Bad Stats Depletes the Public Purse

Paul Marchant PhD, CStat.

Leighton Hall, Leeds Beckett University, United Kingdom LS1 3HE
p.marchant@leedsbeckett.ac.uk
&
Clinical and Population Studies Department,
Faculty of Medicine and Health, University of Leeds, United Kingdom LS2 9JT

RadStats Conference 2020 London

Outline

There is a lot of bad science/statistics around! … seemingly especially in softer science areas … e.g. crime, transport … & where selling is involved.

As a consequences money and resources can go to ineffective 'solutions', rather than things which might work.

The examples are mostly from investigations attempting to find clear evidence for claims of public safety benefit from increasing public lighting, but examples I suspect may be found in many areas.

Some Problems

- Given some data & a computer it is pretty easy to get a 'sciencey' looking result out. ... and if you don't like the look of that 'result' ... try tinkering to get something that is more pleasing. ☺

- I suspect that much nonsense gets through because those who should check and stop it, editors / reviewers of the journals publishing the stuff, do not have the statistical grasp to do so. Also, those receiving bad work being used to get buy-in from governments, public servants, don't have the skills / time. Some of those promoting schemes are perhaps more interested in sales than scientific rigour, e.g. professional bodies associated with an industry. …. 'Marketing under the banner of science' .

- The following disclaimer of a pitch for public money hardly inspires confidence.

  'This work was performed with due care and professional standards … Any representation, statement opinion or advice, expressed or implied … is made in good faith but (*we*) are not liable (whether by reason of negligence, lack of care or otherwise) to any person for any loss whatsoever … in relation to that person taking or not taking … action in respect of any representation, statement or advice referred to here.'

- I suspect that there is plenty of material in many areas, e.g. transport and criminology and other softer areas, where statistics is widely used but seemingly poorly understood. If any one is looking for applied stats projects, these could be used to do 'post-publication review'.

Some solutions?

- More independent investigations of claims.

- It would be of benefit if the incremental roll out of large-scale public projects were to be done as rigorous scientific experiments, with clear protocols and pre-defined measures of success, so that any claimed benefits could be checked as the implementation proceeds. By this means any programme could be stopped or adjusted if it was not delivering its objectives or unwelcome side-effects become apparent.

- Currently much money and resources may be being spent on false promises, rather than interventions which might actually work, so it would be good to test claims with good science.

- More success with policy implementations might arise, if advocates who stand to gain financially by a proposal's acceptance, share some of the financial risk of implementation failure. Doing so might encourage advocates to be careful to only adduce sound scientific evidence for any proposal to spend the public's money.

- Perhaps there should be 'Government Departments of Falsification' to really probe claims …. name after Karl Popper? ;-)   … plenty of jobs for statisticians!

My 'Interest': "...Paul Marchant, statistician at Leeds Metropolitan University who argues that statistics used in the Home Office Study 251 could equally be used to show that street lighting actually increases levels of crime. This is an argument which the APPLG, alongside the ILE, would hope to show as utterly absurd. Of course it is worth noting that Paul Marchant is also an astronomer as well as being a statistician, and that this may lead to some bias in his interpretation of the statistics he refers to."

P56 of the March/April 2004 issue of the Lighting Journal, the magazine of the Institution of Lighting Engineers (ILE).

APPLG=The All-Party Parliamentary Lighting Group, ILE=The Institution of Lighting Engineers. (Now the ILP, P=Professionals)

Update: I actually don't do much astronomy these days, although I worry about the impact of 'turning night into day' on ecology. My concerns have shifted to the effect of 'bad science' generally.

I am not proposing to switch off all lights, just that care is needed for things entered into the environment ... & things done for a reason need to work!

As a taster:

One study purports to be an RCT but does not operate on differencing the treatment group from the control!

The other study recklessly associates correlation with causation and also uses a bit of 'unconventional' statistical technique to generate a very impressive, but illusory 'R^2' of 0.99.

Both are used to encourage spending lots of public money (at least hundred of millions).

In both cases no data is made available. In other cases it is not obvious that data is available.

Research integrity is crucial to all science.

- A failure of research integrity will tend to cause misleading results to infect the scientific record and so mislead 'policy'.

- There are many causes of the failure of research integrity, stemming from the vast range of human failings.

