Parallel Session

Operational Evidence of Fatigue: Shiftwork Operations

Fatigue in Aviation Inspections: Sleep and Performance Data

Colin G. Drury, Ph.D.
University at Buffalo: SUNY

15:10 - 15:35

June 17, 2008

Presented at the FAA Fatigue Management Symposium: Partnerships for Solutions; Vienna, VA: June 17-19, 2008

Colin G. Drury, Ph.D.
Biography

Colin G. Drury is Distinguished Professor Emeritus and Director of Research Institute for Safety and Security in Transportation (RISST) at University at Buffalo, where his work is concentrated on the application of human factors techniques to inspection and maintenance processes. Since 1989 he has been leading a team applying human factors techniques to reduce errors in aviation maintenance and inspection at RISST. He has over 200 publications on topics in industrial process control, quality control, aviation maintenance, security and safety. He is a Fellow of the Institute of Industrial Engineers, the Ergonomics Society, the International Ergonomics Association and the Human Factors & Ergonomics Society, receiving the Bartlett medal of the Ergonomics Society and both the Fitts and Lauer Awards of the Human Factors Ergonomics Society. He recently received that FAA's Excellence in Aviation Research award and the American Association of Engineering Societies' Row Award. He has a private pilot's license.

Fatigue in aviation inspection: sleep and performance data

Colin G. Drury

Industrial and Systems Engineering
University at Buffalo: SUNY

Aviation Fatigue Management Symposium 2008

Presented at the FAA Fatigue Management Symposium: Partnerships for Solutions; Vienna, VA: June 17-19, 2008

Aviation Inspection Failure

- Several accidents involving Inspection failure
  - The 1998 Aloha incident was due to multiple site fatigue damage leading to structural failure.
  - The 1989 Sioux City crash was the result of inspection not finding a crack in an engine disk.
  - The 1996 Pensacola incident was due to a fan hub in the left engine having an undetected crack.
- All had “normal” inspection work duration
- Two questions:
  - Are aircraft inspection tasks fatiguing?
  - Are aircraft inspection tasks sensitive to fatigue?

Temporal Factors in Inspection

- Visual and Non-Destructive Inspection (NDI) techniques require the inspector to work continuously on repetitive tasks for extended periods.
- Such tasks typically occur on all shifts and can involve inspecting at low periods of the human circadian rhythm.
- Therefore inspection tasks have some fatigue potential

Therefore Vigilance?

- Inspection tasks look *a priori* to be analogs of laboratory vigilance tasks. But are they?
- Two Issues:
  - Can we expect the findings from the vigilance literature to apply to aircraft inspection?
  - How well might the control methods for circadian rhythms and cumulative fatigue from shift working apply to aircraft inspection?

Applying to Aircraft Inspection

- Do aircraft inspectors work long hours, have continuous work etc?
- Used survey modified from Folkard's (2002) study of aircraft mechanics
- Gave to 40 NDI inspectors at two US sites
- Results (over) showed typical work period of 2 hours, with ~ 10 min break. Vigilance may apply. Shift work applies.
- Sample older and more senior than BLS data on mechanics (1991) but expected as NDI inspectors recruited from more experienced mechanics (AMT)

	Median	Minimum	Maximum
Hours of work per week	40	30	56
How many hours before a work break?	2.0	1.0	4.0
How many minutes does break last?	12.5	0	45
How many days annual leave?	31	11	40

Describing Inspection

- Systematic analysis of the inspection techniques involved in NDI of aircraft (Drury, 2003): Fluorescent Penetrant Inspection (FPI), Visual inspection, Borescopes, Eddy Current, X-ray, Magnetic Particle and Ultrasonics.
- Each of these NDI techniques exhibited a set of generic functions, seen earlier by Sinclair, Drury, Wang etc.
- Compare these with typical vigilance task

Inspection Description

Function	
Initiate	All processes up to accessing the component. Get and read work card. Assemble and calibrate required equipment. For FPI this includes part preparation steps.
Access	Locate and access inspection area. Be able to see the area to be inspected at a close enough level to ensure reliable detection. For component inspection, the parts are typically brought to the inspector rather than the inspector going to the airframe.
Search	Move field of view across component to ensure adequate coverage. Carefully scan field of view using a good strategy. Stop search if an indication is found.
Decision	Identify indication type. Compare indication to standards for that indication type.
Response	If indication confirmed, then record location and details. Complete paperwork procedures. Remove equipment and other job aids from work area and return to storage. If indication not confirmed, continue search (3).

VIGILANCE TASK ATTRIBUTE	INSPECTION TASK ATTRIBUTE
Time Uncertainty	Defect occurrence is rarely predictable although inspectors often return to the same area of the same aircraft or engine and attempt to predict when defects are likely.
Spatial Uncertainty	Actual occurrence of defects at specific places on specific components is unpredictable, but the inspector has much useful information to guide the inspection process. Training, and shared experiences can help point inspectors to specific locations where defects are more likely.
Low Feedback	Aircraft inspectors do not get good feedback: no easy way to find what truly is a signal, especially a missed signal. Feedback on missed defects only when found at a subsequent inspection, or an operational incident occurs. Even feedback on false alarms is sporadic. Feedback of both Misses and False Alarms is at best severely delayed and of little use to the inspector.
Unrealistic Expectations	For more common defects, expectations from training can translate relatively faithfully into practice. However, for very rare defects, expectation may still be unrealistically high after considerable practice.
Isolated Environment	The hangar and even the shop inspection environment are typically noisy, social and distracting. Both noise and social interaction and even some forms of distraction have been found to <u>improve</u> vigilance performance in laboratory tasks.

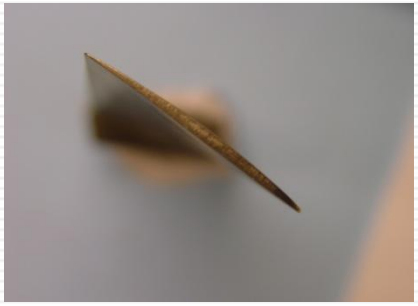
VIGILANCE TASK ATTRIBUTE	INSPECTION TASK ATTRIBUTE
Important Signals	Cracks or other defects that can have direct safety consequences.
Rare Signals	Defects can range from quite common, e.g. corrosion areas on older aircraft, to extremely rare (e.g. cracks in jet engine titanium hubs). Less than 1 out of 10 inspected components will contain a reportable defect.
Low Signal Strength	Most defects are perceptually difficult to detect, often occurring within a background of non-defects, e.g. cracks among dirt marks and scratches.
Long Time on Task	Time on task can vary from a few minutes to about 2 hours without a break. Scheduled breaks are typically four 15-min breaks per shift, but many tasks are self-paced so that inspectors can break early or continue beyond scheduled time to complete an area or component.
High Memory Load	Prototypical defects are usually stored in the inspector's memory, rather than being presented as part of the task. Sometimes typical defects are illustrated on workcards, but workcards are often poorly integrated into the inspection task.
Low Observer Practice	Inspectors are highly skilled and practiced, after 3-10 years as an AMT before becoming an inspector. However, for some rare defects, even experienced inspectors may literally never have seen one in their working lifetime.
Sustained Attention on One Task	Inspectors may have some tasks where just one defect type is the target, but these are often interspersed with other tasks (e.g. different components) where different defects, often less rare defects, are the target.

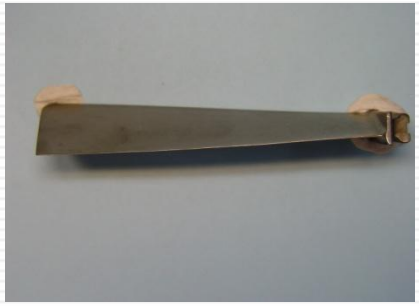
What do we Know?

- Still not clear how closely vigilance mimics aviation inspection tasks
- Quite clear that vigilance tasks are particularly sensitive to the effects of circadian lows and cumulative fatigue from shift working.
- Horowitz, Cade, Wolfe, and Cziesler (2003) found usual effect of peaks and troughs of circadian rhythm on a vigilance task, but none on a simple search tasks performed at similar times.

