Paper #

Paired Design for Treadwear Testing

By Leighton Spadone*
DAAS Inc., Tire Technology and Data Analytics
Beachwood, OH

Presented ITEC Technical Meeting
Akron, OH
September 13-15, 2016

* Speaker

Abstract

Paired Design Treadwear Analysis

Introduction:
A Paired Design-of-Experiment reduces treadwear test cost by reducing the time, vehicles, and tires needed to detect statistically significant treadwear changes.

Background:
Traditional treadwear test design uses one compound per tire and is expensive due to the number of tires and vehicles needed improve the signal to noise ratio and make valid comparisons in the face of the high number of uncontrolled test variables. The analysis used for Multiple Mean comparisons is less powerful than the Paired, One Sample Mean analysis.

Summary:
This example of a Paired Design compared compound treadwear for nine experimental tread compounds vs. control and used only 12 tires and 3 vehicles. The experimental design, assumptions, analysis methodology, caveats, and conclusions are reviewed.

This retreaded passenger tire experiment used 4 different tread compounds per tire. The 4-Way-Treads have different compounds for each quarter of the tread circumference. One of these 4 compounds is always the standard control compound.

The analysis methodology subtracts the wear measurements of the Control on each tire from each Experimental Compound within and across grooves. The mean of these wear differences, taking into consideration their variance, are compared to zero, which is the null hypothesis that asserts there is no difference in wear.

Benefits:
Paired Designs block-out the variation due to tire position, vehicle, driver, route, load, inflation, speed, road surface, and weather conditions during the treadwear testing. This reduced test variation and combined with the power of the One Sample Mean analysis, increases the statistical power to detect wear differences between compounds.

Paired Design Treadwear Experiment and Analysis

Introduction:
Paired Design experiments reduce treadwear-testing cost by reducing the number of test tires needed to detect treadwear improvements.

When two experimental treatments experience the same conditions, like tires kept paired together by axle or when the same tire is tested under different conditions, such as on road and on wheel, these are examples of Paired Designs.

In this Paired Design all the tread compounds are on a single tire and experience the same test conditions. The measurements from paired conditions are subtracted from each other and this difference is compared to zero. A difference from zero is considered significant when its T-test results in a probability of a greater T-value, the Prob>t, is equal to or less than 0.05 or Odds of 20:1 or better.

Background:
Treadwear road testing is expensive propositions because of the presence of uncontrolled variables increases the variation in the test results and makes it more difficult to detect significant differences in wear between compounds without a larger number of tires.

Difficult to control variation is introduced into treadwear testing when tires experience different **drivers, routes, vehicles, tire positions, inflations, loads, speeds, alignments, pavements, ambient temperatures, and precipitation.**

Methodology:
This Passenger Tire Paired Design evaluates the treadwear vs. control for nine experimental tread compounds, using only 12 tires and 3 vehicles in 16K miles. This experiment is intended to rapidly screen tread compounds for improved wear potential.

These retreaded passenger tires used four different tread compounds paired on the same tire. These 4-Way-Tread tires have different experimental compounds for each ¼ of the tread circumference. One of these four tread compounds is always the control compound.

The Paired methodology subtracts the wear measurements of the Control from each experimental compound on a tire and statistically tests to see if the mean differences in the remaining tread depths between compounds are significantly greater than zero. This Single Mean test has maximum statistical power to detect differences.

Benefits:
This Paired Design blocks-out most of the test noise and tire variation. Reducing the variation increases the power of the experiment to detect wear differences between compounds. This methodology is applicable to Passenger, Light Truck, and Medium Truck and Bus tires.

Assumptions:
The wear differences measured are due to **only** to the tread compound formulations.

- The test tires provided uniformity of compound, cure, and footprint shape.
- The tread depth measurement system was accurate and repeatable.
- Tire alignment, camber, toe-in toe-out, rims, etc. was uniform across vehicles.
- The tire rotation schedules were followed.

These comparisons rely on the calculation of T-Values and associated P-Values to make inferences about significant differences between compounds and these procedures assume the measurements are approximately normally distributed.

Caveats:
This screening test was stopped at 16K miles or 25,750 kilometers. Extrapolating these wear improvements to higher mileages is not reliable.

This is a screening experiment that is designed to find promising tread compound candidates for further study. This analysis does not “prove” that one compound is better than another except in regards to the control compound.

Choosing of the correct number of tires and compounds to evaluate in order to provide sufficient test data to detect wear changes is not straightforward and depends on some assumptions and estimates.

The Power to find differences between treatments depends on: 1) The type of statistical comparison, 2) The number of tires, 3) The standard deviation of the wear in each groove or across grooves, and 4) The wear Difference-to-Detect. Often, the Standard Deviation and the Difference-to-Detect may not be well known before the data analysis.

For example, each compound-groove comparison in this study consisted of n= 4 tires and the Power-to-Detect a 0.18mm, 7 mil, wear difference for these single mean comparisons at a P-Value <= 0.05, varied from a estimated Power=0.2 to 1.0, that is, from very poor to excellent, due to the wide range of wear Standard Deviations encountered. (0.250 to 0.020).

Testing Method:
Rotate tire position on vehicle each 1K miles and rotate vehicles each 4K. One wear measurement was reported for each compound's 90-degree section, in each groove, for each tire. Grooves G1 & G6 are at the tire shoulders and grooves G2 to G5 bracket the tread centerline. The test was stopped at 16,000 miles or 25,750 kilometers.

Data:
The detailed Paired analysis was made by groove and compound and required 54 comparisons vs. control. There are 4 compounds per tire, each with 6 groove measurements for the 12 test tires or 288 data points. There are 4 tires in each group of experimental compounds: ABCD, AEFG, AHKL for a total of 12 tires. (**Appendix 2: Table V: & Table VI:**)

A second Matched Pair analysis used 9 compound comparisons that combine data across the 4 central grooves, G2-G5. The measurements for the 4 center grooves are combined for each tire providing n=16 data points, for a total of 192 data points used in this analysis. (**Appendix 3: Table VII:**)

Table I: below is sample of the data table that contains the remaining tread-NS depth measurements in millimeters by tire and groove. Compound A is the control and experimental compounds are labeled BCD EFG HKL. The test tires are labeled 1-12 and the tire grooves are labeled G1-G6.