- The UK Parliament (Commons) Science and Technology Committee, held a long inquiry into Research Integrity in response to the 'Replication Crisis'. The Report (HC350) was published July 2018. https://www.parliament.uk/business/committees/committees-a-z/commons-select/science-and-technology-committee/inquiries/parliament-2017/research-integrity-17-19/

- My contribution(s) to the inquiry are under RES0042 where I mention my 2017 open access IJSL paper 'Why lighting claims may well be wrong'. This paper discusses good scientific practice such as transparency, openness and completeness. It also discusses some statistical issues that need to be considered.

Steps towards getting trustworthy answers in research

- A, Say clearly and precisely what is to be done before it is done.
- B, Say clearly and precisely what has been done after it has been done. Hopefully, this will be very close to A, and with any differences stated and with the reasons for the changes explained. This write-up should happen irrespective of whether the results are pleasing or not.
- C, Be helpful to others wishing to check or extend research findings, for example by providing the data, with its explanation, in a willing and co-operative fashion. Those using the data, of course, should also respect points A, B and C.

- If the above points, A, B and C are not respected in a work, its results should be given little credence. Science must be an open, transparent, non-exclusive process.

The design of a study is vital

- Design is key, especially if we are to do an observational study, rather than a randomised experiment.

- It is important that a study compares like with like, if the aim is to show that the presence or quantity of some agency (e.g. the nature of lighting) is causing a change in something else (number of crimes or road collisions). The design should try to eliminate alternative possible explanations ('confounders') of any result. RCTs can ensure such a valid comparison.

- Observational studies (i.e. no randomised allocation) require assumptions in order to associate correlation with causation.

A Lighting 'Trial' in New York City social housing, by CrimeLabNY

- Two reports from the one study! First one was put boldly on the US university website 2017 but which disappeared, to be replaced by another (2019) (less comprehensive). New one uses LASSO rather than a 'by hand' model-development. Nothing is said about earlier publicised report.

- Done in conjunction with NYC authorities & the police.

- 'Developments' (Social Housing) allocated to 'new lights or not' randomly, stratified by number of outdoor crimes in the previous 2 years. (Very Good!). Amount of lighting varies randomly in the Treatment group (n=39). The Control group has no change of its lighting (n=38). But no protocol for the study is visible. (Bad)

- A host of other variables added to the mix in 1st report, willy-nilly it seems, This is unexpected! Such bad practice easily detects false positives, if one stops at a point of one's own choosing. (This 'p-hacking' is mentioned in the UK Parliament report.). LASSO is used in 2nd report.

- It is not clear that 'overdispersion' is adequately taken into account. (Crime events are not independent). Overdispersion will cause underestimation of uncertainty, so Confidence Intervals produced are too narrow.

- There is no evidence presented that the models fit the data adequately, as could be provided by residual checks.

- Insufficient clear detail is given in the 1st amateurish report & even less in 2nd . There are discrepencies between the two.

- As a result of the claims hundreds of millions dollars are likely to be spent on something that is unlikely to do what it is supposed to.