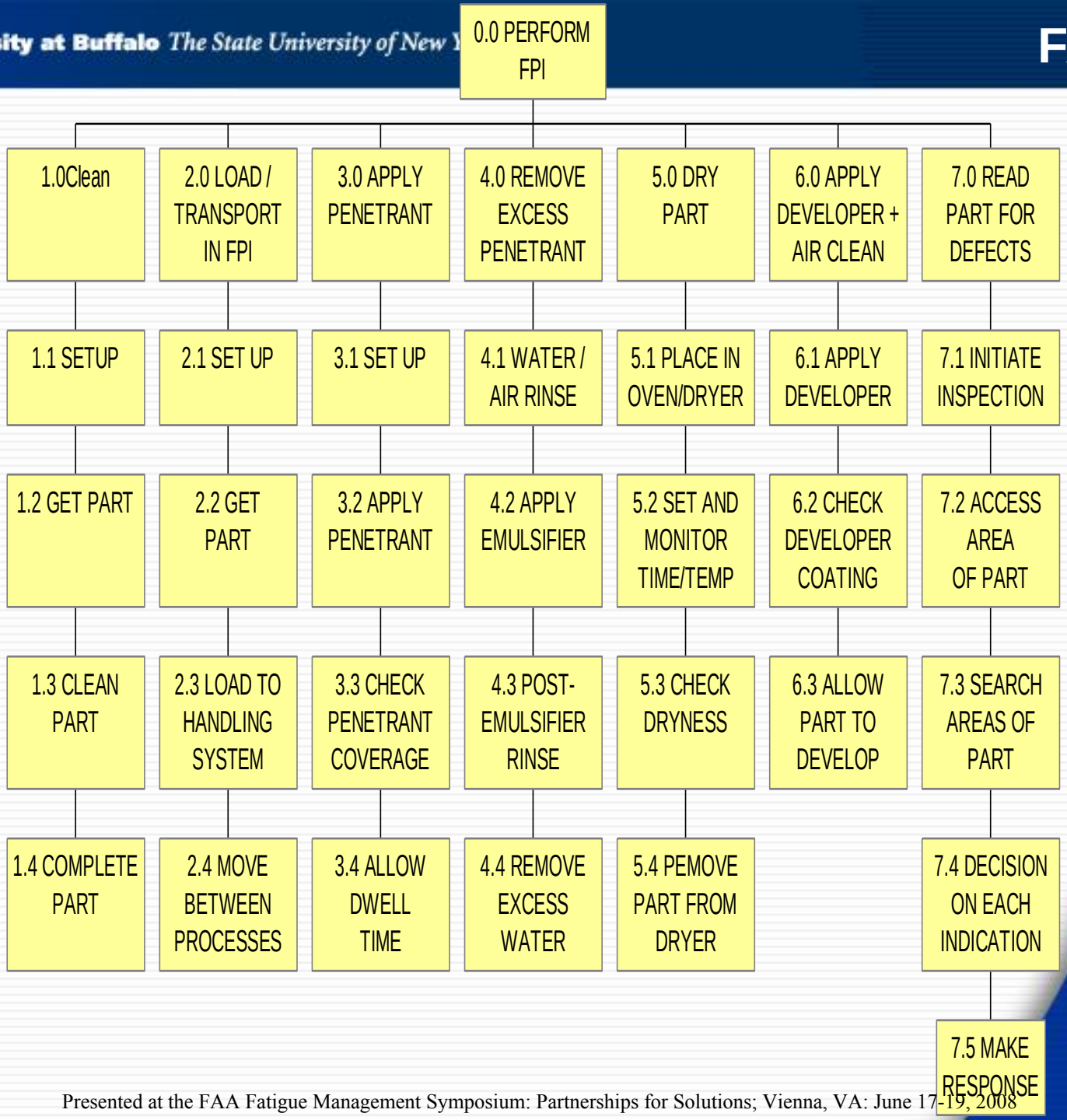
Experiments on NDI Fatigue

- Chose one type of NDI that had failed in Sioux City and Pensacola accidents
- Fluorescent Penetrant Inspection (FPI) used for many years. Involves dipping item into fluorescent liquid, cleaning off excess and examining item for cracks under fluorescent illumination
- Developed a computer simulation based on photos of 64 blades from 6 angles (over)
- Validated program using airline inspectors





FAA 2008


0.0 PERFORM FPI
1.0Clean
2.0 LOAD / TRANSPORT IN FPI
3.0 APPLY PENETRANT
4.0 REMOVE EXCESS PENETRANT
5.0 DRY PART
6.0 APPLY DEVELOPER + AIR CLEAN
7.0 READ PART FOR DEFECTS
1.1 SETUP
2.1 SET UP
3.1 SET UP
4.1 WATER / AIR RINSE
5.1 PLACE IN OVEN/DRYER
6.1 APPLY DEVELOPER
7.1 INITIATE INSPECTION
1.2 GET PART
2.2 GET PART
3.2 APPLY PENETRANT
4.2 APPLY EMULSIFIER
5.2 SET AND MONITOR TIME/TEMP
6.2 CHECK DEVELOPER COATING
7.2 ACCESS AREA OF PART
1.3 CLEAN PART
2.3 LOAD TO HANDLING SYSTEM
3.3 CHECK PENETRANT COVERAGE
4.3 POST-EMULSIFIER RINSE
5.3 CHECK DRYNESS
6.3 ALLOW PART TO DEVELOP
7.3 SEARCH AREAS OF PART
1.4 COMPLETE PART
2.4 MOVE BETWEEN PROCESSES
3.4 ALLOW DWELL TIME
4.4 REMOVE EXCESS WATER
5.4 PEMOVE PART FROM DRYER
7.4 DECISION ON EACH INDICATION
7.5 MAKE RESPONSE

FAA 2008

FAA 2008

FAA 2008

Two Studies, as Designed

- Idea was classic Design of Experiments sequence: 2^n factorial design to find factor structure, followed by series of parametric experiments to get response surface
- Used industrially - experienced population recruited via advert in local paper: should be good analogue of inspectors (see previous SUNY studies)

Two Studies, with Plan B

- Idea was classic Design of Experiments sequence: 2^n factorial design to find factor structure, followed by series of parametric experiments to get response surface
- Used industrial population recruited via advert in local paper: should be good analogue of inspectors (see previous SUNY studies)
- Did not work out quite that way!
- Experiment 1 started as 2^5 factorial design but lost one factor
- Experiment 2 re-designed as validation using aircraft NDI inspectors

Experiment 1: Hypotheses and Design

- Is vigilance a valid model of complex inspection tasks?
  - Tested Time on Task (1 vs. 2 hours), Shift (0300 vs. 0900 start) and Defect Rate (7% vs. 15%) all known effects in vigilance literature
  - Also used Time Period (20 min) intervals
- What interventions prevent fatigue effects?
  - Tested Breaks (3 min break every 20 min vs. no breaks) and Illumination (50 vs. 500 Lux)

Design of Experiment 1

- Between participants except for Shift (each participant tested 0300 and 0900 and at each 20-min time period
- Recruited 80 participants from the local community
- Good age range and all had working experience
- Trained ~ 3 hours on simulation using different cracks

