 unique shifts)
egen shift_id = group(id trial_week)

/**** I need to keep one row per minute for each individual each week even if the
temperature is missing.Between 9:30 and 14:00 there are 271 minutes ****/
*Fill in the gaps
fillin id trial_week work_time

* Fill any constant variable
foreach v of varlist trial_group trial_arm shift_id carryover {
    bysort id trial_week (`v'): replace `v' = `v'[1] if missing(`v')
}

* Declare the dataset as panel data
gen work_minutes = round((work_time - 9.5) * 60)
tsset shift_id work_minutes
```


Stanford MEDICINE
School of Medicine
Biostatistics Core
Division of Nephrology

Change temperature artifacts to missing

Artifact: where the rate of change exceeds the body's thermal inertia (> 0.1C or ~0.18F /min).

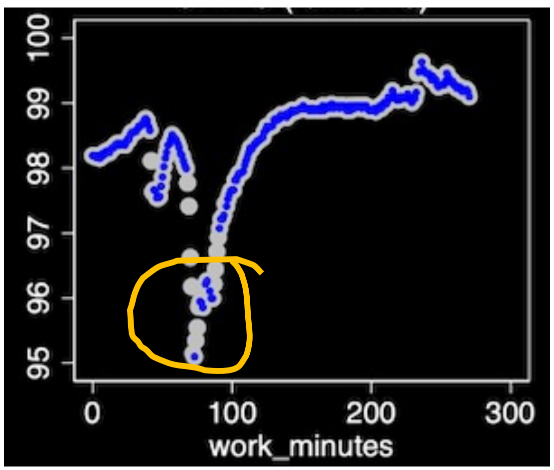
```
/* Flag temperature artifacts */
gen temp_diff = D.temperature
gen is_artifact = (temp_diff > 0.18) & !missing(temp_diff)

/* This requires manual inspection or a rule-based truncation at the last valid temp */
gen cleantemp_chips = temperature
replace cleantemp_chips =. if is_artifact
```


Stanford MEDICINE School of Medicine
Biostatistics Core
Division of Nephrology

Mitchell et al, 2017. California Heat Illness Prevention Study (CHIPS)

Correction of “islands in the data”

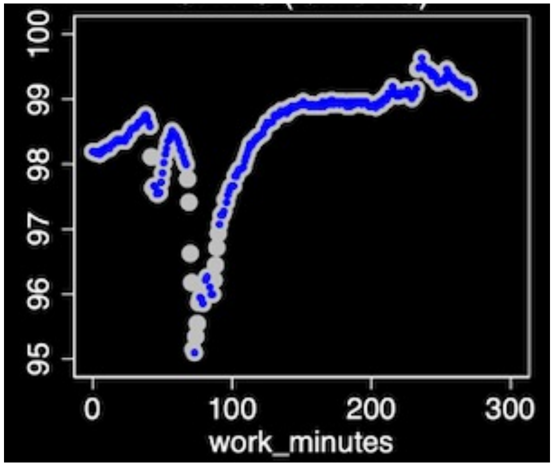
- Issue arises when algorithm correctly flags ”artifacts” but maintains temperatures at low levels because those are stable
- Look for islands of data surrounded by missing values and low temperatures. I restrict to 6-minute intervals.



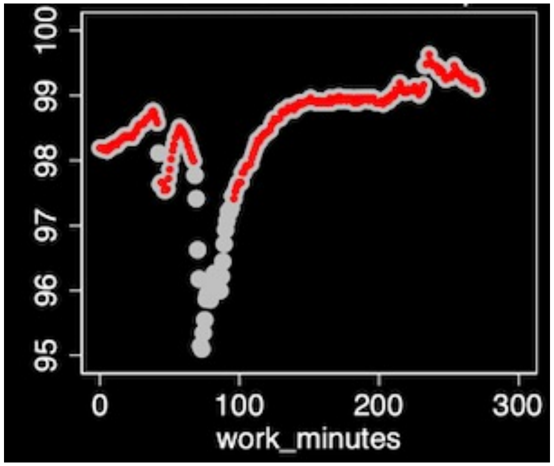

```
*** After creating cleantemp_chips variable
* Create a grouping variable for consecutive valid data points
gen valid = (is_artifact == 0)
bysort id (time_numeric): gen run_id = sum(is_artifact)

* Count how many valid points are in each "run"
bysort id run_id: egen run_length = count(cleantemp_chips) if valid == 1