Table Number I

Tire	Cmpd	Tire Cmpd	G1 mm	G2 mm	G3 mm	G4 mm	G5 mm	G6 mm
1	A	01A	0.965	1.397	2.362	2.159	1.575	0.889
1	B	01B	1.168	1.575	2.591	2.413	1.626	0.889
1	C	01C	0.965	1.88	2.718	2.743	1.676	0.838
1	D	01D	0.94	1.575	2.565	2.413	1.549	0.813
2	A	02A	1.295	1.473	2.261	2.184	1.321	0.914
2	B	02B	1.245	1.676	2.54	2.464	1.499	0.813
2	C	02C	1.219	1.626	2.438	2.515	1.727	0.864
2	D	02D	1.194	1.397	2.311	2.438	1.499	0.838
3	A	03A	1.168	1.346	2.159	2.21	1.422	0.838
3	B	03B	1.168	1.397	2.159	2.261	1.549	0.787
3	C	03C	1.016	1.397	2.057	2.261	1.448	0.787
3	D	03D	1.194	1.295	2.388	2.261	1.372	0.635

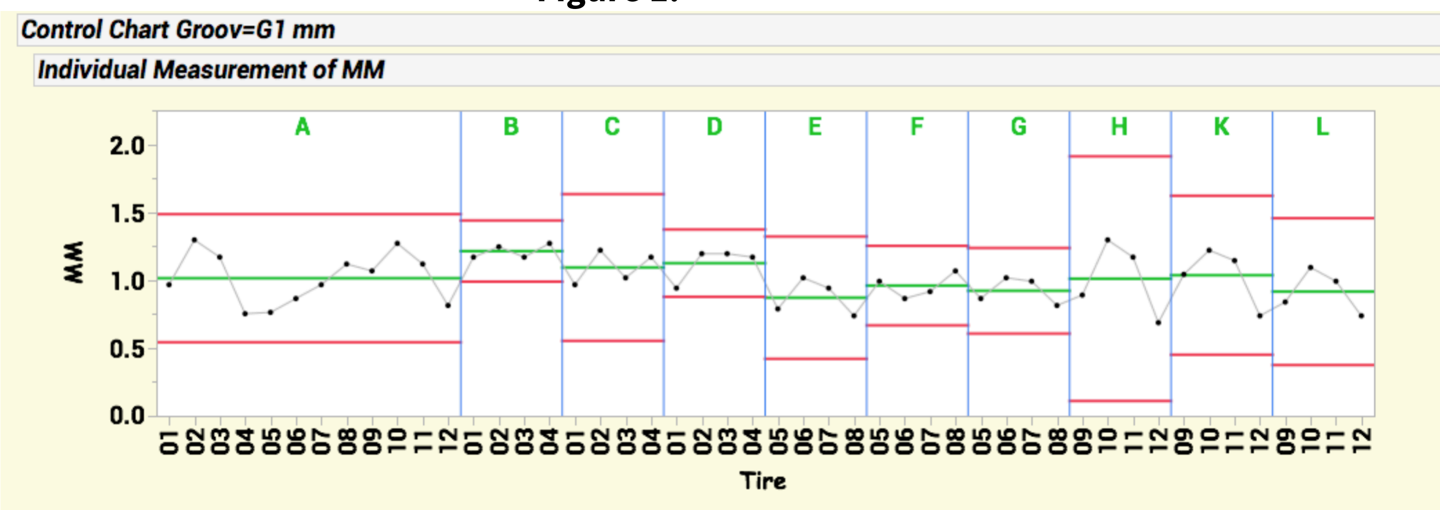
Table I: Sample data in millimeters of remaining tread NS by tire, compound and groove. Non-Skid measurements of 48 rows by 6 columns = 288 data points.

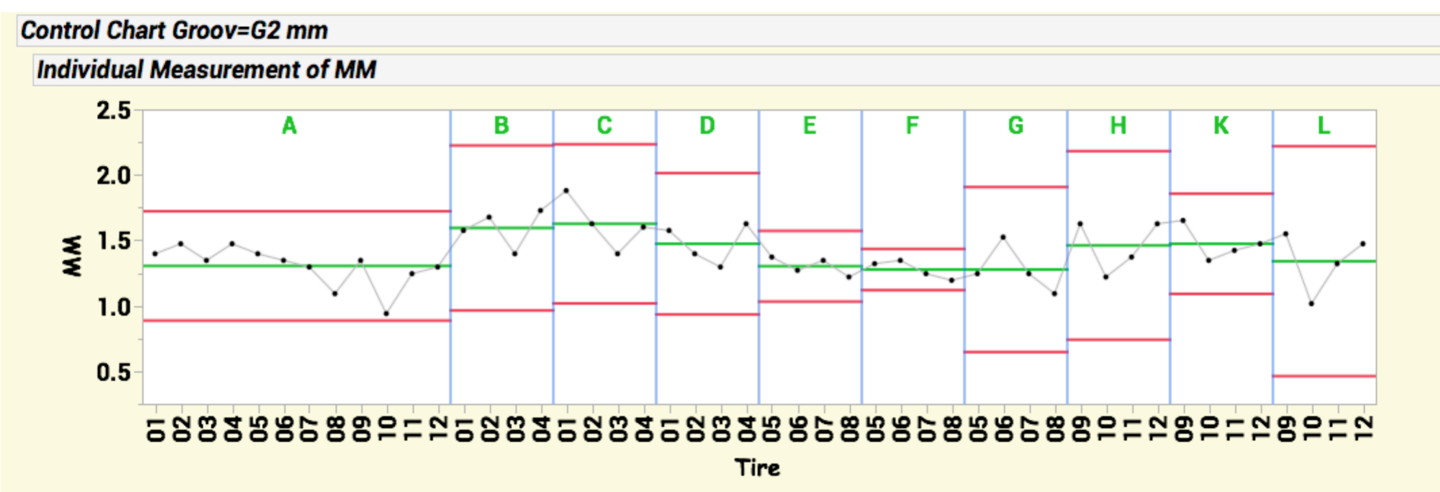
Analysis Precondition:
Comparing unstable wear processes would not provide useful information to predict compound wear. As a prerequisite to analysis, the wear process should be analyzed and shown to be stable and predictable and to be able to support valid wear comparisons.