From the first report issued … showing a dubious discontinuous slope …. from a quadratic model!
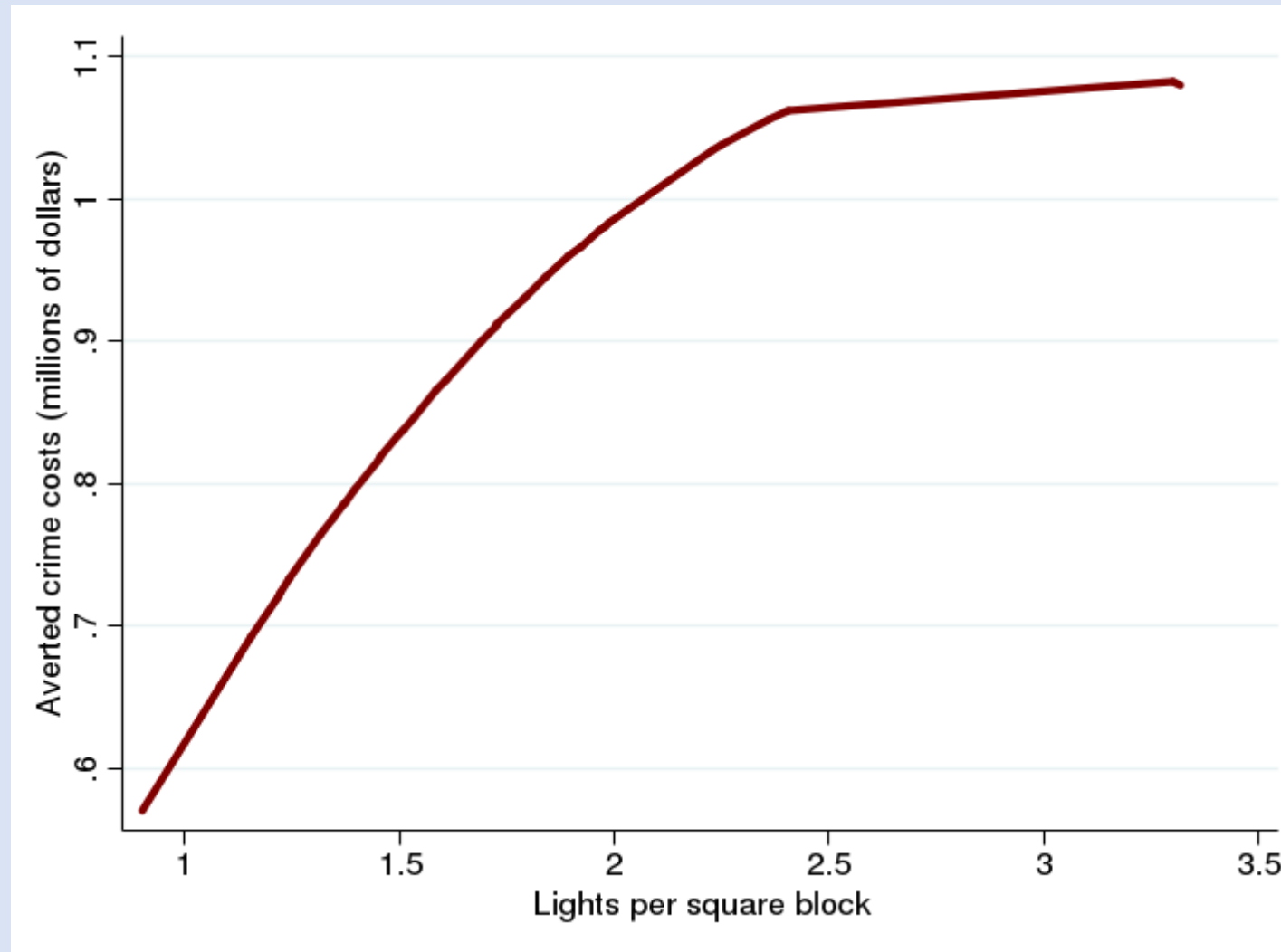
However NOT analysed as a RCT!

The control group is not taken into account in the result! They should analyse the treatment and control groups together, so the uncertainty from both groups is combined. (As a consequence any 'signal', if present, is harder to detect above the increased 'noise'.)

Incorporating control group information directly to make the comparison
1st Issue:

Using 'index' crime at night time, as the example, an estimate of the proper comparison result, can be found, using the statistics in Tables1 & 4 of the 1st report:

Treatment estimated coefficient = -0.48 with z-statistic = -2.59, so the estimate of the 'standard error' (SE), (SE is the uncertainty in the estimate) = -0.48 / -2.59 = 0.1853.

Control estimated coefficient = -0.12 with z-statistic = -1.55, so the estimate of its SE = -0.12 / -1.55 = 0.0774

Therefore, differencing the 'effects' & combining the SEs, the z-statistic for the difference.

= (-0.48- (-0.12)) / $\sqrt{(0.1853^2 + 0.0774^2)}$ = -1.79, so not statistically significant at the conventional 5% level (z = 1.96).

Similarly for second report

Incorporating control group information directly to make the comparison

2nd Issue:

Using 'index' crime at night time, as the example, an estimate of the proper comparison result, can be found, using the statistics in Table 2 of the report:

Treatment estimated coefficient = -0.89. The standard error (SE) = 0.34 (gives z = 2.62)

Control estimated coefficient = -0.15 with SE = 0.35

Therefore, differencing the 'effects' & combining the SEs, the z-statistic for the difference.

= (-0.89- (-0.15)) / √(0.34^2 + 0.35^2) = -1.52, so not statistically significant at the conventional 5% level (z = 1.96)

A similar result to the 1st issue of report.

I suspect that if overdispersion were adequately considered the statistical evidence for any effect would be reduced further.

'LANTERNS' project. An evaluation of the effect of **Reduced** Street lighting but ….. also added change to white light.

