

- Main conclusion
 - Authors state: *"Dirt and gravel trail running surfaces do not have lower tibial accelerations or greater shock attenuation than paved surfaces."*
 - The phrasing in this statement is expressed with more certainty than the data supports. Further, the study didn't include enough people (n=15) to confidently say that small differences between surfaces don't exist
 - The results support the claim only within the context of the experiment's specific conditions, not that no differences exist at all outside of this study population
 - The conclusion should have been phrased to reflect the limitations of the study and uncertainty in the data
- Sample size, variability
 - The study included 15 participants, which reduces the statistical power and limits the ability to detect subtle differences between surfaces
 - Table 1 shows large standard deviations relative to the mean values, which indicate wide variability in the responses (ex. Lateral = 3.9 ± 2 ; Braking = 9.1 ± 3.5)
 - The large standard deviations result in wide confidence intervals, which means that the data is too variable to accurately estimate the true mean and limits the experiment's ability to distinguish small differences between surfaces
 - The mean differences between surfaces are smaller than the variabilities, so the ANOVAs lack the ability to detect subtle effects. Because the variability is greater than the differences themselves the tests are more likely to return "no significant differences", however, this doesn't mean none exist.
- Effects of averaging the data
 - Averaging the data could potentially hide important individual differences
 - For example, if one runner showed higher tibial acceleration on gravel while another shows lower acceleration, these opposing effects cancel out and create the appearance of there not being a change when averages, despite there being variability within the group
 - Since our data is highly variable, tests like ANOVA become less sensitive because they are designed to detect consistent patterns across the whole group (homo assumption), meaning they would be best when participants respond in the same way. However, when individual responses vary widely, these differences cancel and the test may conclude that no significant difference is found.
 - This study overlooks individual effects that could be relevant
- External factors
 - The variability may reflect the influence of external factors that have not been accounted for in the study such as differences in running form or fatigue levels.
 - Because these factors were not measured or controlled, it is difficult to say whether the variability is random noise or related to the unmeasured factors
 - This further limits the strength of the conclusion: if the variability can be explained through some unmeasured factor(s), then the current study's results reflect lack of information rather than lack of surface effect

- Stated limitations vs conclusion
 - The authors did acknowledge the limitations of their study in the discussion, including the small sample size, standardized footwear, short runs, and lack of surface stiffness measurement. However, the language of their conclusion does not reflect these limitations. Instead, the tone of the abstract and results summary is more absolute than the evidence supports.
- Overall
 - While the conclusion reflects the statistical findings well, it is phrased with more certainty than the evidence supports
 - The combination of having a small sample size, high variability, averaging, and possible unmeasured factors means that the study cannot conclude that surface type as no impact on tibial acceleration or shock attenuation

Other:

- In the discussion, the authors claim leg stiffness adjustments may explain the lack of differences across surfaces; however this was not measured. This explanation remains speculative rather than definitive. On a similar note, the discussion also introduces ideas about fatigue effects, but these were also not measured in the experiment. Bringing these unmeasured variables into the interpretation of the data could possibly extend beyond what the current data supports; could be seen as misleading (?)