* If the run is too short (e.g., 6 minutes or less), it's an island
gen cleantemp_chips2 = cleantemp_chips
replace cleantemp_chips2 = . if is_artifact == 1 | run_length <= 6
```

Correction of “islands in the data”

- Issue arises when algorithm correctly flags ”artifacts” but maintains temperatures at low levels because those are stable

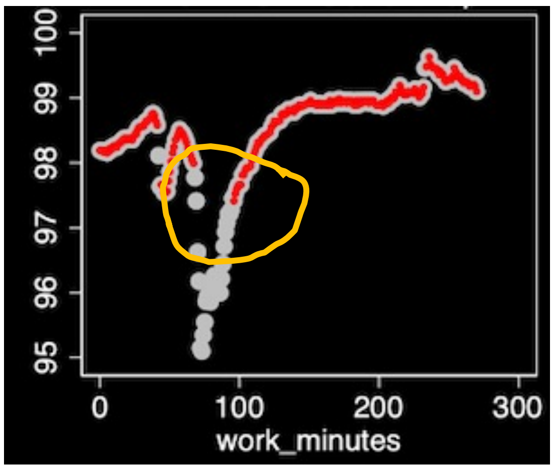


- Look for islands of data surrounded by missing values and low temperatures. I restrict to 6-minute intervals.



Missing data interpolation

- After removing the “island”

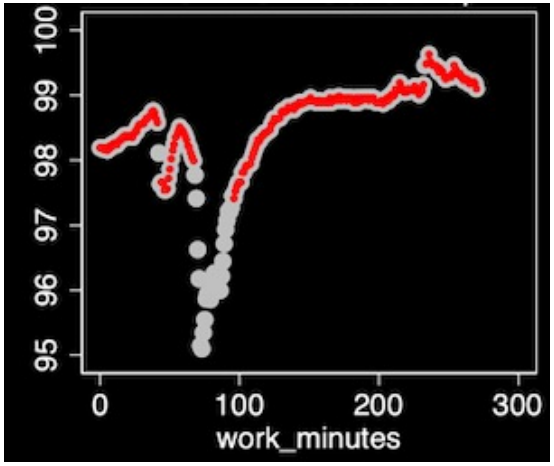


- Interpolate missing temperatures using natural cubic splines
