Discussion of

“Testing an Elaborate Theory of a Causal Hypothesis”

by Dylan Small

Peter Bühlmann

Seminar for Statistics, ETH Zürich

Excellent talk, Dylan!

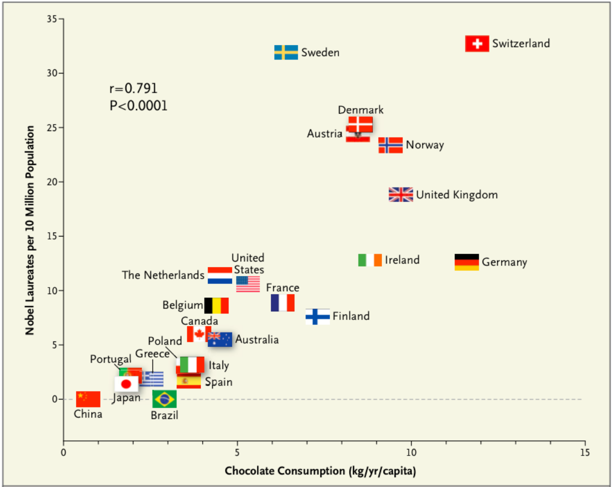
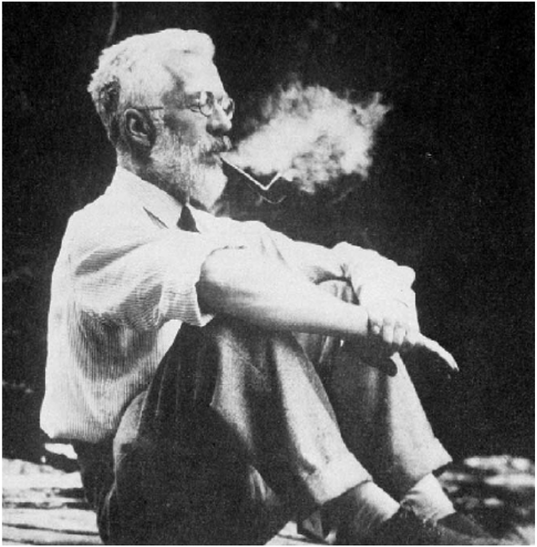
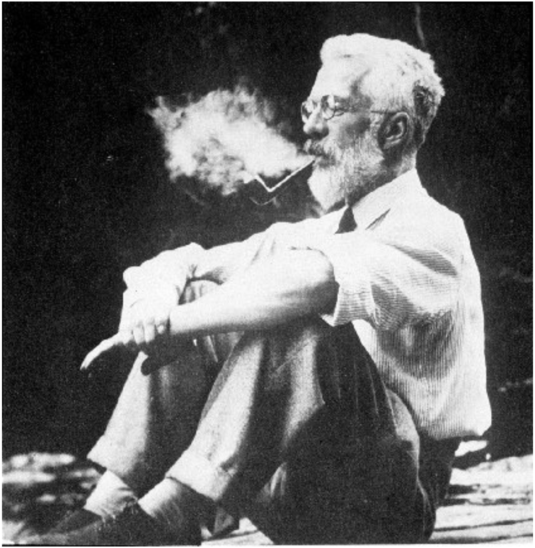


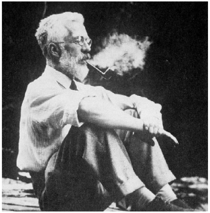
Figure 1. **Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.**

F. H. Messerli: *Chocolate Consumption, Cognitive Function, and Nobel Laureates*, N Engl J Med 2012

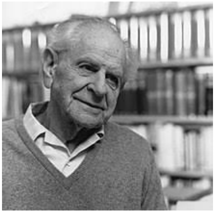




Dylan Small's talk is based on a beautiful paper (Karmakar & Small, 2018) connecting to great scientists


Fisher


Cochran


Popper

with some German from Popper: "Grad der Bewährung" translated by Popper to "degree of corroboration" (correcting the inattentive first translation "degree of confirmation")

the broader context of the talk:
Dylan Small is very careful about the process of collecting evidence – rather than quickly claiming confirmatory results...
$\leadsto$ much desired for improving "data-driven science"

(Slightly) more technical

Setting up a multi-phase elaborate theory

– and the issue of ordering

An elaborate theory is a set of ordered statistical tests

methodological trick: near independence of evidence factors! Cute!