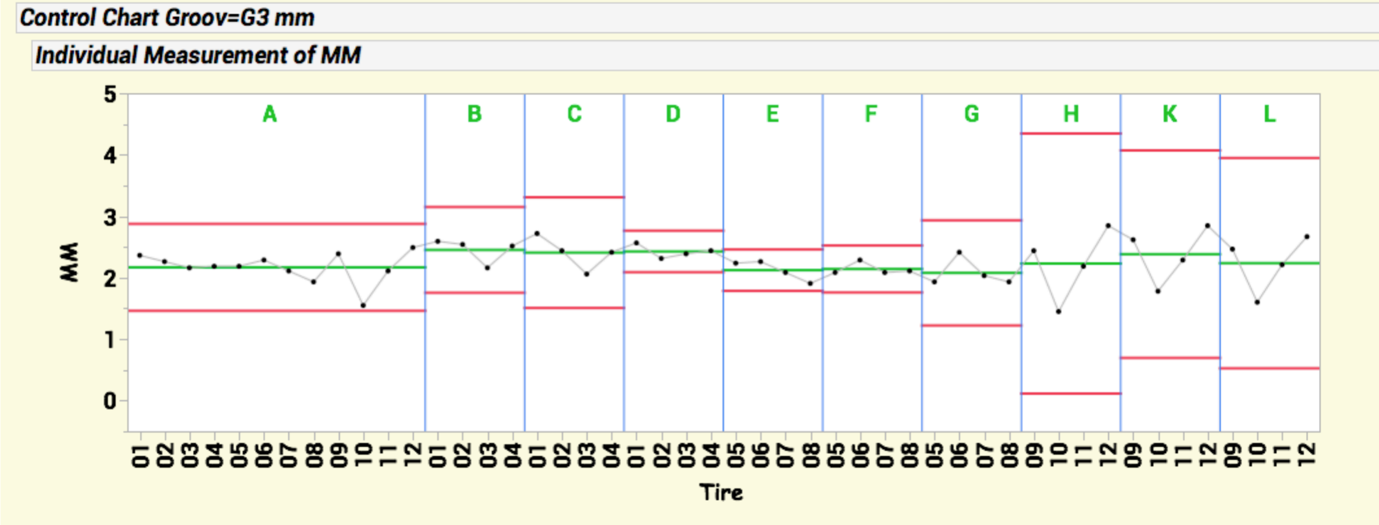
Shewhart process control charts for the treadwear measurements are shown in Fig 1 below. Non-Skid measurement is shown by compound for each groove for all 12 tires. The charts show that the wear process is stable for all grooves and all tires. That is, the wear measurements are within the upper and lower red control limit lines. This type of analysis does not assume normally distributed data.

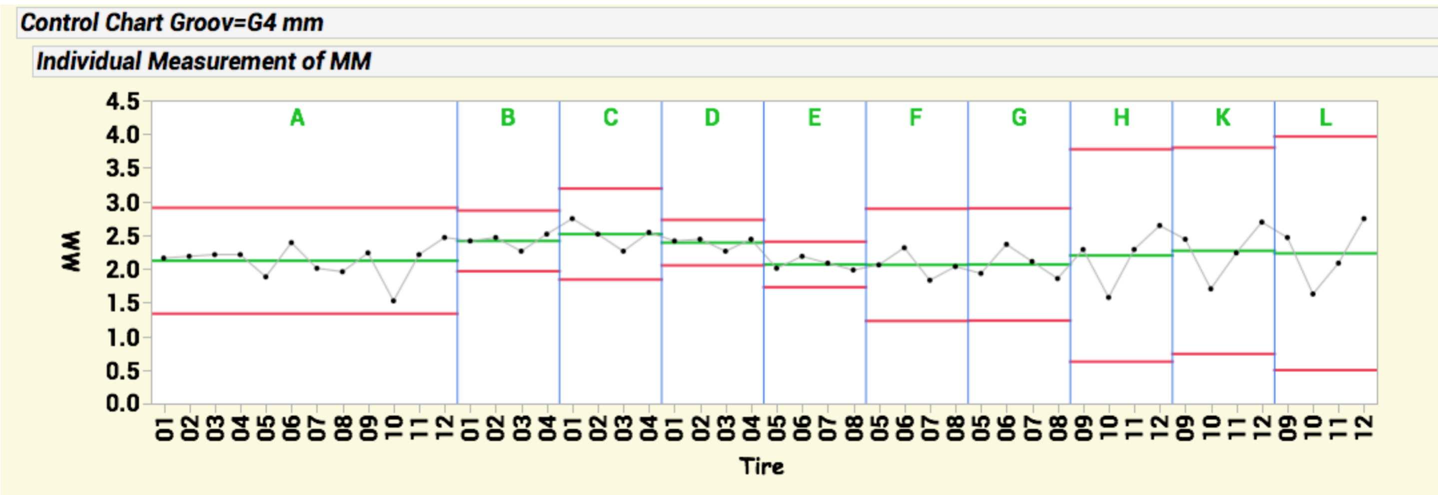
Fig1: Remaining Tread (Non-Skid) Measurements Shewhart Process Control Charts