- MIPOLATE: Stata module to interpolate yvar on xvar for missing values of yvar
- Take care to specify panels if needed

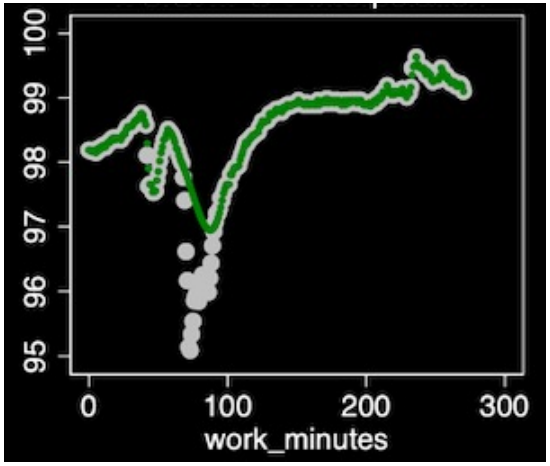
```
bys id trial_week: mipolate cleantemp_chips time_numeric, ///
spline gen(cleantemp_chips_csi)
```

Missing data interpolation

- After removing the “island”

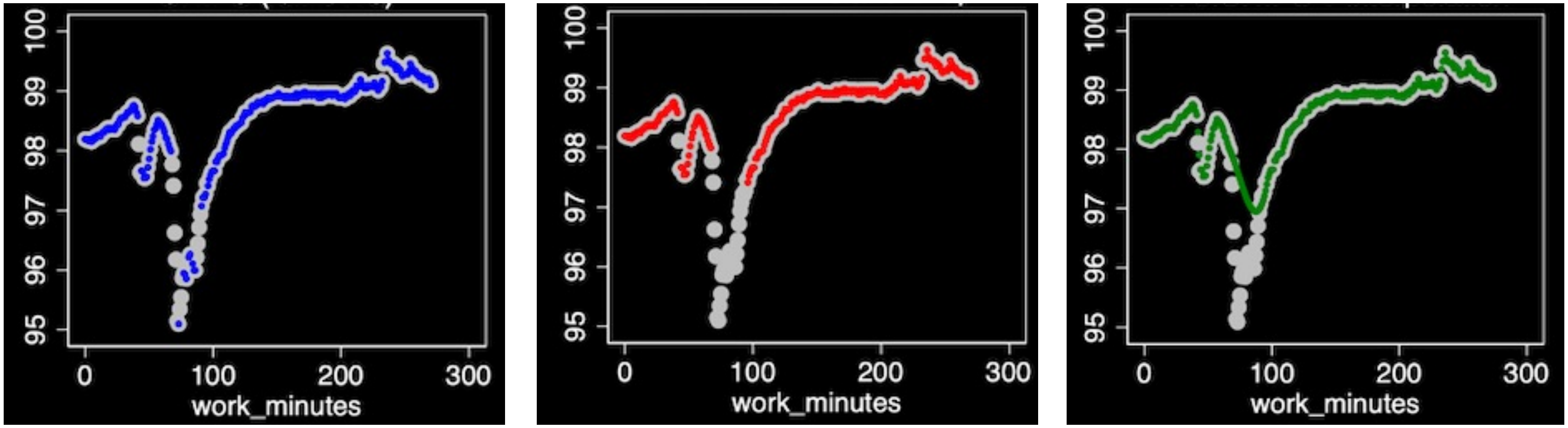


- Interpolate missing temperatures using natural cubic splines

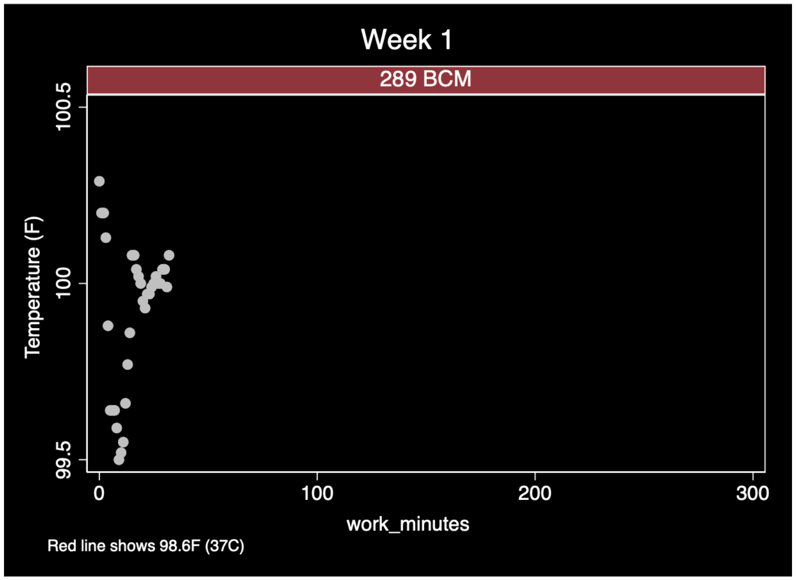


Note that interpolation requires the missing value to be in between observed values

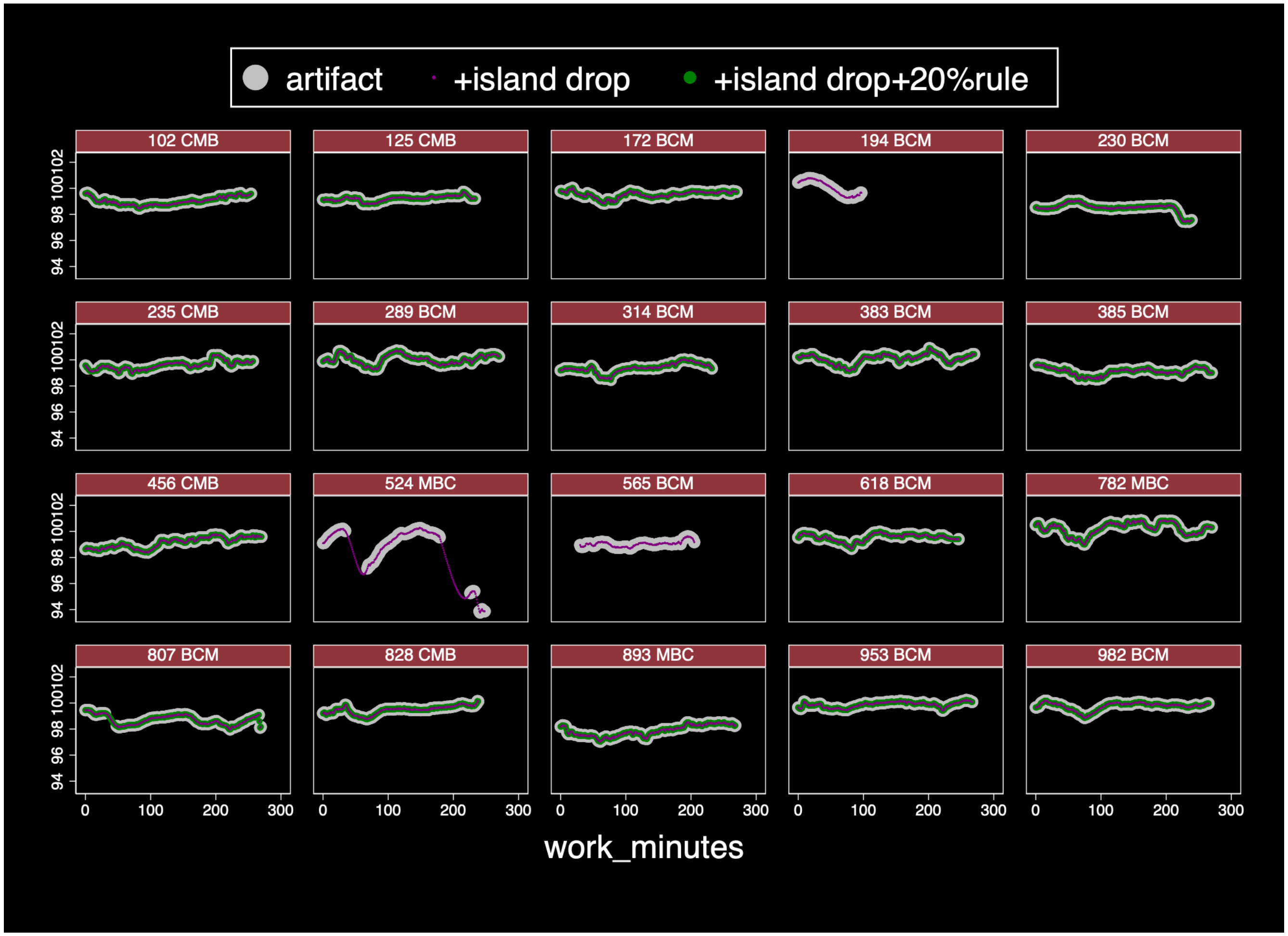
Complete sequence



Handling large amounts of missing data: 20% rule



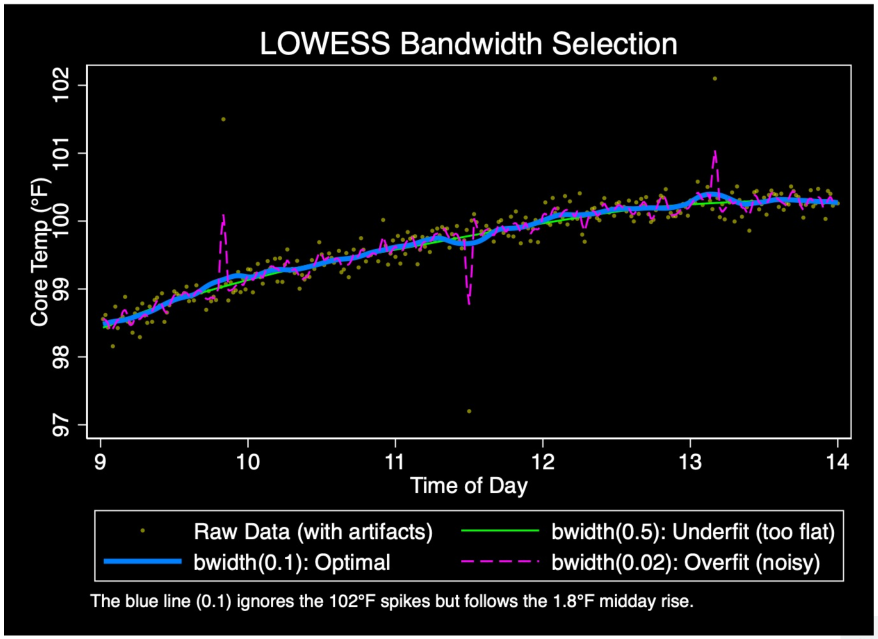