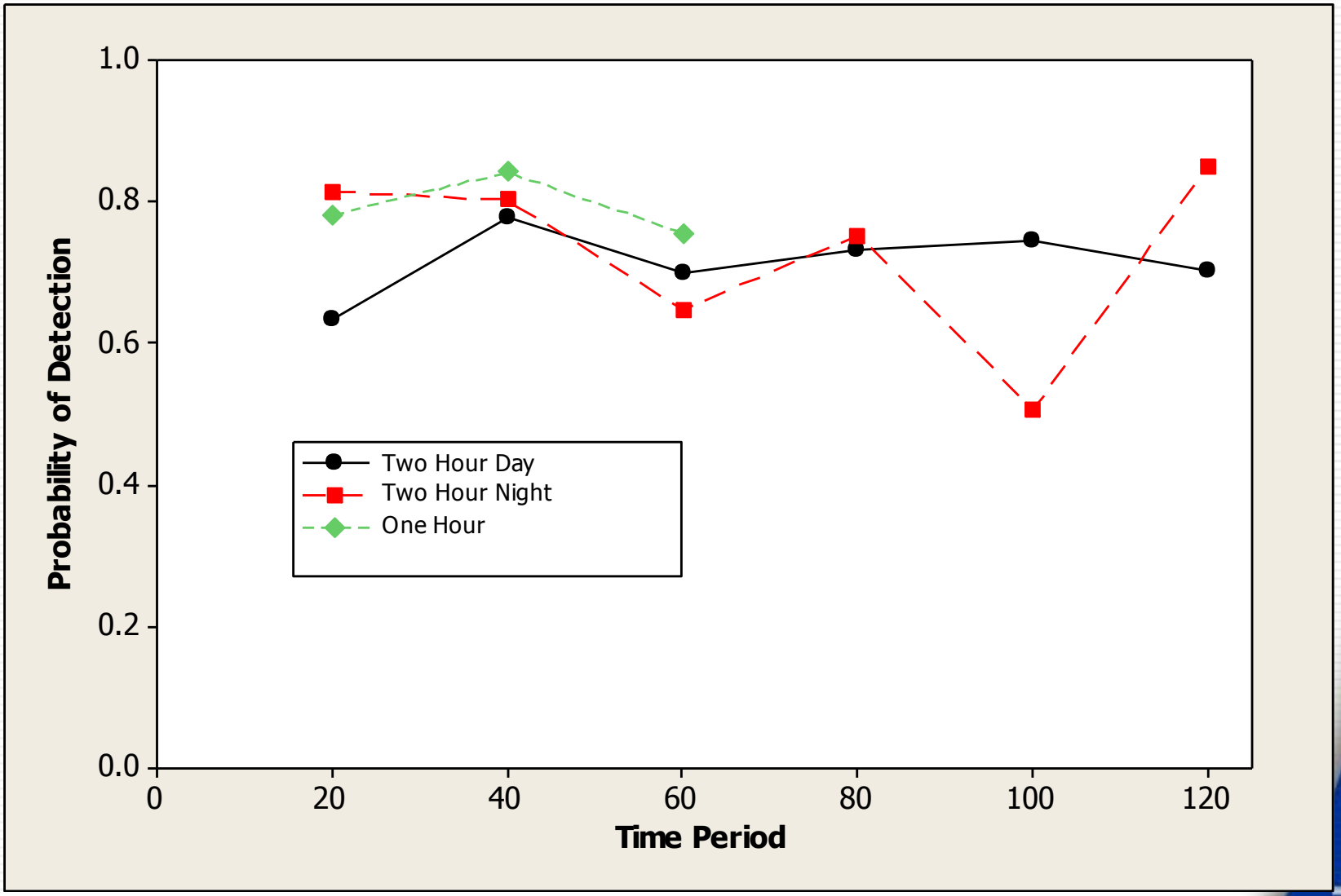
Measures

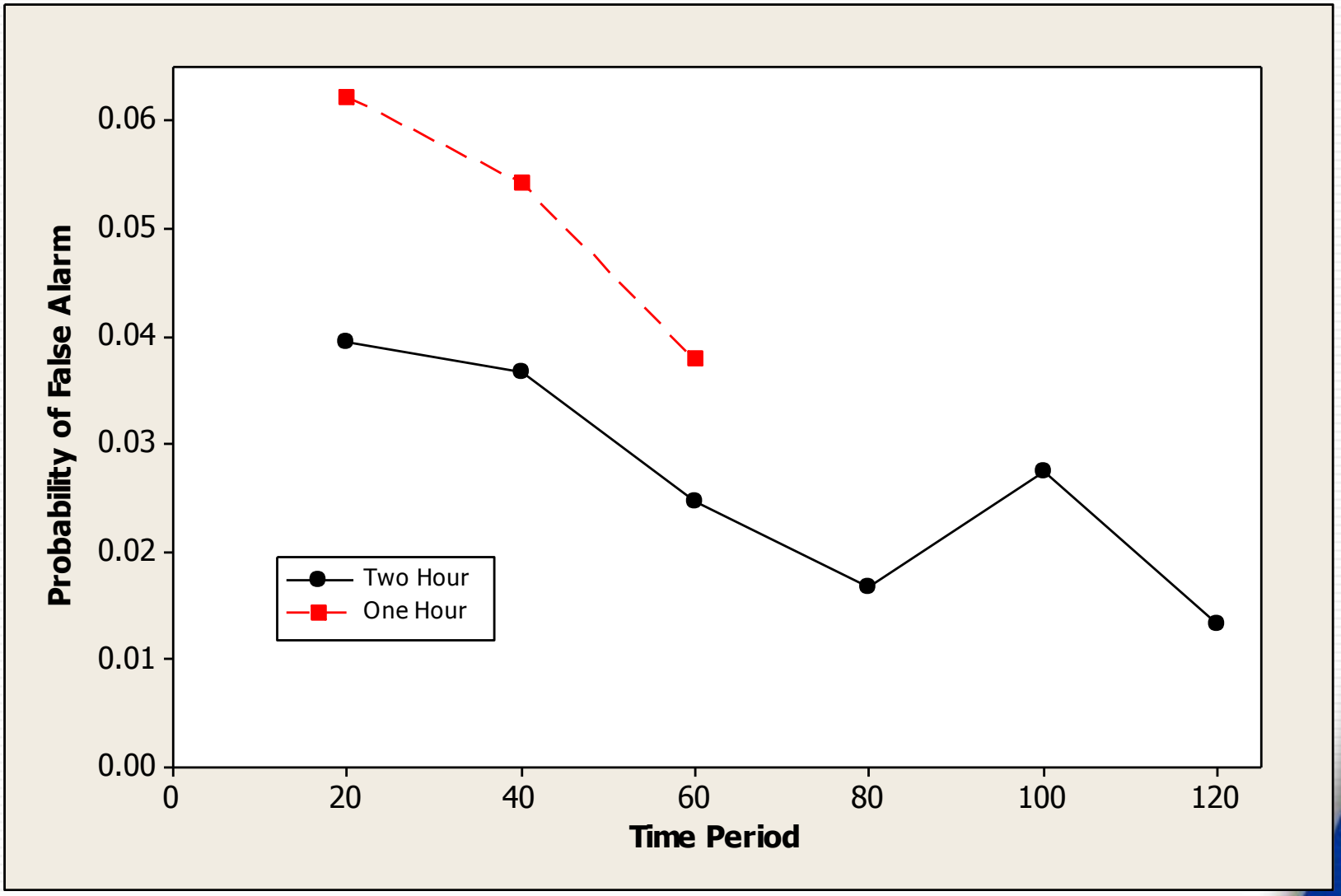
- Performance: Accuracy and speed
- Hits, misses, FAs, correct rejections, d'
- Times for each decision, plus total blades
- Collected sleep results for prior 24 hours with ActiWatch ®: Hours since sleep, sleep length
- Scale responses for sleepiness / fatigue scales: SOFI and SSS before and after task
- Workload with TLX
- (Task Used)

Experiment 1: Time on Task Results

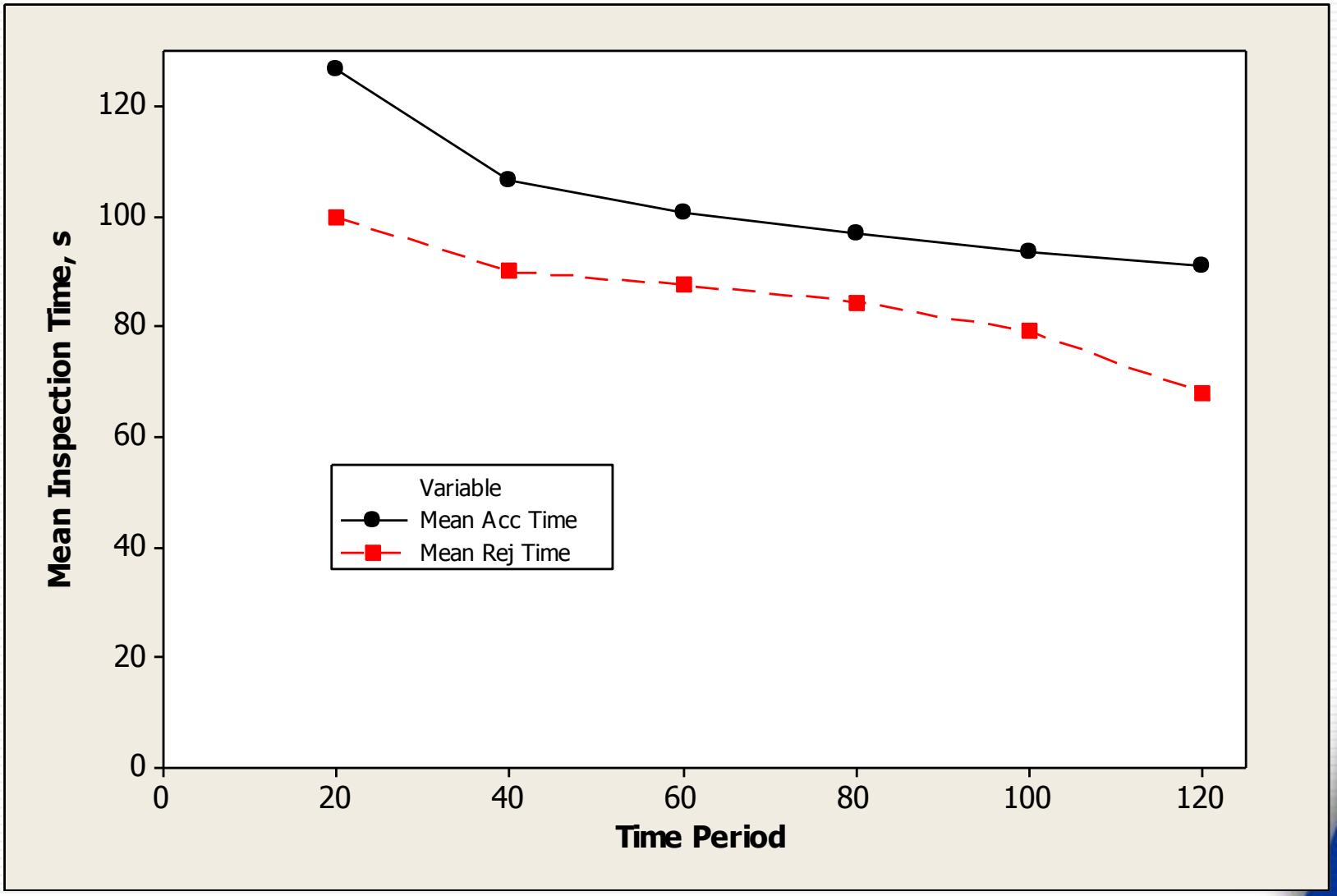
- No main effect of Time Period on PoD but interaction with Shift for 2-hour task
- Good improvement on False Alarms and Speed over Time Period
- Other interesting effects, including that performance, sleep and reported fatigue / sleepiness did not correlate
- Also Shifts X Breaks interaction: faster performance with breaks in day, not night