- A big project: data from 62 of the 174 England & Wales local authorities, cost £414,315.32. A longitudinal study… At the end of the study period, in 2013, around 40,000km of road had experienced lighting changes. There were 859,935 road crashes in the dataset.

- The protocol is on the National Institute of Health Research Public Health Research (NIHR PHR) website along with the report. (Good!). The NIHR PHR project number is 03110

- No change to public safety was detected, whatever was done to streetlighting (Switch-off, Dimming, Part-night lighting or Changing to white light) in this longitudinal study.

- 'Undeclared variation of protocol' occurred: Change to white light was examined too.

- I don't doubt the null result, but I would like to have the data and apply my (superior!) multilevel method. However, the data of positions and changes to the streetlights (public assets) is apparently confidential! Nothing is said about this in the protocol! I mention this lack of 'openness' in my Research Integrity submission, RES0042.

- Their Cost Benefit Analysis only examines Part Night Lighting & shows PNL to be beneficial. But is silent on the cost dis-benefit of changing to white light.

- Little publicity has been given to the fact that £billions spent on changing to white light, yet there is no evidence that it has enhanced public safety, despite prior claims that the change would.

LANTERNS results: No effect from any change.

Road Collision
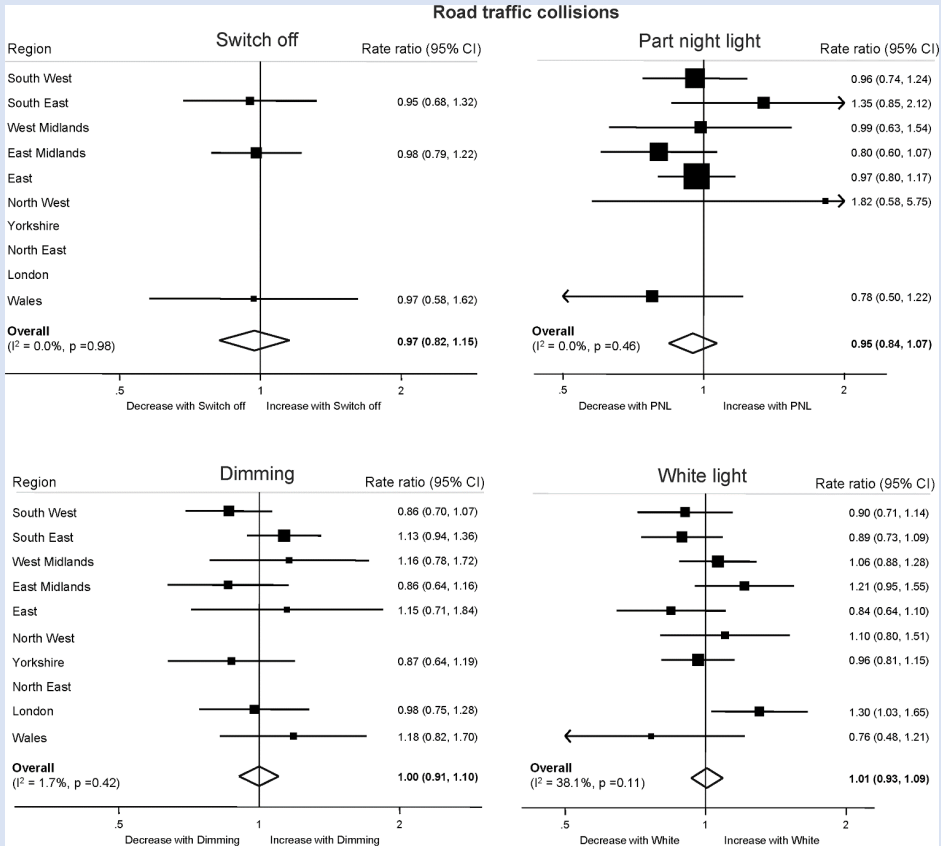
Crime results
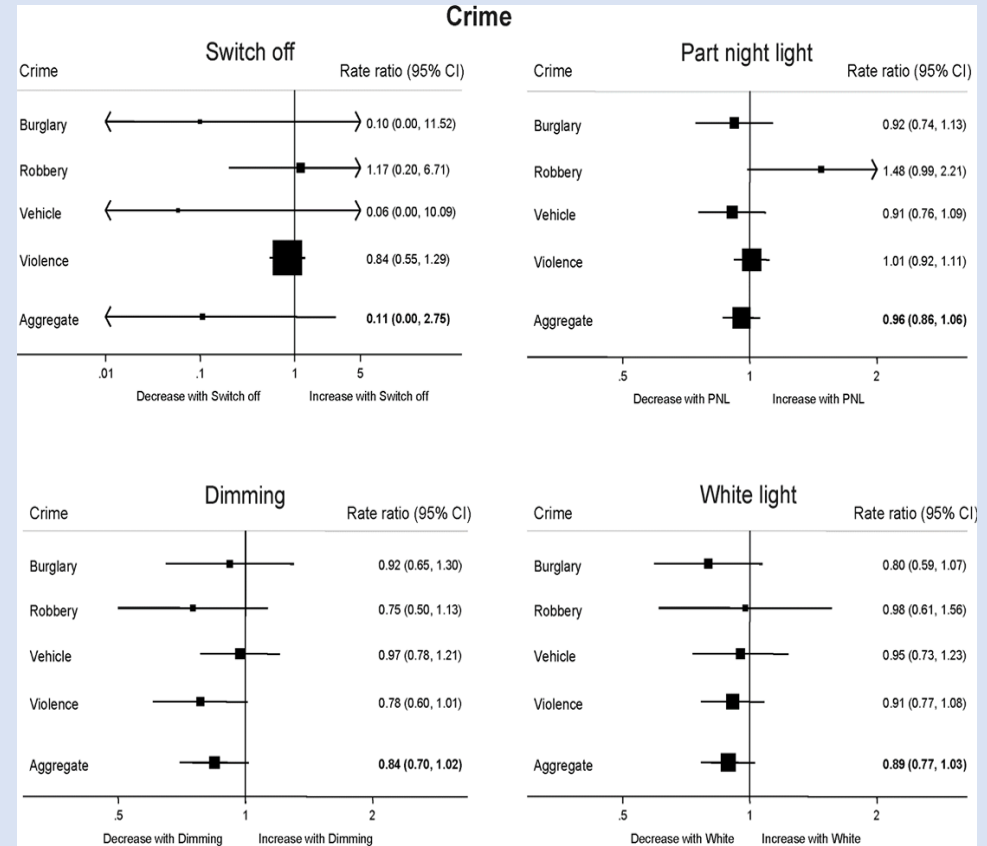
Correlation of itself does not imply Causation