```
*Protocol: CHIPS + island drop + 20% rule
gen miss_tempchips1=missing(cleantemp_chips1)
bysort id trial_week: egen miss_count1 = sum(miss_tempchips1)
bysort id trial_week: gen total_obs1 = _N
gen pct_missing1 = (miss_count1 / total_obs1) * 100
gen cleantemp_chips3=cleantemp_chips1
replace cleantemp_chips3=. if pct_missing1 > 20
```


artifact
+island drop
+island drop+20%rule
102 CMB
125 CMB
172 BCM
194 BCM
230 BCM
235 CMB
289 BCM
314 BCM
383 BCM
385 BCM
456 CMB
524 MBC
565 BCM
618 BCM
782 MBC
807 BCM
828 CMB
893 MBC
953 BCM
982 BCM
94 96 98 100102
0 100 200 300
work_minutes

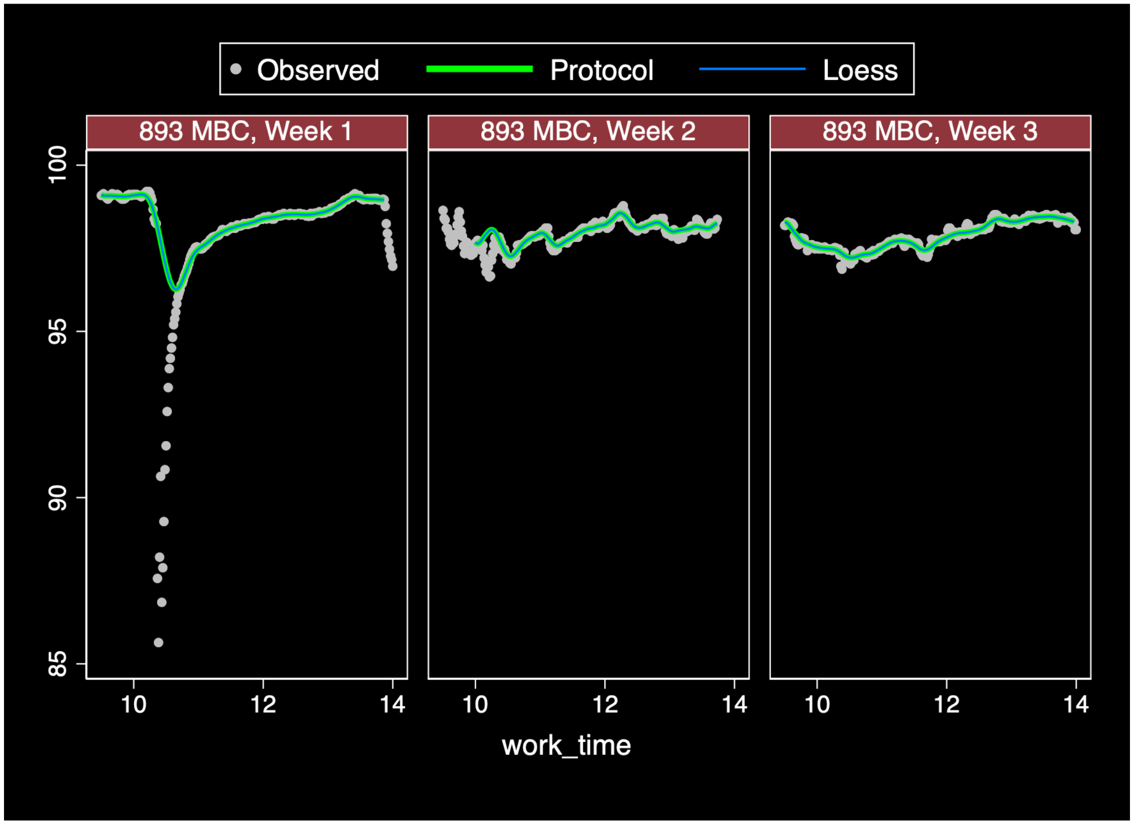
Signal Smoothing

- We apply LOWESS to capture the broad functional form
  1. Removing remaining measurement artifacts
  2. Acts as a "low-pass filter," removing the noise while preserving the meaningful "functional arch" of the worker's shift
  3. Reduce the residual variance, leading to more stable and reliable parameter estimates
  4. Accurate integration for AUC: area better estimated with a continuous and representative trajectory

Signal Smoothing