FAA 2008

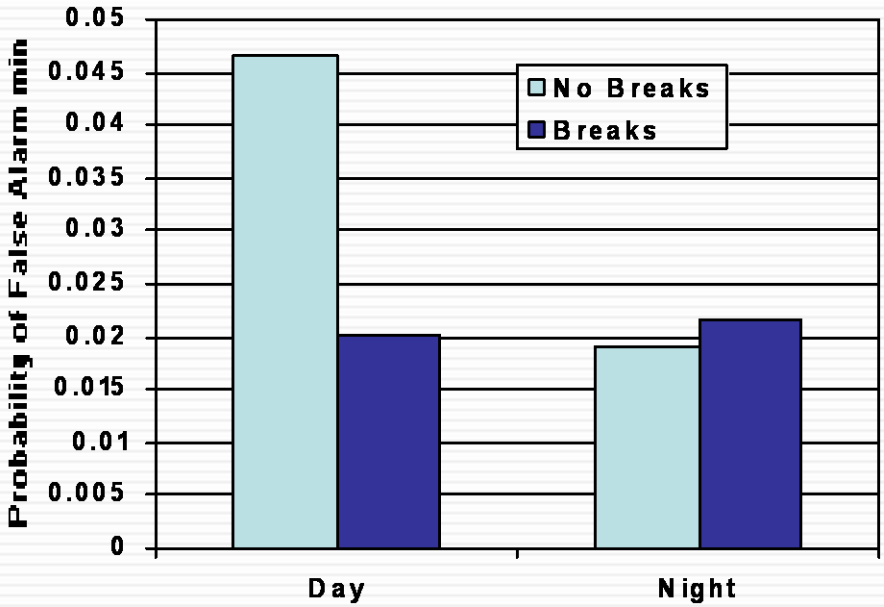

1.0
0.8
0.6
0.4
0.2
0.0
Probability of Detection
0
20
40
60
80
100
120
Time Period
Two Hour Day
Two Hour Night
One Hour

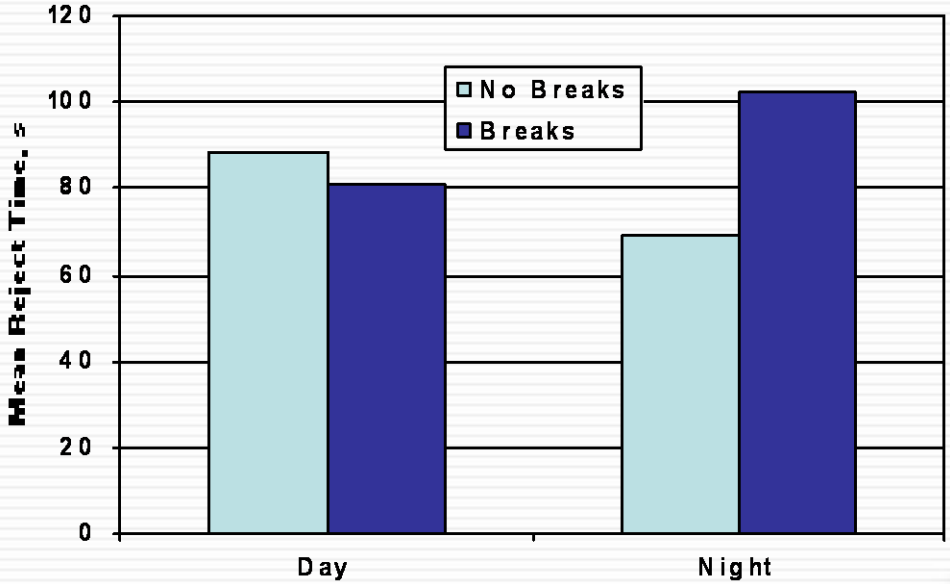

Probability of False Alarm
0.06
0.05
0.04
0.03
0.02
0.01
0.00
0
20
40
60
80
100
120
Time Period
Two Hour
One Hour

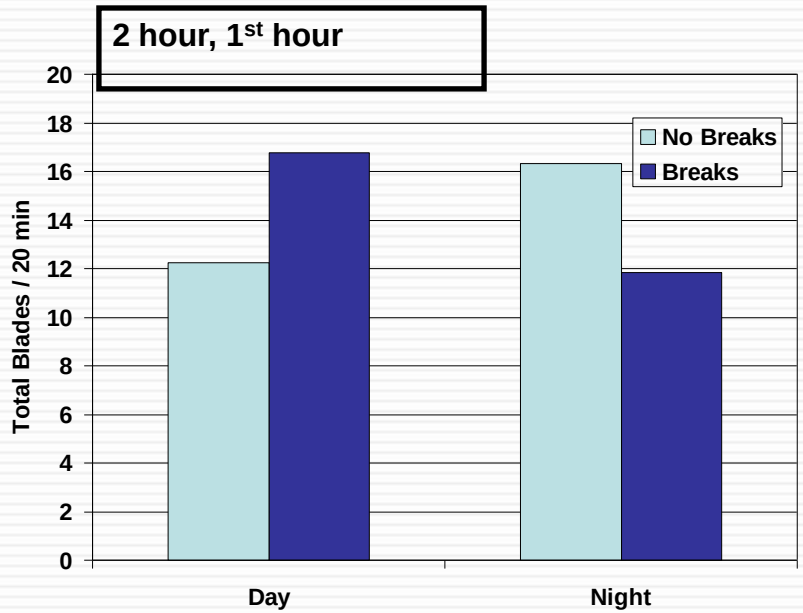
FAA 2008

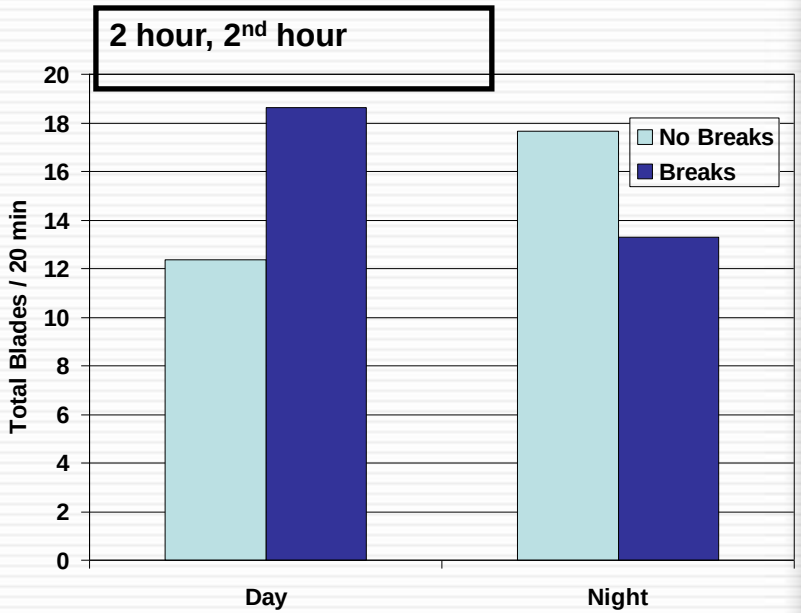

Mean Inspection Time, s
120
100
80
60
40
20
0
0
20
40
60
80
100
120
Time Period
Variable
Mean Acc Time
Mean Rej Time