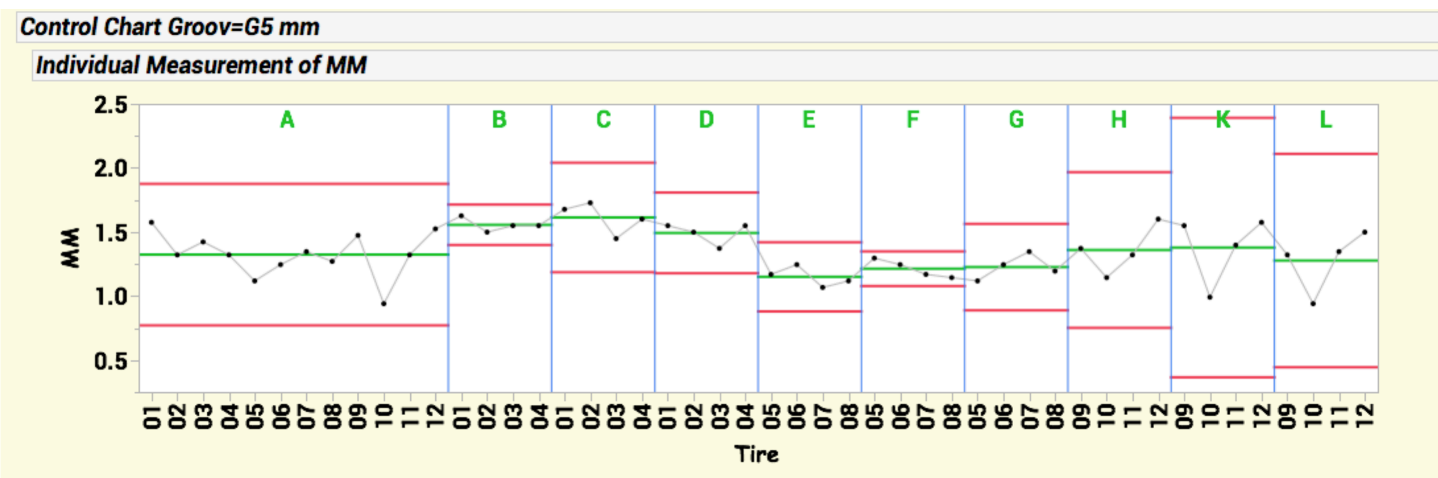
Figure 1:






Control Chart Groov=G3 mm
Individual Measurement of MM
MM
Tire
A B C D E F G H K L


Control Chart Groov=G4 mm
Individual Measurement of MM
MM
Tire
A B C D E F G H K L


Control Chart Groov=G5 mm
Individual Measurement of MM
MM
Tire
A B C D E F G H K L

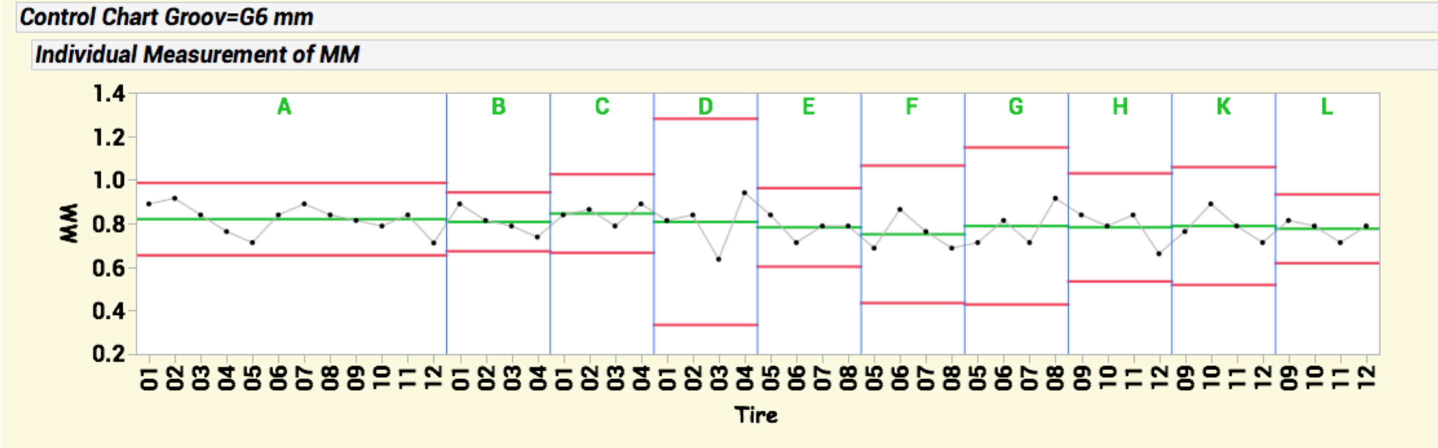


Fig1: SAS JMP Control Chart Tire Wear Process by compound for each tire by groove.

Wear Analysis by Compound-Groove:

Fig 2: and Table II: below highlights those experimental compounds by groove, whose combination of mean wear difference from control and their associated standard deviations allowed the rejection of the null hypothesis, that the wear difference is zero.