```
gen cleantemp_chips4_loess =.
levelsof shift_id, local(shifts)
foreach s of local shifts {
    tempvar tmp
    * bwidth(0.1) uses 10% of the day's data for each local regression point
    capture lowess cleantemp_chips4 time_numeric if shift_id == `s', gen(`tmp')
bwidth(0.1) nograph
    if _rc == 0 {
        replace cleantemp_chips4_loess = `tmp' if shift_id == `s'
    }
}
```




```
*** Lowess line
sort id time_numeric
twoway (scatter temperature work_time, msize(small) mcolor(gs12)) ///
            (line cleantemp_chips5 work_time, lwidth(thick)) ///
            (line cleantemp_chips5_loess work_time) ///
            if (id==893), ///
            legend(rows(1) order(1 "Observed" 2 "Protocol" 3 "Loess")) ///
            by(id trial_week, rows(1) note(" ") legend(position(12)))
```

What was the weather like during those 3 weeks?

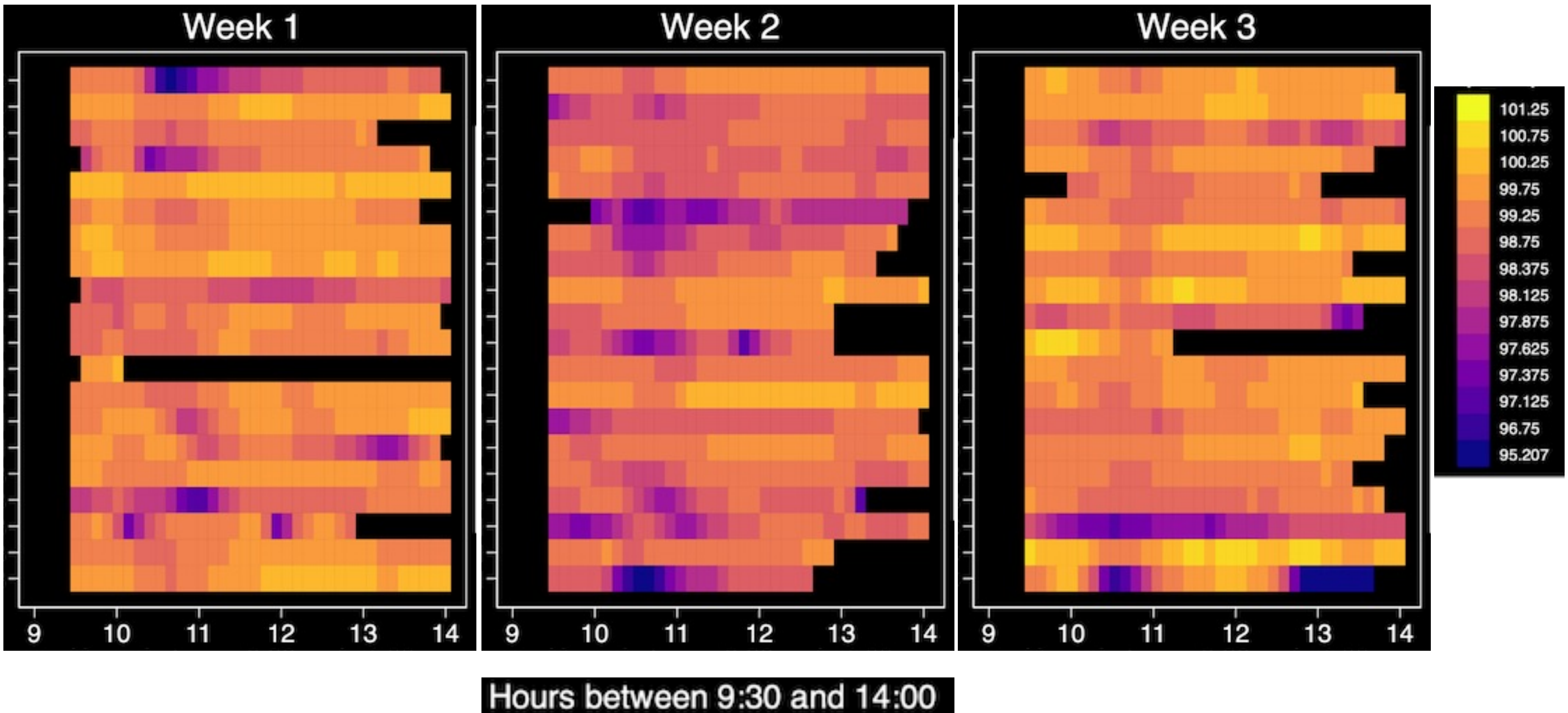
Observed Max Temps

Week 1: 77.4 - 87.7 F
25.2 – 30.9 C

Week 2: 73 – 81 F
22.3 – 27.2 C

Week 3: 73.2 – 78.2 F
22.9 – 25.7 C

Santa Rosa averaged 79.9F highs compared to 87.9F in 2024
26.6C 31.1C