Road crash cross-sectional studies. Jackett & Frith (2013), is like Hargroves & Scott (1979), and both claim their study shows that brighter lighting improves safety.

The studies look at: relationship between, night to day crash-ratio and luminance (cd/m^2) on a sample of roads by standard Poisson regression (where the 'link function' is logarithm).

Finds the night to day crash-ratio is much reduced with higher luminance.

However safety may vary considerably between roads night and day, nothing to do with lighting, but just correlated with it, e.g. traffic flows, types of vehicle and driver. (Busier roads tend to be brighter.)

Longitudinal studies are more reliable, as these compare crash-ratios on the same road before and after when lighting is changed & so indicate the effect of changed lighting directly for these roads. Use a large sample of roads with (ideally differently) changed lighting (like 'LANTERNS').

Note, both x-sectional studies (2013 & 1979) would not have been included in the Beyer & Ker (2010) Cochrane Systematic Review on 'Street lighting for preventing road traffic injuries', because of the lack of 'control'. The review, for good reason, only allowed RCT, Quasi RCT, & Controlled Before After (CBA) studies. (However, this review is blind to Publication Bias & RTM).

Estimates of the parameters from the Jackett and Frith models for night to day crash ratio. Just 152 road section totalling 270km.

Summary results of three models using the Poisson Multiplicative Model to predict the number of night time crashes.

		Independent variables				
Model No.	Constant term (a)	L_{avg}, Average Luminance	TI, Threshold Increment	Uo, Overall Uniformity	Ul, Longitudinal Uniformity	Colour (White=1)
1	−0.84	−.038**	1.08*	0.07	−0.08	0.35*
2	−0.81	−0.38**	0.95*			
3	−0.62	−0.44**				

Notes: The number of * indicates the significance of the parameter. * = two standard errors (significant at p<=0.05), ** = three+ standard errors (highly significant)

Problems with Jackett and Frith (2013)

- It is unclear how the road segments entered into the study.

- No descriptive statistics of any of the data.