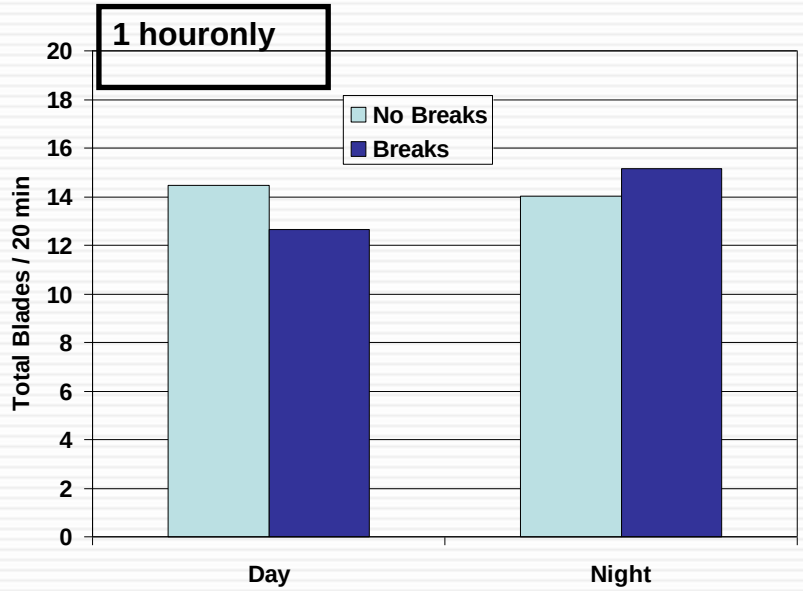
Breaks X Shift



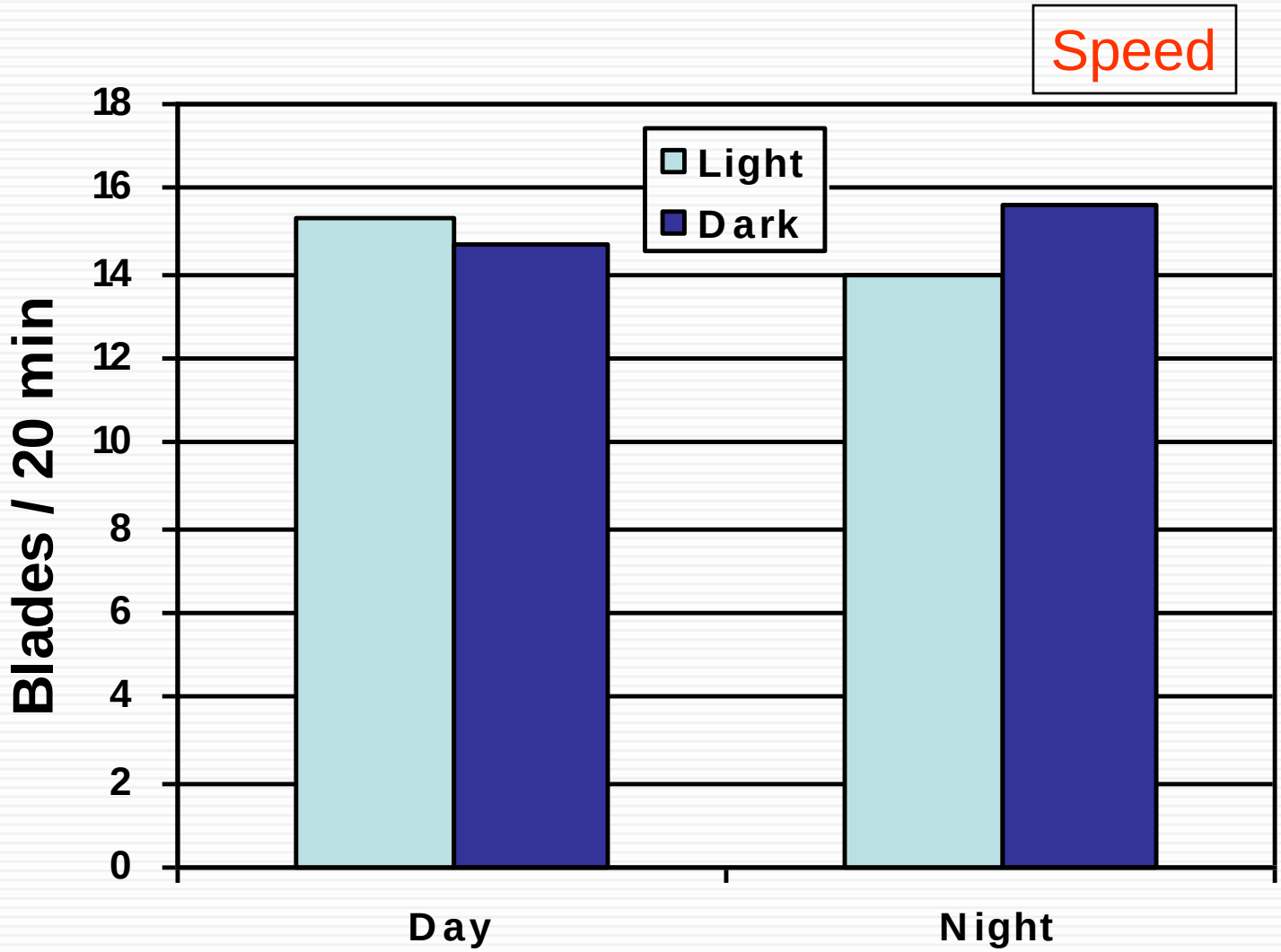



2 hour, 1st hour
Total Blades / 20 min
0
2
4
6
8
10
12
14
16
18
20
No Breaks
Breaks
Day
Night

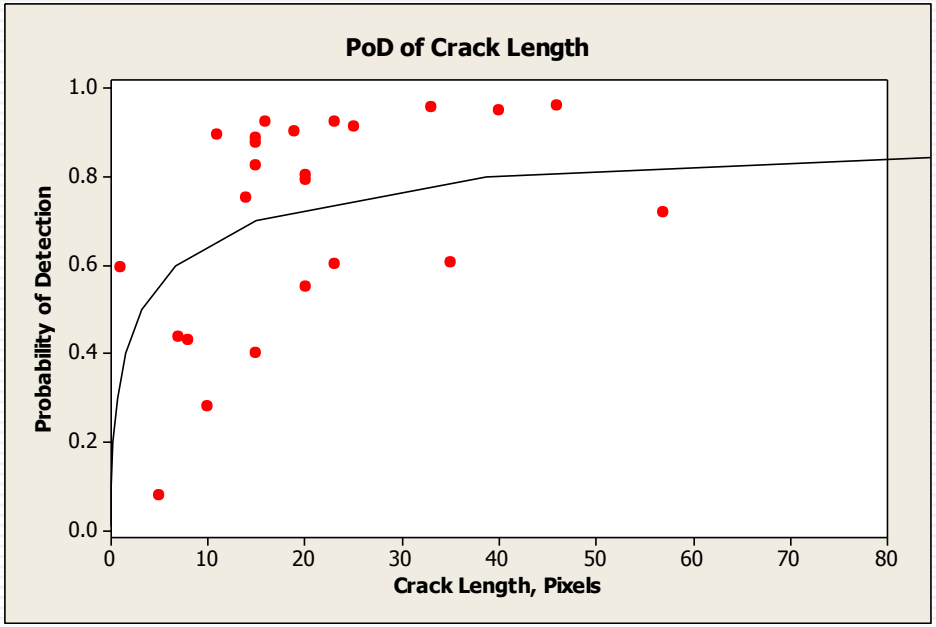

2 hour, 2nd hour
Total Blades / 20 min
0
2
4
6
8
10
12
14
16
18
20
No Breaks
Breaks
Day
Night


1 houronly
Total Blades / 20 min
0
2
4
6
8
10
12
14
16
18
20
No Breaks
Breaks
Day
Night