Is there a "recipe" to set-up an elaborate theory?

Or is this "the art of statistical science"?

Does the ordering of the tests matter?

The problem with hidden confounding

addressed by elegant sensitivity analysis with freq. err. control!

general nice idea of sensitivity analysis (cf. Rosenbaum 1987, 2002)

perhaps more realistic than trying to perform "full deconfounding" (Wang & Blei; Cevid, PB & Meinshausen; Shah, Frot, Thanei & Meinshausen; ...)

but, obviously, relying on some assumptions:

here binary treatment: linear logistic treatment assign. model

How about continuous treatments?

$\leadsto$ many more models and model misspec. (for sensitivity analysis)

Robustness against confounding

under no hidden confounding, Fisher's or truncated product aggregation are Bahadur slope optimal

again, under the assumed treatment assignment model

in practice: How much confounding (Γ-value) do we allow for?

$\Gamma\downarrow$	$\overline{P}_{5,\Gamma}$	$\overline{P}_{4,\Gamma}$	$\overline{P}_{3,\Gamma}$	$\overline{P}_{2,\Gamma}$	$\overline{P}_{1,\Gamma}$
1	0.420036	**0.009441**	0.095923	**0.00381**	**0.00007**
1.2	0.470253	0.013512	0.128619	**0.006773**	**0.000263**
1.4	0.512934	0.017814	0.161157	0.010557	**0.000688**
1.6	0.549884	0.022219	0.192914	0.015089	**0.001425**
1.8	0.582428	0.02672	0.223553	0.020268	**0.002525**
2	0.611224	0.031257	0.252909	0.025994	**0.004007**
2.2	0.636902	0.035769	0.280914	0.032177	**0.005867**
2.4	0.659949	0.040228	0.307565	0.038738	**0.008085**
2.6	0.680756	0.044615	0.332889	0.045607	0.010632
2.8	0.699635	0.048916	0.356935	0.052721	0.013472
3	0.716841	0.053123	0.379764	0.060029	0.016569
4	0.784073	0.072632	0.477894	0.098608	0.034756
4.8	0.822295	0.08707	0.541509	0.130282	0.051015
5	0.830333	0.090589	0.555832	0.138152	0.055166
$\Gamma\downarrow$	$P_\Gamma^{5\|5}$	$P_\Gamma^{4\|5}$	$P_\Gamma^{3\|5}$	$P_\Gamma^{2\|5}$	$P_\Gamma^{1\|5}$
1	1	0.193477	**0.017172**	**0.000795**	**0.000002**
1.2	1	0.24579	**0.027005**	**0.001965**	**0.000012**
1.4	1	0.297852	**0.037873**	**0.003864**	**0.000052**
1.6	1	0.348663	**0.049288**	**0.006544**	**0.00016**
1.8	1	1	0.149304	**0.026378**	**0.001114**
2	1	1	0.161532	**0.033879**	**0.002024**
2.2	1	1	0.17212	**0.041805**	**0.003305**
2.4	1	1	0.181238	0.050012	**0.004979**
2.6	1	1	0.191565	0.05838	**0.007052**
2.8	1	1	0.205224	0.066817	**0.009511**
3	1	1	0.219255	0.075254	**0.012336**
4	1	1	0.293328	0.116672	**0.030932**
4.8	1	1	0.354142	0.148886	**0.049496**
5	1	1	0.369251	0.156729	0.054454

Multiple testing

a clever combination of frequentist multiple testing adjustment for partial conjunction hypotheses (cf. Benjamini and Heller, 2008)

the number of tests is small

Is the choice of the adjustment method important? Especially when protecing against strong confounding (large Γ value)?