```
use perminutedt_t5, clear
forvalues w=1(1)3 {
	preserve
		keep if trial_week == `w'

		* Create a string version of the ID labels
		decode id, gen(id_string)

		* Mask the sorting variable with the string names
		* ssc install labutil (if not already installed)
		labmask id_arm, values(id_string)

		* Run your Heatmap code
		heatplot cleantemp_chips5_loess i.id_arm work_time, ///
			cut(96.5 97 97.25 97.5 97.75 98 98.25 98.5 99 99.5 ///
			100 100.5 101 101.5) color(plasma) ///
			title("Week `w'") ///
			ytitle("Individual ID") xtitle("Hours between 9:30 and 14:00")
		graph export heatmap_temp_week`w'.png, replace

		**** get mean temperatures by arm at 9:45 (before lunch)
			and 12:45 (before 2nd break)
		keep if work_time==9.75 | work_time==12.75
		bys trial_arm: tabstat cleantemp_chips5_loess, by(work_time) ///
			statistics(n mean sd min max) columns(statistics)

	restore
}
```


Stanford MEDICINE
School of Medicine
Biostatistics Core
Division of Nephrology

Analysis: Are the Cal OSHA standards enough?

Pilot Study: 3-arm Crossover Trial

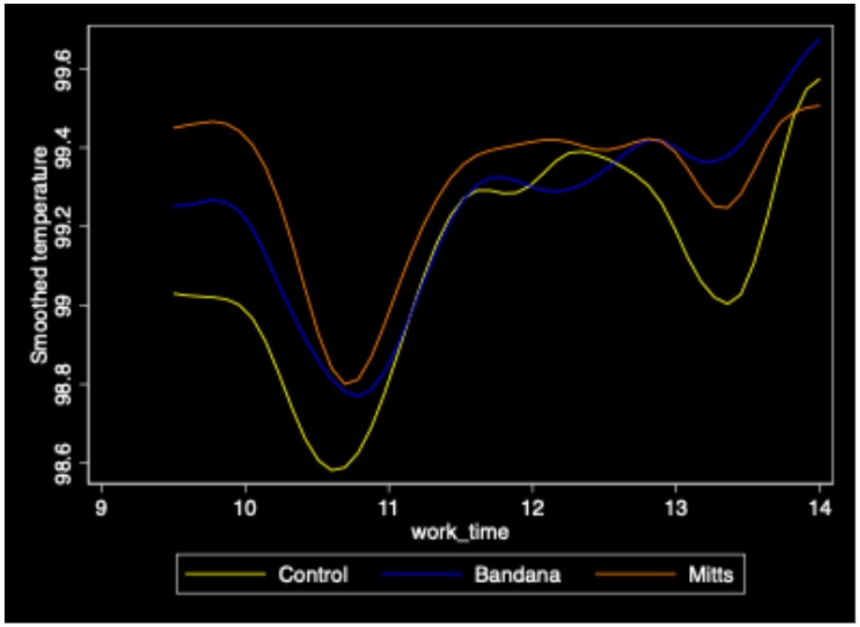
	Week 1	Week 2	Week 3
Group 1	Control	Mitts	Bandana
Group 2	Mitts	Bandana	Control
Group 3	Bandana	Control	Mitts

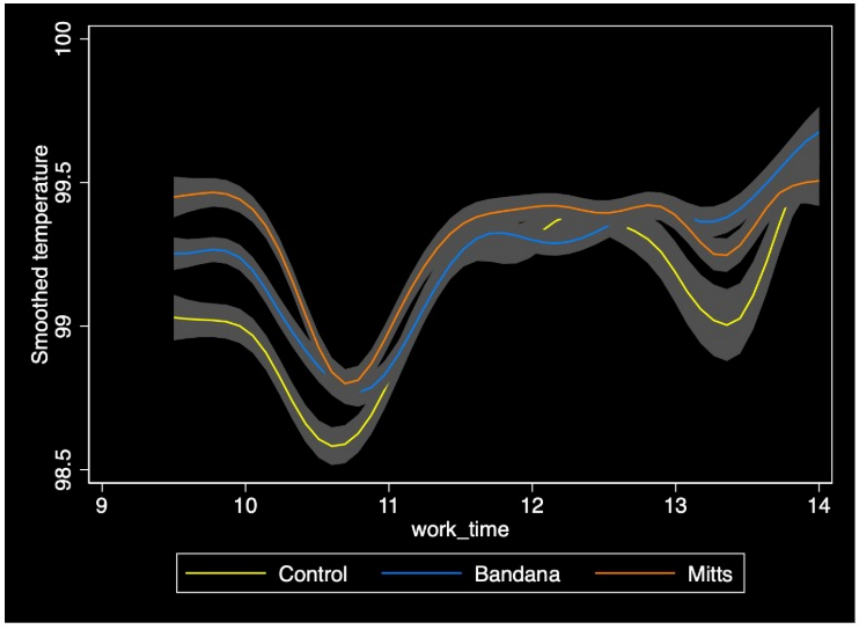
Control: Cal OSHA standards

Bandana

Mitts

Kdensity weighted local polynomial estimates

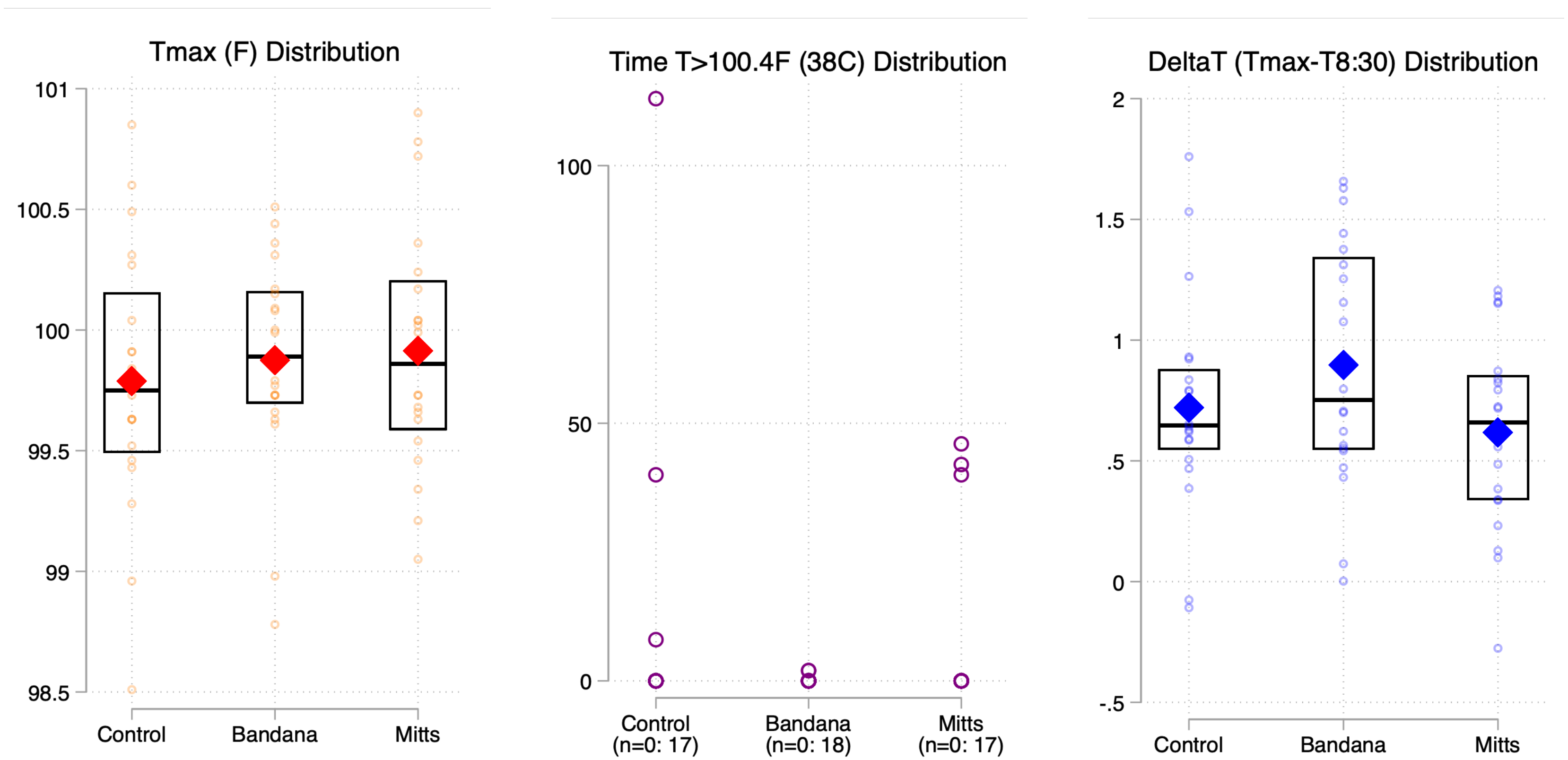