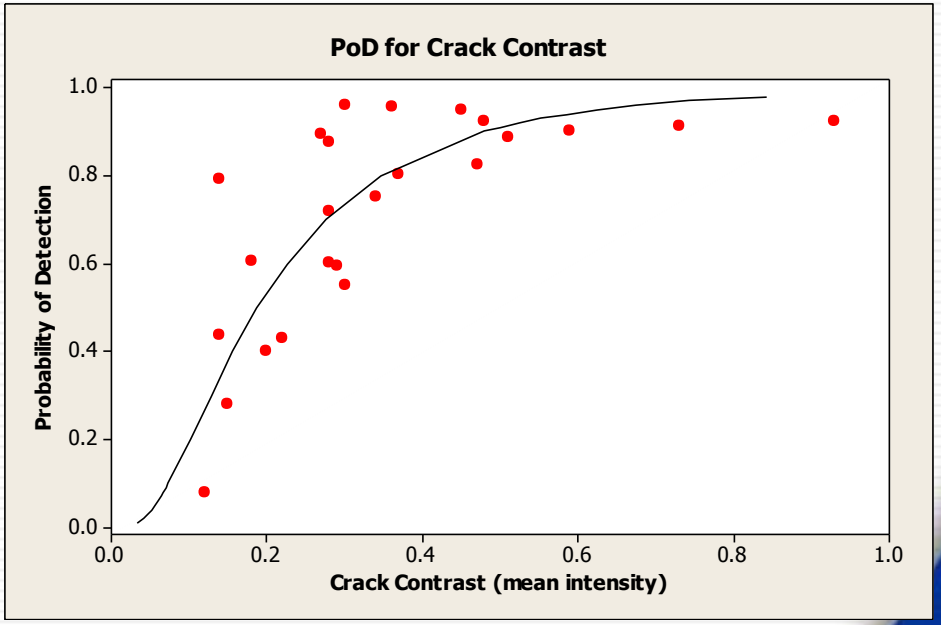
Speed



Illumination X Shift


PoD of Crack Length
Probability of Detection
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
50
60
70
80
Crack Length, Pixels

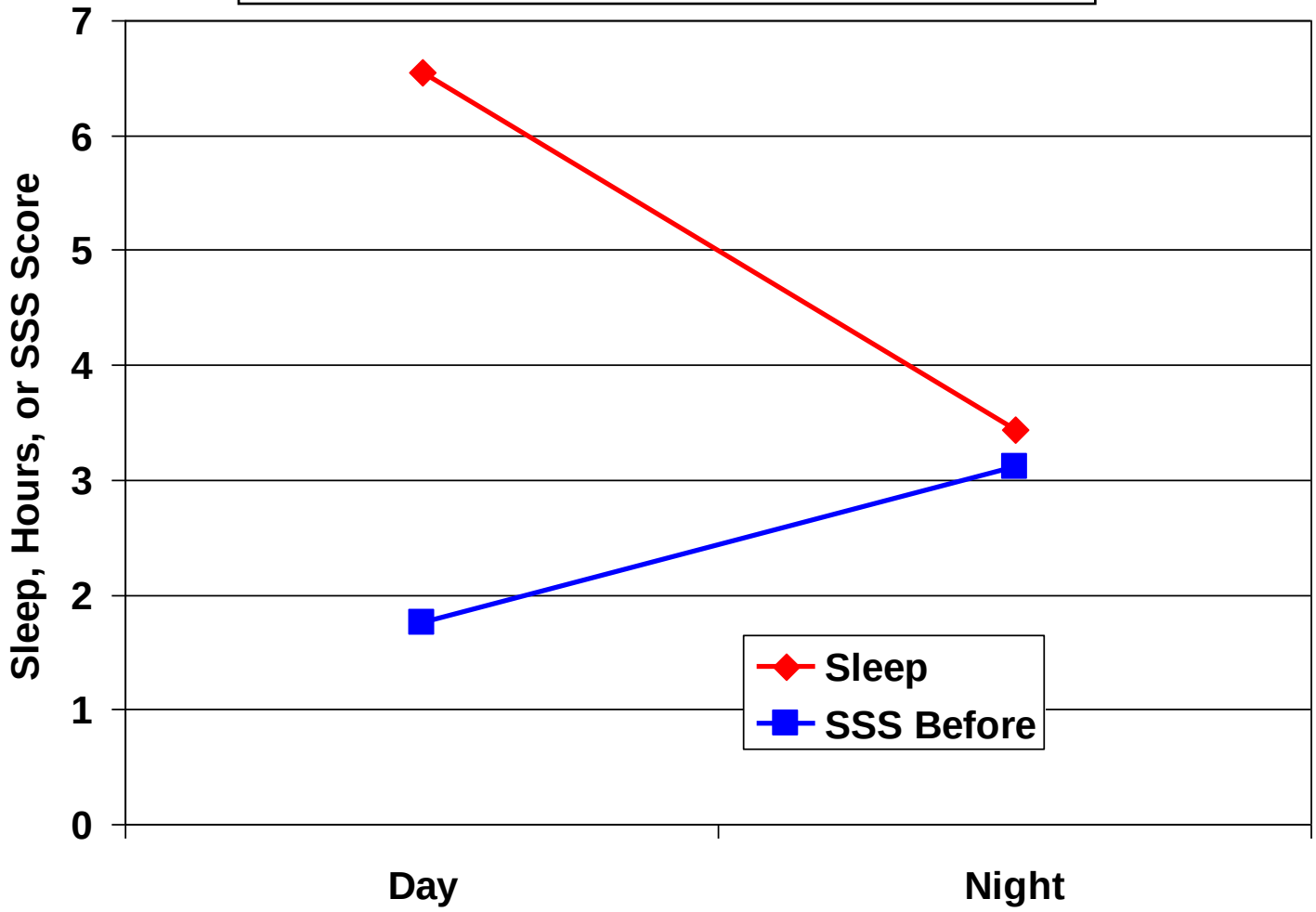
PoD Curves


PoD for Crack Contrast
Probability of Detection
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Crack Contrast (mean intensity)

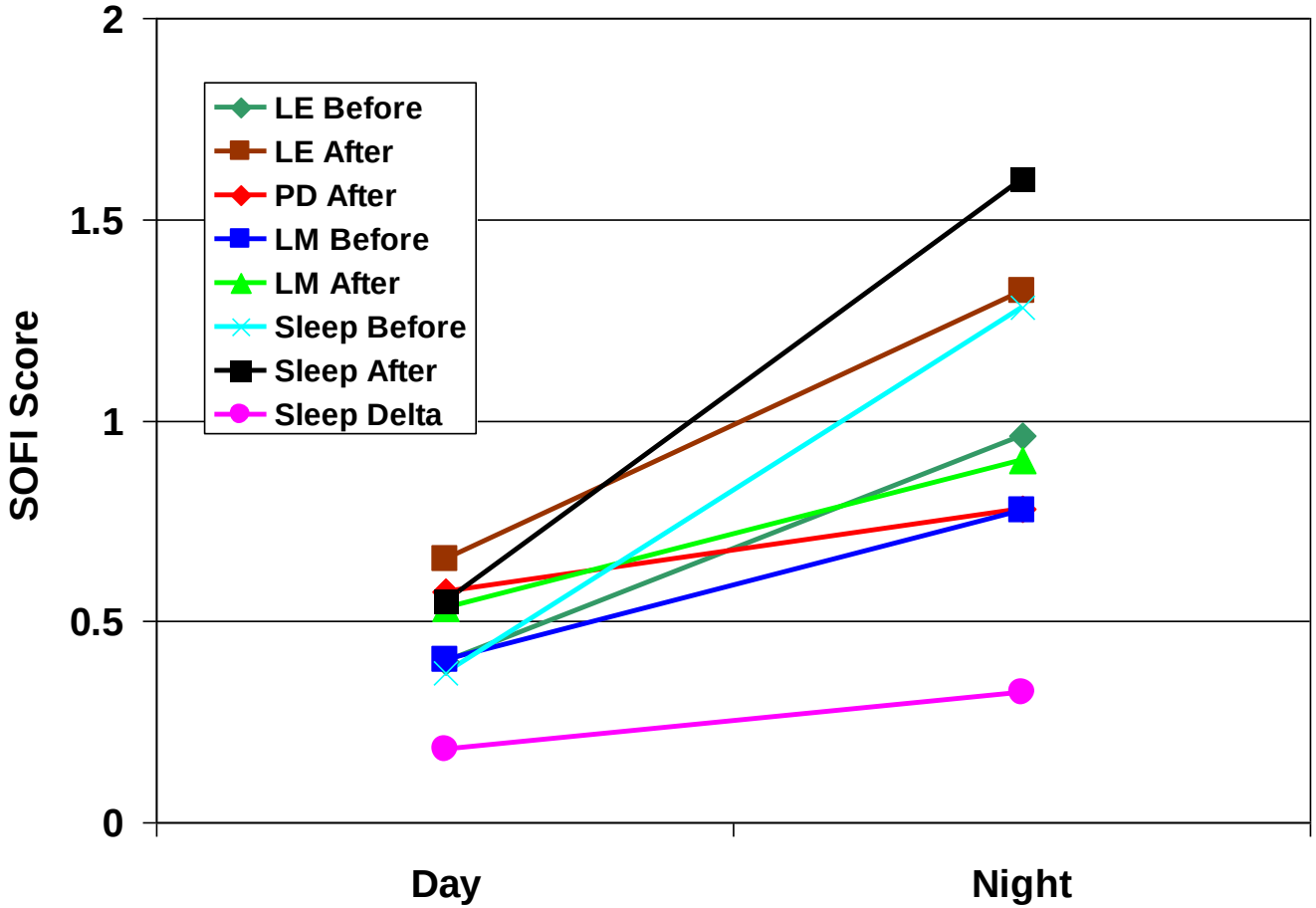
Results: Sleep Measures

- Longer sleep before task when task at 0900 than when at 0300 (graph 1 in red)
- People did not seem to adjust their bedtime for the 0300 start
- Confirmed by higher sleepiness score 0300 (graph1 in blue)
- Many measures from SSS and SOFI were worse at night (0300) than day (0900)