Figure 2

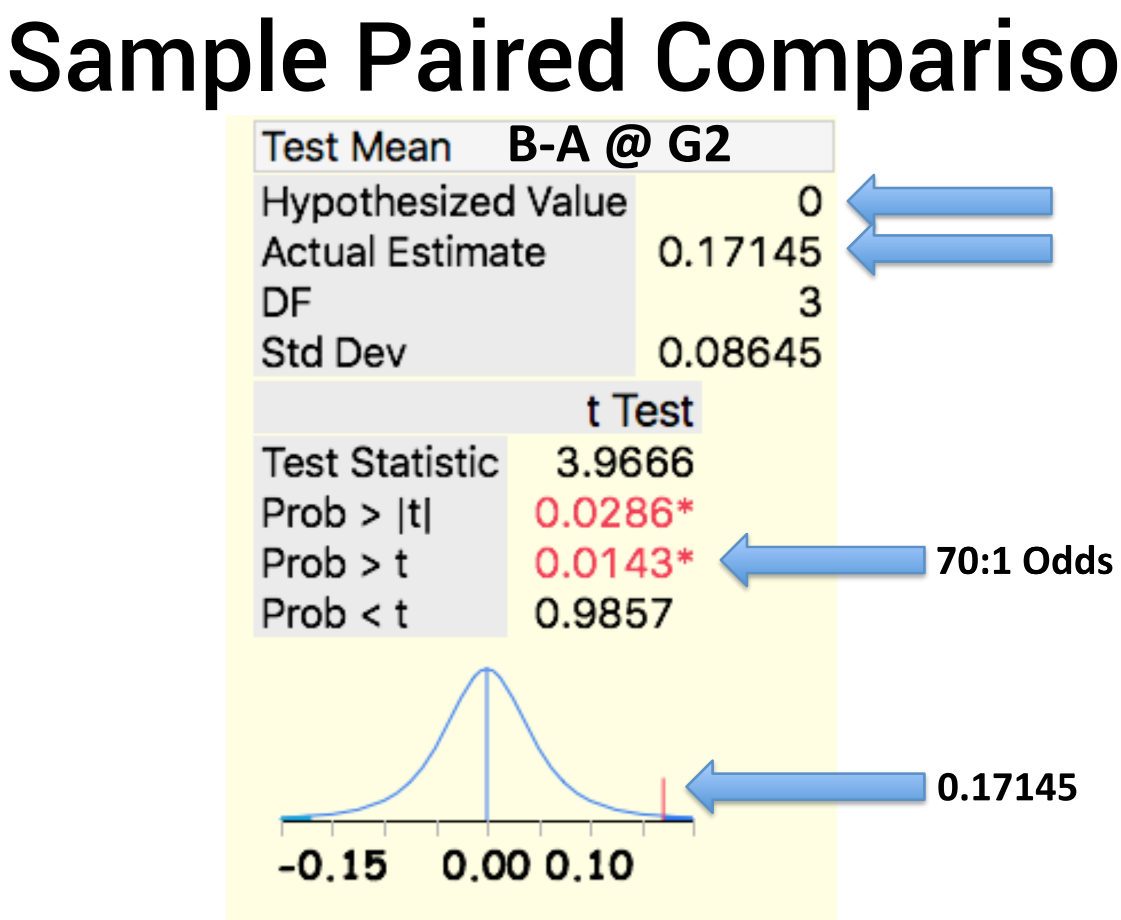


Fig2: Sample One-Mean test statistic of the wear difference between Compound B and the Control compound at Groove 2.

The highlighted compounds below are improved treadwear candidates since the null hypothesis, that the difference in wear is equal to zero, has been rejected at a Prob>t < 0.05. That is, 20 to 1 odds or better bet against the idea that this large of an increase in wear performance may have happen by chance. The 20 to 1 odds are a good bet, but with 54 comparisons, there is still a 90% chance that some of these comparisons are false positives and not as significant. See **Appendix 2: Fig 3:** for a P-Value's Log-Worth plot vs. T-Value that could be used to choose a smaller, alternative, P-Value to reduce the false positive rate.

Table Number II

	G1	G2	G3	G4	G5	G6
B		+B	+B	+B	+B	
C				+C	+C	
D			+D	+D		
E						
F						
G						
H		+H		+H		
K		+K	+K	+K	+K	
L		+L	+L			

Table II: The candidate compounds for wear improvements by groove and compound across N=4 tires. See **Apendix2** for the comparison's P-Values for this table.

Statistical significance, P-Value<=0.05, is not always synonymous with practical significance, since the size of the effect, the wear difference, also needs to be considered. Therefore the difference from the mean wear by groove is graphically summarized in **Appendix1: Fig 3:**

Matched Pairs Analysis combined across Grooves:

Table III: below highlights the significant experimental compounds when evaluated by combining center grooves G2-G5 in a second Matched Pairs analysis. These 9 comparison, each with n=16 data points, 4 tires x 4 grooves, have a false positive rate of less than 1% using at a Prob>t =< 0.001 or 1000:1 Odds or better. Also included below in **Table IV:** is the rank order of the mean wear improvement by compound.

Table Number III

	G2	G3	G4	G5
B+	(red)	(red)	(red)	(red)
C+	(red)	(red)	(red)	(red)
D+	(red)	(red)	(red)	(red)
E				
F				
G				
H+	(red)	(red)	(red)	(red)
K+	(red)	(red)	(red)	(red)
L+	(red)	(red)	(red)	(red)

Table Number IV

Rank Order

Compare Cmpd	mm Mean Difference
C-A	0.224
B-A	0.187
K-A	0.184
D-A	0.129
H-A	0.121
L-A	0.079

Table III: The experimental compounds with statistically significant wear improvements across grooves G2-G4 with N=16 data points. **Table IV** is the Rank Order of compound mean wear differences. **Table IV** rank order by wear difference only tested the difference between the experimentals and the control. See **Appendix 3: Table VII:** for the comparison P-Values.

Summary

The wear process was stable and in-control for all compounds by tire and groove.

Shoulder grooves G1 and G6 show fast wear, that is, a lower NS-tread depth, than the central G2-G6 tire grooves (**Fig 1**). These shoulder grooves show no significant wear difference between the control and any of the experimental compounds (**Table II**).

Focusing just on the center grooves, G2-G5 in the Paired Analysis by groove, compounds B, C, D, H, K, & L each were candidates for wear increase vs. control in two or more grooves. (**Table II). Appendix 2: Table V & Table VI:** tabulates these significance tests for all grooves.

Again considering just the central grooves, G2-G5, the second Matched Pairs Analysis across these four grooves by compound with N=16 data points, show experimental compounds B, C, D, H, K, & L all being candidates for wear improvements compared to the control (**Table III**). **Appendix 3: Table VII** tabulates these significance tests across grooves G2-G5 combined.

The Paired Design screening experiment and associated analysis detected treadwear differences vs. control among the 9 experimental compounds by testing only 12 tires on 3 vehicles after 16,000 miles.

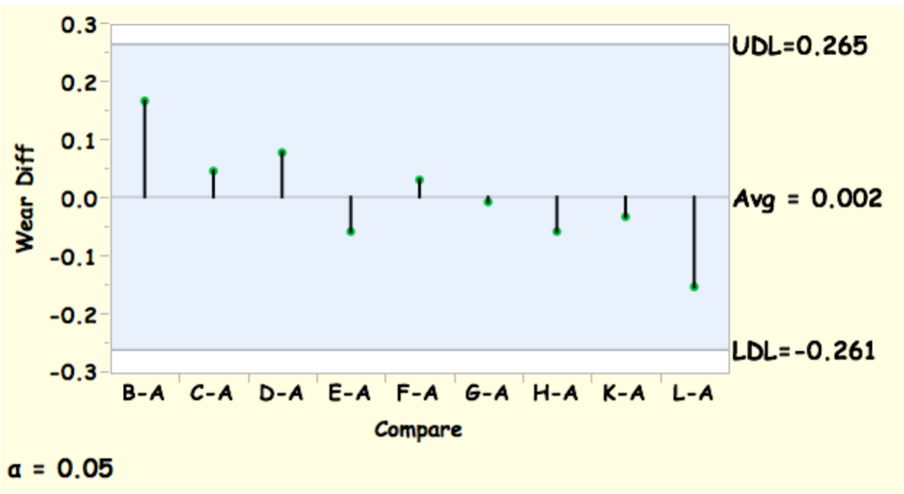
Conclusions
The Paired Design reduces treadwear test cost by detecting significant compound wear improvements with fewer tires and vehicles than traditional, one compound per tire designs.