```
* Grand Mean per arm
twoway (lpoly cleantemp_chips5_loess work_time if trial_arm == 0, lcolor(yellow)) ///
       (lpoly cleantemp_chips5_loess work_time if trial_arm == 1, lcolor(blue)) ///
       (lpoly cleantemp_chips5_loess work_time if trial_arm == 2, lcolor(orange)), ///
       legend(rows(1) order(1 "Control" 2 "Bandana" 3 "Mitts")) ///
       xtitle("work_time") ytitle("Smoothed temperature")
```

Limitations of "Snapshot" Metrics

- Maximum Core Temperature (T_max)
  - Measure of acute safety
- Time Above Threshold
  - Duration: Minutes where Temp > 100.4F
- Delta Temperature
  - Net change from morning baseline to day’s peak (T_max - T_start)




```
*ssc install stripplot
stripplot tmax, box over(trial_arm) vertical fxsize(60) ///
    msymbol(oh) mcolor(orange%30) /// Individual data points (optional)
    addplot(scatter arm_tmaxmean trial_arm , msymbol(D) mcolor(red) msize(large)) ///
    title("Tmax (F) Distribution") ytitle("") xtitle("")
```

Limitations of "Snapshot" Metrics

- Maximum Core Temperature (T_max)
  - High intensity
  - but ignores how long it lasted

- Time Above Threshold
  - Captures duration
  - but ignores how high the temperature went

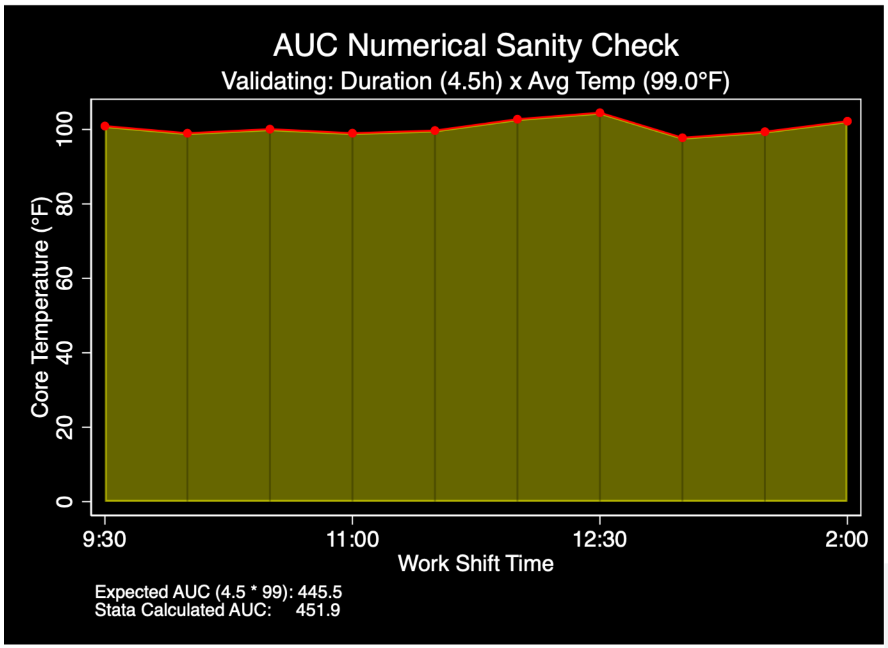
- Delta Temperature
  - Captures individual change
  - but ignores the accumulated temperature effect throughout the shift

Quantifying Thermal Load: Area Under the Curve (AUC)

- Measure of cumulative "heat-strain"
  - Captures both the **intensity AND duration** of heat exposure rather than a single peak value

How to calculate AUC?

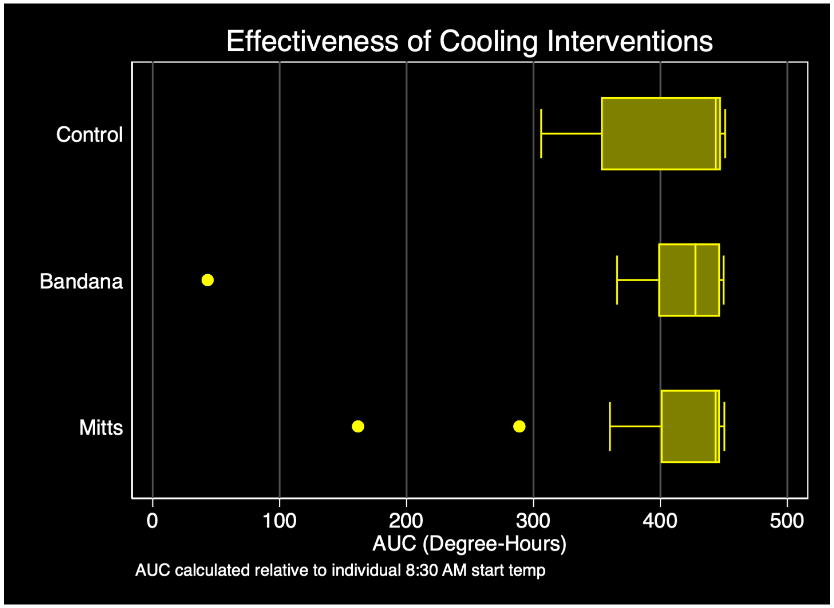
The Trapezoidal Rule: area = height (Temperature) x width (Time interval)




```
* 3. Calculate AUC using the trapezoidal rule
integ temp time, trapezoid
local actual_auc : display %9.1f r(integral)
local expected_auc = 4.5 * 99.0
```

Results: AUC basic statistics

trial_arm	N	Mean	SD	Min	p25	p50	p75	Max
Control	20	409.7429	53.75247	306.0421	353.1722	443.5562	447.5872	450.9585
Bandana	20	403.8233	88.78955	43.2831	398.2985	427.5407	446.9536	449.7645
Mitts	20	410.978	71.39056	161.7578	400.2701	443.4323	446.91	450.36
Total	60	408.1814	71.55725	43.2831	393.507	440.5646	447.3425	450.9585




```
* Basic Stats by Trial Arm
tabstat auc_val, by(trial_arm) ///
    statistics(n mean sd min p25 p50 p75 max) ///
    columns(statistics)

* Basic Box Plot by Trial Arm
graph hbox auc_val, over(trial_arm) ///
    ytitle("AUC (Degree-Hours)") ///
    title("Effectiveness of Cooling Interventions")
graph export boxplot_auc_perarm.png, replace
```


Stanford MEDICINE
School of Medicine
Biostatistics Core
Division of Nephrology

Effect of interventions on mean AUC

- Mixed effects model
  - To account for the nested structure
    - Observation *i* nested within worker *j*
  - Random intercept: allows each worker to have their own baseline

```
* Model 1: Arm only
mixed auc_val i.trial_arm || id: , reml
margins r.trial_arm, post
estimates store m1
```