Hours of Sleep & SSS Before



SSS and SOFI Measures



Results: Correlations

- 5 Performance measures, 3 sleep measures, 15 sleep scales
- No significant correlations out of 15 between 3 sleep and 5 performance measures at $p > 0.05$
- Only 2 out of 90 significant correlations between 15 sleep scales and 6 performance measures at $p > 0.05$
- Conclude that sleep not related to performance on this task

Experiment 2 Considerations

- No real need for parametric tests, e.g. more times of day, more lighting levels, more break conditions
- Real need to confirm / deny this lack of vigilance decrement found in Experiment 1
- Replicated part of Experiment 1 with airline and repair station NDI inspectors
- Much travel to sites and night-shift working (all done by Brian Green and Jui-Feng Lin!)

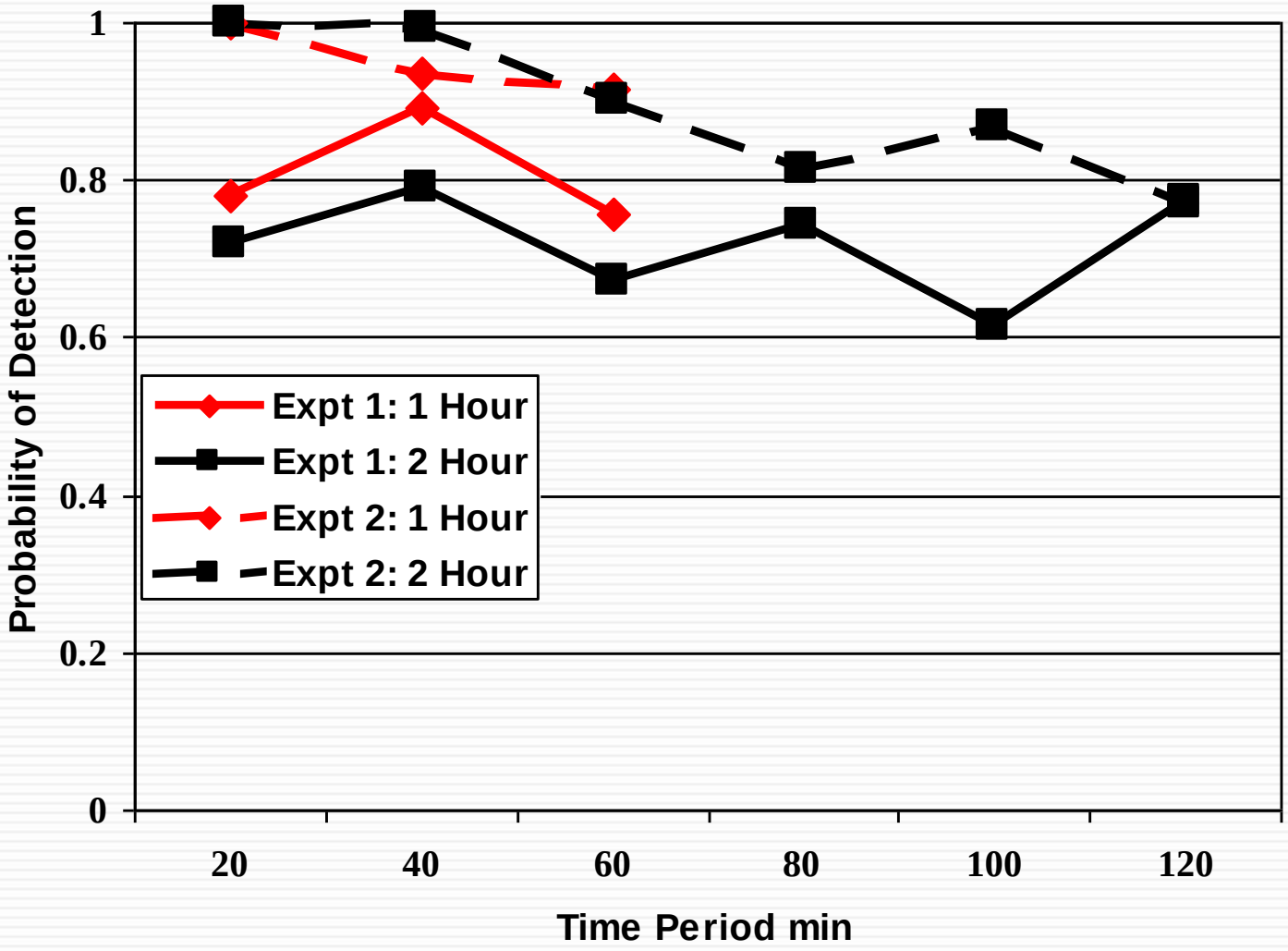
Experiment 2 Design

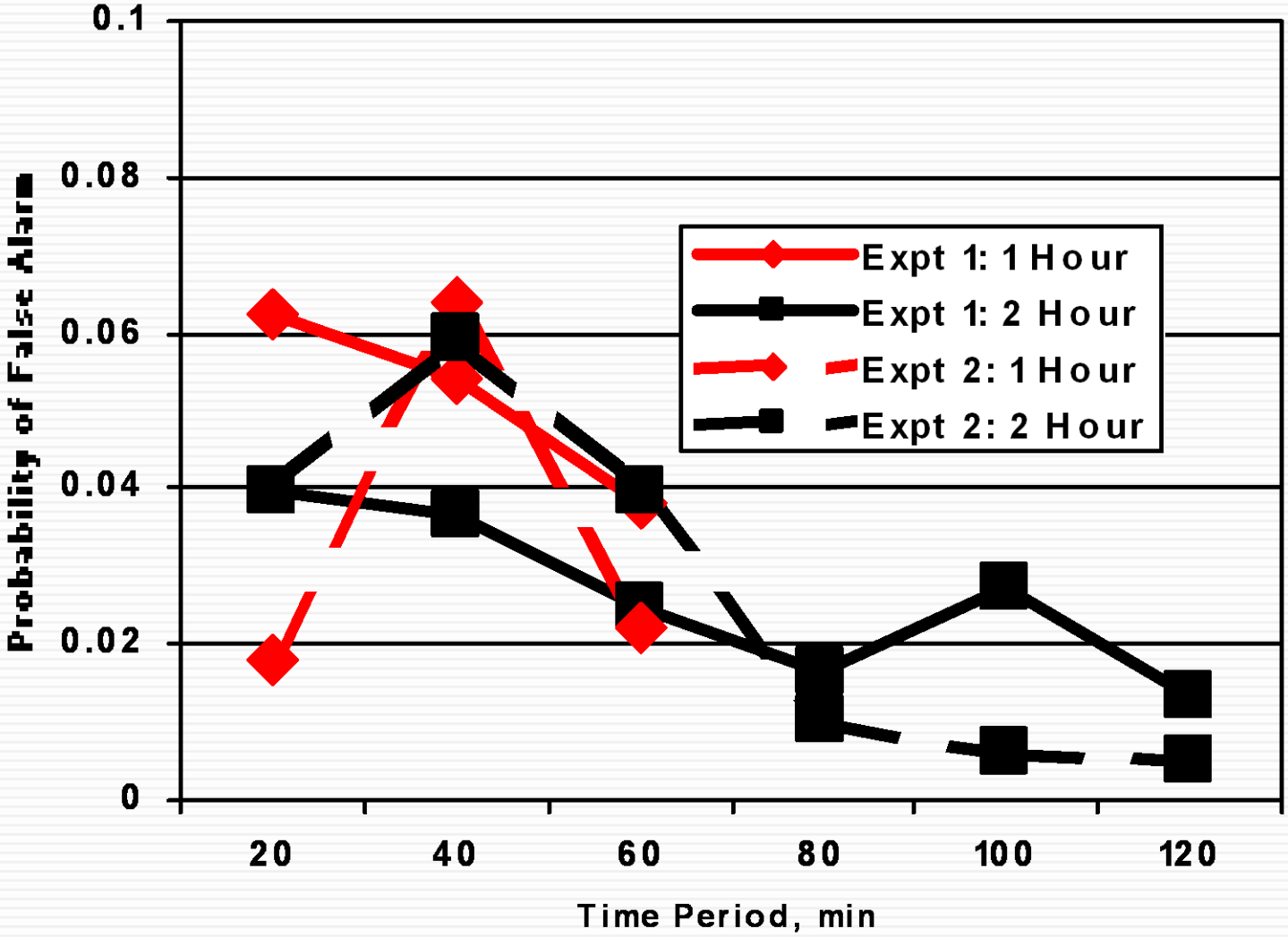
- Tested Shifts, Hours, Breaks for 70 inspectors
- Shifts was between-participants this time to fit with shift system of sites
- Some inspectors had to do night test at 0000 rather than 0300 (no difference in results)
- Resulted in a 2^3 between-participants factorial design plus Time Period within participants
- Same measures as Experiment 1: no ActiWatch
- Change in design made direct comparisons between experiments difficult.