- Data is modified. *'Some sites were subsequently shortened, subdivided or deleted to improve homogeneity'*. No statement of how much of this ad-hoc manipulation was done. No sensitivity analysis of the effect of this cavalier action is presented. *'Improving homogeneity'* will reduce the estimate of uncertainty in the results and therefore make finding 'statistical significance' more likely.

- The two stat. sig. coefficients, 'Threshold Increment' (glare) and 'White Light'. These show a <u>worsening</u> of crashes by a factor:
    - of nearly 3 for a Threshold Increment increase of 1
    - and 1.42 for White Light …. but this is not taken further. Both terms simply removed, so as to end with model3, containing <u>only</u> Average Luminance.

- No evidence is given that (even) model1 fits the data, as no model checking is referred to.

- No protocol is visible, no data is offered, none given when requested, no declaration of interest statement is made.

- It is an example of bad science 'with legs', as the paper is used in a pitch for a A$ billion relighting programme. Even the Royal Society of New Zealand have cited it in a report.

- The claim is further 'enhanced' by producing grouped data plots (ignoring uncertainty/variation in both X & Y) and putting a simple least squares regression line through the plot!

Cosmetic redraft: Grouping data, then fitting and putting a line through it! Other such plots are produced and used for further analysis.

All reported crashes

$y = 0.53e^{-0.43x}$

$R^2 = 0.99$
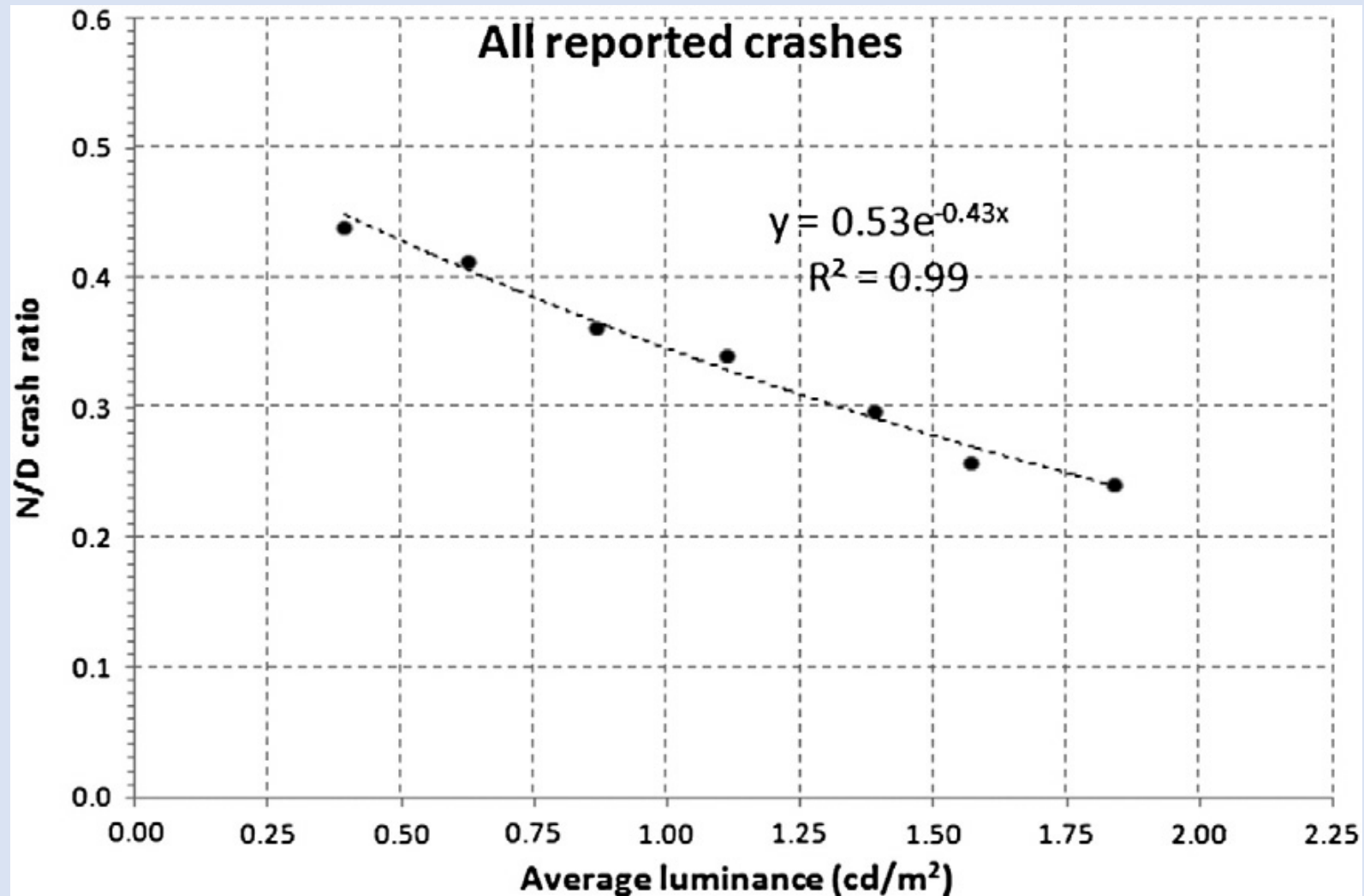
A graphical example of the spurious effect of grouping and averaging data using data from Hargroves & Scott (1979) which does give its data.