```
Mixed-effects REML regression                   Number of obs     =         60
Group variable: id                              Number of groups  =         20
                                                Obs per group:
                                                              min =          3
                                                              avg =        3.0
                                                              max =          3
                                                Wald chi2(2)      =       0.11
Log restricted-likelihood = -329.71908          Prob > chi2       =     0.9462
```

auc_val	Coefficient	Std. err.	z	P>|z|	[95% conf. interval]	
trial_arm						
Bandana	-5.919567	22.99965	-0.26	0.797	-50.99806	39.15893
Mitts	1.23506	22.99965	0.05	0.957	-43.84343	46.31355
_cons	409.7429	16.26321	25.19	0.000	377.8676	441.6182

Random-effects parameters	Estimate	Std. err.	[95% conf. interval]	
id: Identity				
var(_cons)	2.09e-11	6.91e-08	0	.
var(Residual)	5289.841	991.046	3664.13	7636.851

LR test vs. linear model: chibar2(01) = 0.00   Prob >= chibar2 = 1.0000

Sensitivity to Model Specifications

Model 1: Arm only

Model 2: + Trial Week

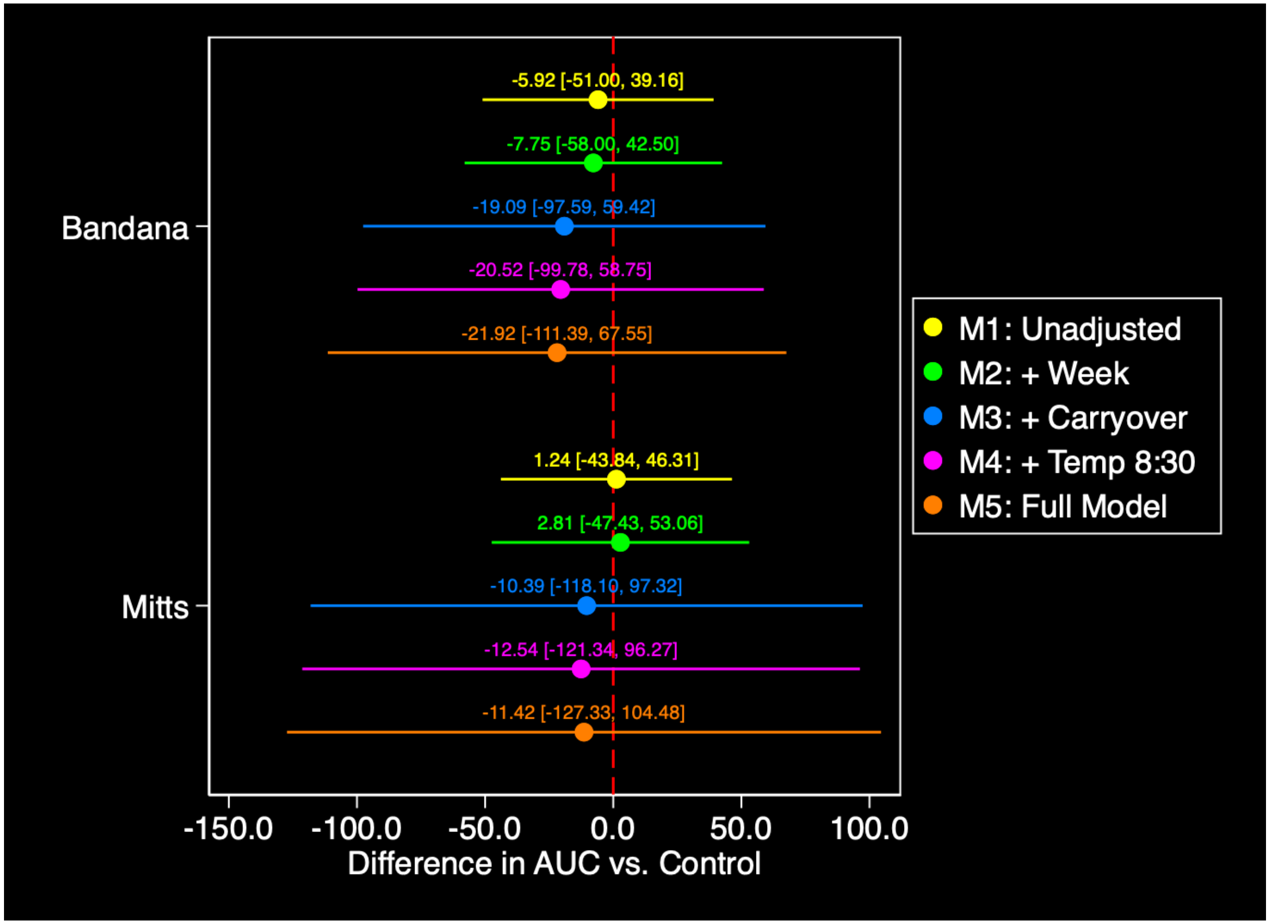
Model 3: + Carryover

Model 4: + Baseline Temp

Model 5: Full Model (Demographics)

```
**** forest plot of AUC analysis
coefplot (m1, label("M1: Unadjusted")) ///
         (m2, label("M2: + Week")) ///
         (m3, label("M3: + Carryover")) ///
         (m4, label("M4: + Temp 8:30")) ///
         (m5, label("M5: Full Model")), ///
    horizontal xline(0, lcolor(red) lp(dash)) ///
    drop(_cons) ///
    format(%9.1f) ///
    mlabel(string(@b, "%9.2f") + " [" + string(@ll, "%9.2f") + ", " + string(@ul,
"%9.2f") + "]") mlabsize(vsmall) mlabposition(12) ///
    xtitle("Difference in AUC vs. Control") ///
    grid(none) legend(position(3) cols(1)) ///
    ylabel(1 "Bandana" 2 "Mitts")
graph export forest_auc.png, replace
```


Stanford MEDICINE
School of Medicine
Biostatistics Core
Division of Nephrology



-5.92 [-51.00, 39.16]
-7.75 [-58.00, 42.50]
Bandana
-19.09 [-97.59, 59.42]
-20.52 [-99.78, 58.75]
-21.92 [-111.39, 67.55]
1.24 [-43.84, 46.31]
2.81 [-47.43, 53.06]
Mitts
-10.39 [-118.10, 97.32]
-12.54 [-121.34, 96.27]
-11.42 [-127.33, 104.48]
-150.0
-100.0
-50.0
0.0
50.0
100.0
Difference in AUC vs. Control
M1: Unadjusted
M2: + Week
M3: + Carryover
M4: + Temp 8:30
M5: Full Model

Summary

- Code for signal Processing to handle temperature dips, islands, missingness and data smoothing
  - Used *fillin* and *mipolate* to ensure our time-series was continuous and ready for analysis
- Use of *lowess* for signal smoothing
- AUC to capture the cumulative "thermal debt"
- Use of stripplot and coefplot to enhance visualization
- Leveraged the Stata community (SSC) who graciously provide user-written packages like heatplot and labmask

Command/Function	Source/Package	Description
fillin	Built-in	Rectangularizes the dataset by creating observations for missing worker-shift-time combinations.
tsset / xtset	Built-in	Declares the data to be time-series or panel data, enabling lead/lag operators and longitudinal analysis.
mipolate	ssc install mipolate	Performs linear interpolation to fill gaps in telemetric pill data before smoothing.
lowess	Built-in	Applies Locally Weighted Scatterplot Smoothing to isolate physiological trends from sensor noise.
labmask	ssc install labutil	Assigns variable values (like worker IDs or shift dates) as labels to another variable for cleaner axis formatting.
heatplot	ssc install heatplot	Creates "Lexis" style grids or spatio-temporal maps of temperature intensity over the course of a shift.
stripplot	ssc install stripplot	Enhanced alternative to dotplot and boxplot; used here to overlay means and jittered raw data.
integ	Built-in	Uses the trapezoidal rule to calculate the Area Under the Curve (AUC) for cumulative heat strain.
mixed / margins	Built-in	Fits multi-level models to account for nested data; margins calculates predicted thermal loads per arm.
coefplot	ssc install coefplot	Visualizes model coefficients and confidence intervals, making the mixed-effects results easy to interpret.

Thank you!

MMRATH@STANFORD.EDU