Experiment 2 Results

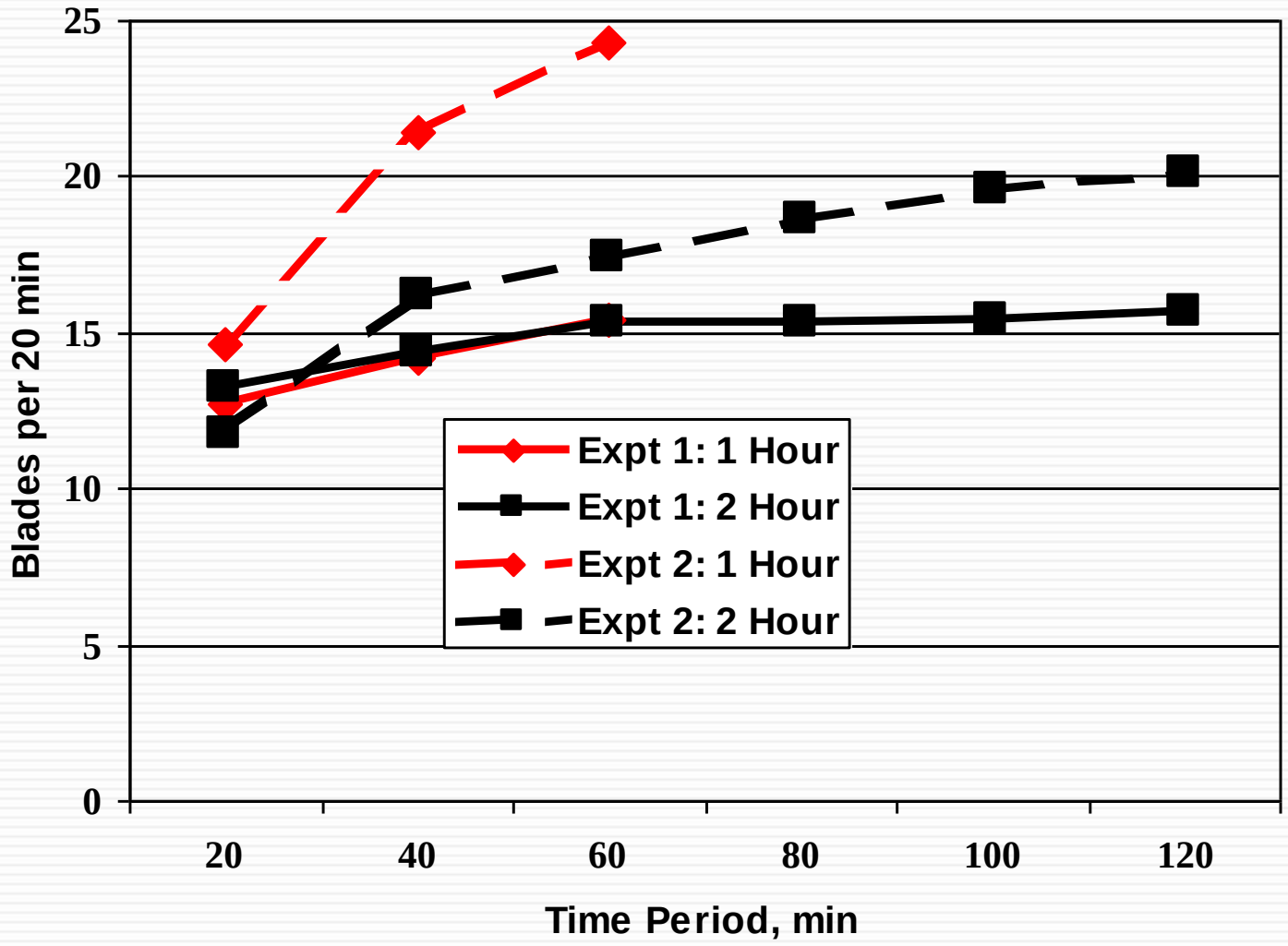
- **Again mainly Time Period effects presented Significant performance effects:**
  - Overall, inspectors faster (36 blades per hour in Experiment 1 vs 60-75 in Experiment 2) with higher PoD in Experiment 2 (but response changed a little in Experiment 2)
  - Shifts X Breaks as in Experiment 1
  - Same decrease in p(FA) with Time Period, but now a *decrease* in PoD with Time Period
- **Again, few significant correlations between performance, sleep and fatigue / sleepiness**

FAA 2008


1
0.8
0.6
0.4
0.2
0
Probability of Detection
20
40
60
80
100
120
Time Period min
Expt 1: 1 Hour
Expt 1: 2 Hour
Expt 2: 1 Hour
Expt 2: 2 Hour


0.1
0.08
0.06
0.04
0.02
0
Probability of False Alarm
20
40
60
80
100
120
Time Period, min
Expt 1: 1 Hour
Expt 1: 2 Hour
Expt 2: 1 Hour
Expt 2: 2 Hour

FAA 2008


25
20
15
10
5
0
Blades per 20 min
20
40
60
80
100
120
Time Period, min
Expt 1: 1 Hour
Expt 1: 2 Hour
Expt 2: 1 Hour
Expt 2: 2 Hour

So what changed, what did not?

- Only high level analysis presented here: 86 + 70 participants, 5 performance measures, 5 self-report scales, many sub-factors and covariates
- Similarities:
  - Breaks help performance (speed) but only by day and mainly for 2-hour condition
  - Few correlations between types of measures
- Differences:
  - Mainly a significant Vigilance decrement using experienced inspectors but not industrial population

…and Why?

- Some hypotheses but more than a single difference between the experiments
- Note careful simulation, validated by inspectors and by those in Experiment 2.
- Note many significant effects found, so experimental power in either experiment not a problem
- Obvious difference is between the two participant populations: experienced inspectors vs. general public with some industrial experience

…more Differences

- Experiment 1 trained participants for ~3 hours until task could be performed well. Experiment 2 used people who knew aviation inspection (not always FPI though) with short training on the simulation task only.
- Experiment 1 performed in isolated room, with recording of hangar noise. Experiment 2 performed in area of hangar chosen by site: usually break room or office
- Mean ages quite similar, but no female inspectors vs. 22% females in Experiment 1 (but no gender effect there, as in most inspection studies)

Results: Expertise Comparison

	Novices	Inspectors	t-value	Probability
Probability of Detection	0.742	0.906	6.56	p<0.001
Probability of False Alarm	0.348	0.029	0.56	p = 0.578
Blades / 20 min	11.04	18.2	5.47	p<0.001
Accept Time, s	134.0	100.0	1.70	p = 0.091
Reject Time, s	94.3	56.0	4.40	p<0.001
Workload, TLX Day 1	52.3	49.1	1.11	p = 0.271
Workload TLX Day 2	49.0	*		

Practical Conclusions

- Treat vigilance decrement as real for this complex inspection task
  - (Here I eat my words about the applicability of vigilance findings to complex inspection tasks, HFES, 2004-2006)
- Provide inspection breaks during day shift
- Use caution about linking vigilance and inspection

- …and be VERY careful about design of experiments using other than target participants!

Fatigue in aviation inspection: sleep and performance data

Colin G. Drury

Industrial and Systems Engineering
University at Buffalo: SUNY

Aviation Fatigue Management Symposium 2008

Presented at the FAA Fatigue Management Symposium: Partnerships for Solutions; Vienna, VA: June 17-19, 2008