- The dark to day crash ratio plotted against the mean group luminance values for mean and individual data.

- Left hand graph is based on the group means, the right hand one is on the individual data points

- LH Using the mean gives an impressive squared correlation coefficient, $R^2 = 0.936$

- RH Using the individual 89 data points gives $R^2 = 0.062$ and a totally different, more realistic, view!

- Note, it is far from ideal to create the night to day crash ratio and run standard regressions, as is done on the right. The correct way to proceed is to use a model in which the count nature of the response is properly treated and that respects Generalised Linear Model assumptions.

-
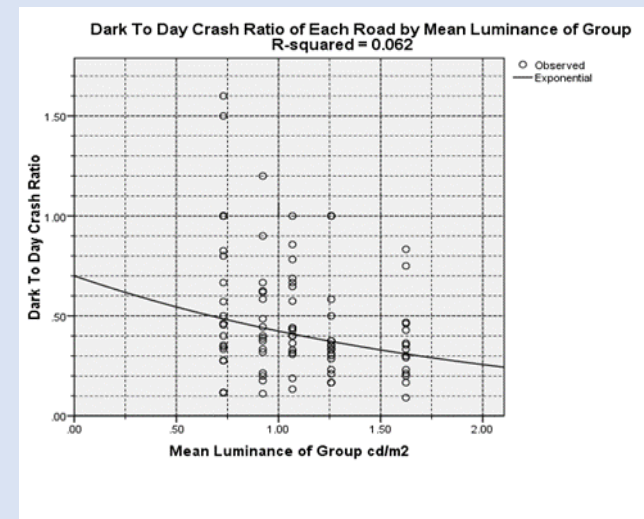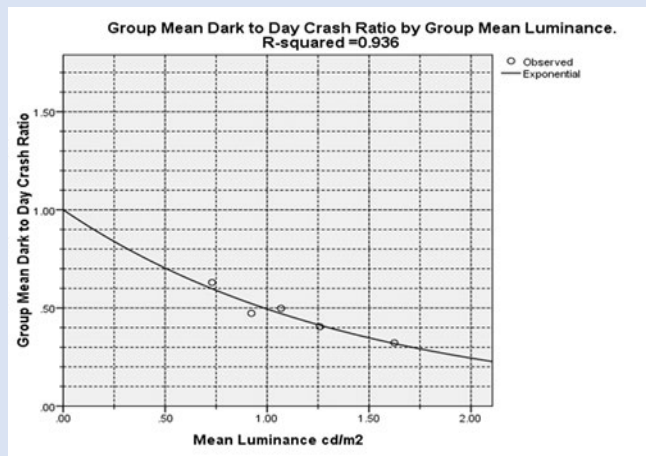
No liability for consequences of reports

- The Institute of Public Works Engineering Australasia report, titled 'SLSC Roadmap: Smart Lighting Smart Controls' uses the JF paper as key evidence in a pitch to obtain A$1+ billion to increase lighting on Australasia's roads.

Disclaimer

This work was performed with due care and in accordance with professional standards. However, the views expressed in the document are solely those of IPWEA and its advisers, Strategic Lighting Partners and Next Energy. Any representation, statement, opinion or advice, expressed or implied in this publication is made in good faith but on the basis that IPWEA, Strategic Lighting Partners and Next Energy are not liable (whether by reason of negligence, lack of care or otherwise) to any person for any damage or loss whatsoever which has occurred or may occur in relation to that person taking or not taking (as the case may be) action in respect of any representation, statement or advice referred to here.