Recommendations
A flat mold retreading process for multi section treads would allow Paired Comparison of compounds within a tread design or tread designs within compound with lower cost and higher confidence than traditional one-compound or one-design per tire wear testing.

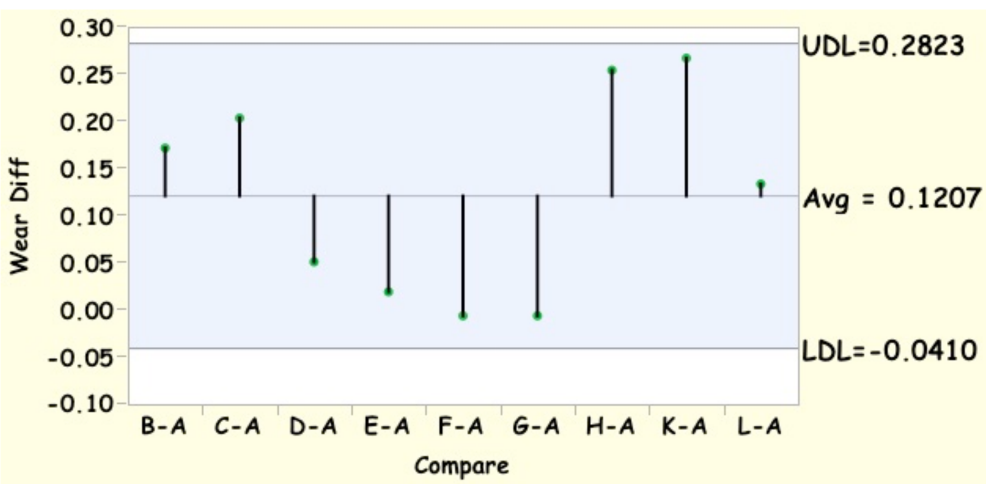
Appendix1: Fig 3: Wear Difference distribution from overall mean by Groove & Compound.
Note that the mean wear midline value and scale changes from chart to chart.
The UDL and LDL limits below are not applicable for this Paired Design experiment.
Charts were created by SAS JMP Analysis of Means (ANOM) procedure.