- Another UK transport study costing £1 million (nothing to do with lighting)

…. lead consultant…. assumes no responsibility to any other party in respect of or arising out of or in connection with this document and/or its contents. …. The information or guidance in this document (including third party information, products and services) is provided by DfT on an 'as is' basis, without any representation or endorsement made and without warranty of any kind whether express or implied.

Conclusion

- I have never seen good scientific evidence that public lighting does much to enhance public safety. I have seen a lot of bad & 'spun' stuff!

- Increasing public lighting may enhance public safety, but not by very much I suspect (…or may also make matters somewhat worse, for all we really know.) When lighting is changed, it needs to be implemented so that its effects can be assessed scientifically.

- Scientific standards in lighting research tend to be, so far, very poor indeed and much good scientific practice ought to be adopted.

- Challenging shoddy work may (hopefully) assist in driving the improvement that is desperately needed.

- Of course, similar bad science issues are likely to occurring in other fields of research. These will also mislead.

References

- Beyer F R & Ker K (2010) Street lighting for preventing road traffic injuries. The Cochrane Library https://www.cochrane.org/

- Chalfin A, Hansen B, Lerner J & Parker L (2019) Reducing Crime through Environmental Design: Evidence from a Randomized Experiment of Street Lighting in New York City https://urbanlabs.uchicago.edu/labs/crime-new-york

- Hargroves R A & Scott P P (1979) Measurements of Road Lighting and Accidents, Public Lighting pp213-221, December

- The Institute of Public Works Engineering Australasia, SLSC Roadmap: Smart Lighting Smart Controls 2016/17, downloadable from http://www.slsc.org.au/slsc-publications/slsc-roadmap

- Jackett M & Frith W (2013) Quantifying the impact of road lighting on road safety — A New Zealand Study IATSS Research 36 (International Association of Traffic and Safety Sciences) pp 139-145 https://www.sciencedirect.com/science/article/pii/S0386111212000325?via%3Dihub

- Marchant P R, (2017) Why Lighting Claims Might Well Be Wrong, International Journal of Sustainable Lighting 19 69-74 https://doi.org/10.26607/ijsl.v19i1.71   http://www.lightingjournal.org/index.php/path/article/view/71/79

- Marchant P, (2019) Do brighter whiter street lights improve road safety? Significance Vol 16, Issue 5 pp8-9, a joint publication of the Royal Statistical Society and the American Statistical Association https://doi.org/10.1111/j.1740-9713.2019.01313.x

- Perkins, C, Steinbach, R, Tompson, L, Green, J, Johnson, S, Grundy, C, Edwards, P et al. (2015) What is the effect of reduced street lighting on crime and road traffic injuries at night? A mixed-methods study. Available from: https://www.journalslibrary.nihr.ac.uk/programmes/phr/11300402/#/.

- The Royal Society of New Zealand, (November 2018) 'Impacts of Artificial Blue Light on Health and the Environment, Evidence Summary', downloadable from https://royalsociety.org.nz/major-issues-and-projects/blue-light-aotearoa/

- Scott P P, (1980) The Relationship Between Road Lighting Quality and Accident Frequency, in: TRRL laboratory report LR 929, https://trl.co.uk/reports/LR929

- Steinbach R, Perkins C, Tompson L, Johnson S, Armstrong B, Green J, Grundy C, Wilkinson P, Edwards P. (2015) The effect of reduced street lighting on road casualties and crime in England and Wales: controlled interrupted time series analysis. Journal of Epidemiology and Community Health 2015 69(11):1118-1124. Available from: https://jech.bmj.com/content/69/11/1118

- UK Parliament (Commons) Science and Technology Committee, (July 2018) Report into Research Integrity, HC350. https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/350/350.pdf The inquiry website is https://www.parliament.uk/business/committees/committees-a-z/commons-select/science-and-technology-committee/inquiries/parliament-2017/research-integrity-17-19/ The Committee page is https://www.parliament.uk/business/committees/committees-a-z/commons-select/science-and-technology-committee/

- World Health Organization (2018) Road Traffic Injuries Factsheet https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries

Thanks for listening!

BTW A question for you! … has anyone used the Centre for Open Science https://cos.io/ or other system to record and constrain flexibility in procedure for an investigation?

Any thoughts on talk???

Paul Marchant

p.marchant@leedsbeckett.ac.uk