Figure 3

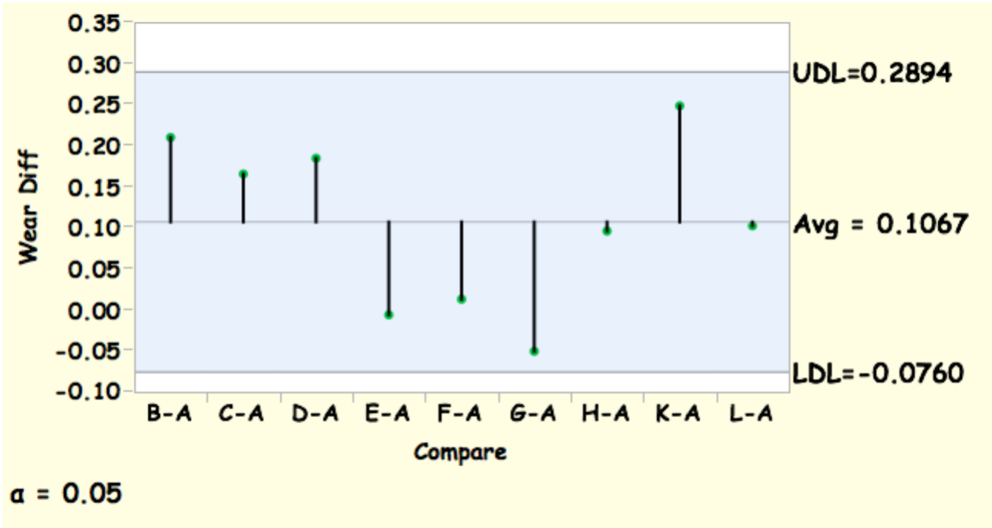
Groove 1



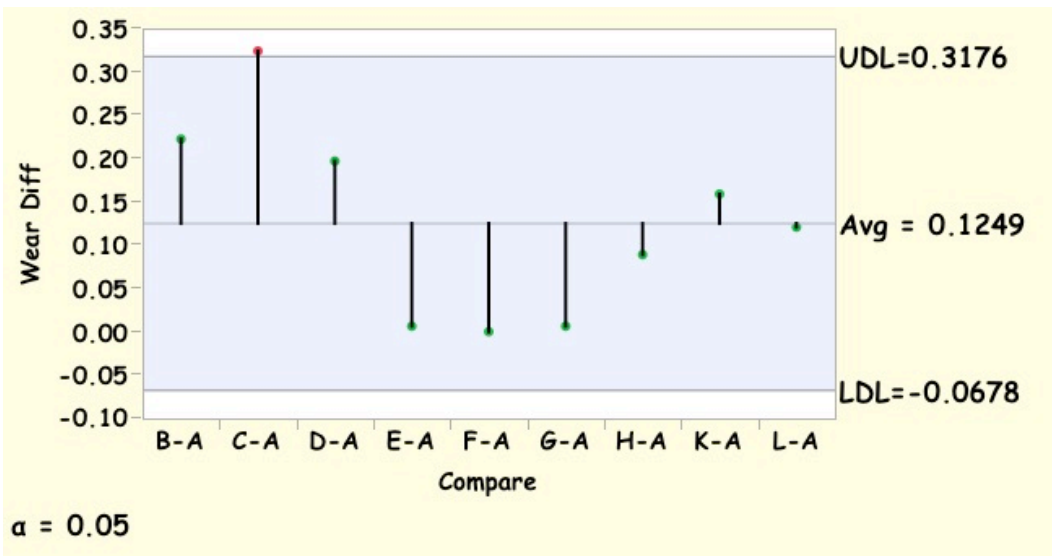
Groove 2



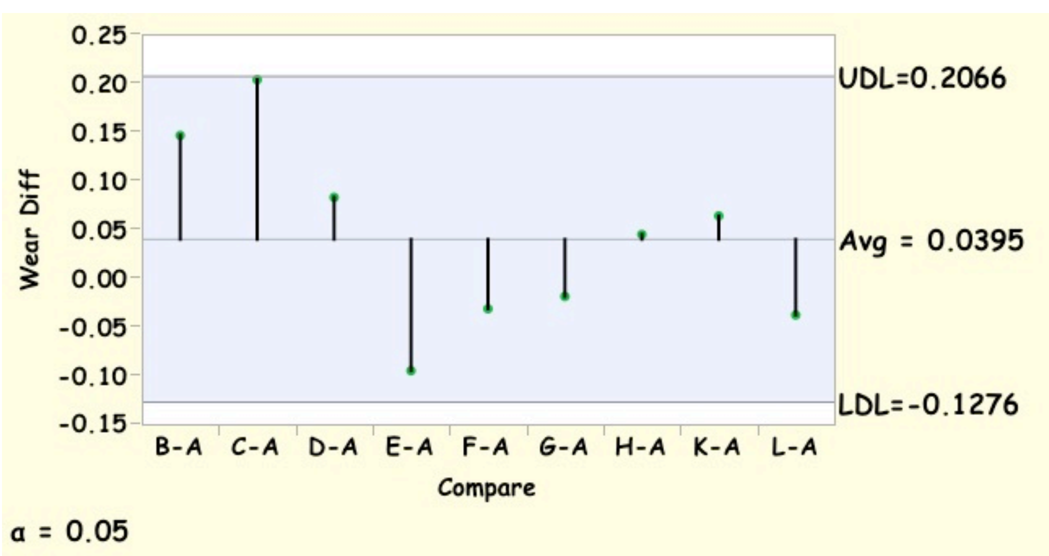
Groove 3



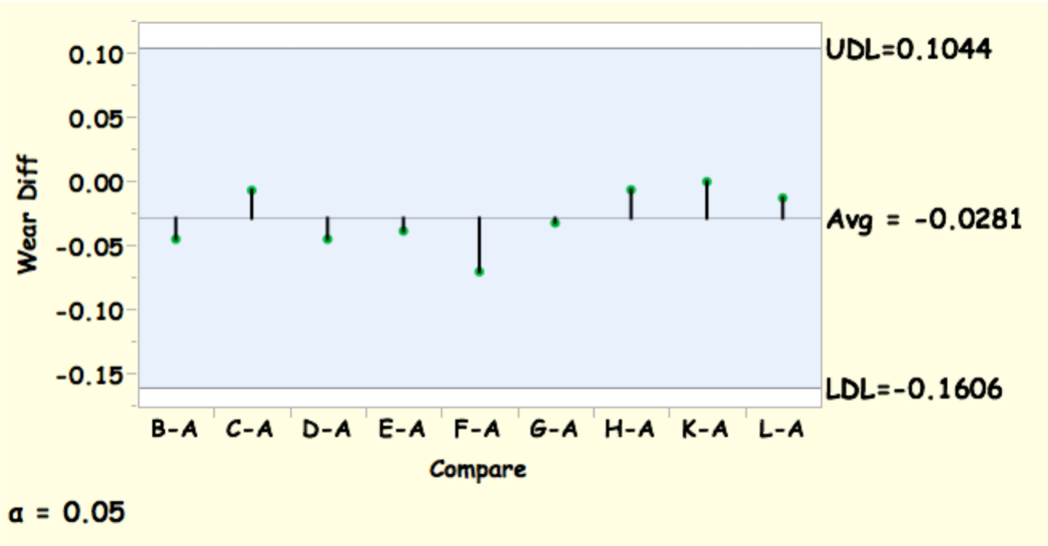
Groove 4



Groove 5



Groove 6



0.10
0.05
0.00
-0.05
-0.10
-0.15
Wear Diff
UDL=0.1044
Avg = -0.0281
LDL=-0.1606
B-A
C-A
D-A
E-A
F-A
G-A
H-A
K-A
L-A
Compare
α = 0.05

Appendix 2: Significant Compounds by Groove with N=4 tires. (Ref **Table II:**)
Paired Analysis Comparisons by groove (1-27 of 54) using JMP Distribution Platform.

Table Number V:

Groove	Compare	Mean Difference	Std Dev	t-Value	PValue P>t
G1 mm	B-A	0.168	0.258	1.299	0.142
G2 mm	B-A	0.171	0.086	3.967	0.014
G3 mm	B-A	0.210	0.146	2.876	0.032
G4 mm	B-A	0.222	0.116	3.826	0.016
G5 mm	B-A	0.146	0.076	3.851	0.015
G6 mm	B-A	-0.044	0.043	-2.049	0.934
G1 mm	C-A	0.047	0.254	0.370	0.368
G2 mm	C-A	0.203	0.191	2.125	0.062
G3 mm	C-A	0.165	0.193	1.712	0.093
G4 mm	C-A	0.324	0.218	2.973	0.029
G5 mm	C-A	0.203	0.172	2.359	0.050
G6 mm	C-A	-0.006	0.089	-0.143	0.552
G1 mm	D-A	0.079	0.231	0.681	0.272
G2 mm	D-A	0.051	0.133	0.765	0.250
G3 mm	D-A	0.184	0.091	4.035	0.014
G4 mm	D-A	0.197	0.098	4.013	0.014
G5 mm	D-A	0.083	0.141	1.169	0.163
G6 mm	D-A	-0.044	0.160	-0.556	0.692
G1 mm	E-A	-0.057	0.228	-0.500	0.674
G2 mm	E-A	0.019	0.089	0.429	0.349
G3 mm	E-A	-0.006	0.038	-0.333	0.620
G4 mm	E-A	0.006	0.146	0.087	0.468
G5 mm	E-A	-0.095	0.150	-1.269	0.853
G6 mm	E-A	-0.038	0.115	-0.665	0.723
G1 mm	F-A	0.032	0.133	0.476	0.333
G2 mm	F-A	-0.006	0.079	-0.162	0.559
G3 mm	F-A	0.013	0.118	0.215	0.422

Appendix 2: Significant Compounds by Groove with N=4 tires. (Ref **Table II:**)

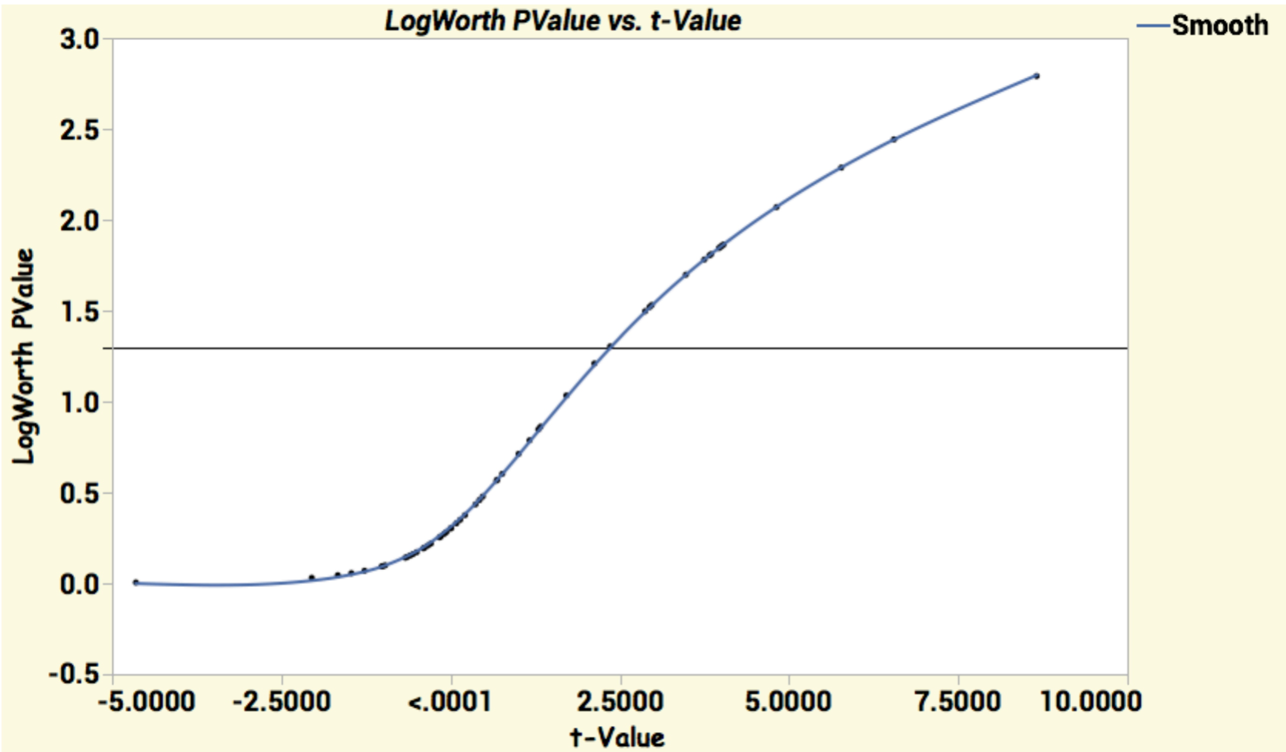
Paired Analysis Comparisons by groove (28-54 of 54) using JMP Distribution Platform.

Table Number VI:

Groove	Compare	Mean Difference	Std Dev	t-Value	PValue P>t
G4 mm	F-A	0.000	0.158	0.000	0.500
G5 mm	F-A	-0.032	0.158	-0.401	0.642
G6 mm	F-A	-0.070	0.084	-1.665	0.903
G1 mm	G-A	-0.006	0.206	-0.062	0.523
G2 mm	G-A	-0.006	0.138	-0.092	0.534
G3 mm	G-A	-0.051	0.159	-0.638	0.716
G4 mm	G-A	0.006	0.089	0.143	0.448
G5 mm	G-A	-0.019	0.038	-1.000	0.805
G6 mm	G-A	-0.032	0.107	-0.596	0.704
G1 mm	H-A	-0.057	0.112	-1.017	0.808
G2 mm	H-A	0.254	0.088	5.774	0.005
G3 mm	H-A	0.096	0.190	1.007	0.194
G4 mm	H-A	0.089	0.060	2.941	0.030
G5 mm	H-A	0.044	0.128	0.692	0.269
G6 mm	H-A	-0.006	0.031	-0.381	0.636
G1 mm	K-A	-0.032	0.043	-1.464	0.880
G2 mm	K-A	0.267	0.111	4.818	0.009
G3 mm	K-A	0.248	0.076	6.552	0.004
G4 mm	K-A	0.159	0.091	3.478	0.020
G5 mm	K-A	0.064	0.015	8.660	0.002
G6 mm	K-A	0.000	0.072	0.011	0.496
G1 mm	L-A	-0.152	0.066	-4.648	0.991
G2 mm	L-A	0.133	0.067	3.993	0.014
G3 mm	L-A	0.102	0.054	3.751	0.017
G4 mm	L-A	0.121	0.181	1.331	0.138
G5 mm	L-A	-0.038	0.079	-0.965	0.797
G6 mm	L-A	-0.012	0.085	-0.290	0.605

Fig 4: Log-Worth of P-Value, = -Log10 (P-Value), vs. the T-Value for the 54 comparisons. A Log-Worth of 1.3 is equivalent to a P-Value= 0.05, the critical P-Value used in this part of the study. Using Log-Worth values avoids the difficulty of graphing very small P-Values.

Figure 4:



Appendix 3: Table VII: Significant wear improvement compounds compared across central grooves G2-G5 with N=16 measurements for the 9 Compounds (Ref **Table III:**).
Generated by JMP Matched Pairs procedure.

Table Number VII:

Grooves	Cmpd	Mean Difference	Std Error	Lower 95%	Upper 95%	t-Ratio	Prob>t
G2-G5	B-A	0.1873	0.0257	0.1325	0.2421	7.2882	0.0000
G2-G5	C-A	0.2238	0.0461	0.1256	0.3221	4.8559	0.0001
G2-G5	D-A	0.1286	0.0310	0.0626	0.1946	4.1512	0.0004
G2-G5	E-A	-0.0191	0.0283	-0.0793	0.0412	-0.6742	0.7448
G2-G5	F-A	-0.0064	0.0299	-0.0701	0.0574	-0.2122	0.5826
G2-G5	G-A	-0.0175	0.0265	-0.0740	0.0390	-0.6587	0.7400
G2-G5	H-A	0.1208	0.0349	0.0463	0.1952	3.4568	0.0018
G2-G5	K-A	0.1843	0.0277	0.1252	0.2433	6.6541	0.0000
G2-G5	L-A	0.0795	0.0299	0.0156	0.1433	2.6539	0.0090