Controlled experiments

Marie-Abele Bind

STAT 140

Spring 2026

Original title

A week when Florian actually appeared to care about statistics

Quote #1

"You don't really need statistics to see an effect in neuroscience."

Quote #1

"You don't need [no stinking] statistics to see an effect in neuroscience."

Quote #1

"You don't need [no stinking] statistics to see an effect in neuroscience."

Florian, MCB Retreat 2018

Quote #1

"You don't need [no stinking] statistics to see an effect in neuroscience."

Quote #1

"You don't need [no stinking] statistics to see an effect in neuroscience."

Paper submitted to Science Advances (April 2021)

Collective behavior emerges from genetically controlled simple behavioral motifs in zebrafish

Ariel C. Aspiras*[1], Roy Harpaz*[2,3], Sydney Chambule[1], Sierra Tseng[1], Florian Engert[2,3], Mark C. Fishman[1#§] & Armin Bahl[2,3,4§]

1 Harvard Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge 02138, USA.

2 Department of Molecular and Cellular Biology, Harvard University, Cambridge 02138, USA.

3 Center for Brain Science, Harvard University, Cambridge, Massachusetts 02138, USA.

4 Centre for the Advanced Study of Collective Behaviour, University of Konstanz, Konstanz 78464, Germany.

Quote #2

- "The authors repeatedly and systematically misinterpret the p-values calculated for their statistical tests and consequently make a large number of false and misleading claims about the meaning of their results.

Quote #2

- "The authors repeatedly and systematically misinterpret the p-values calculated for their statistical tests and consequently make a large number of false and misleading claims about the meaning of their results.
- The most charitable explanation for this is that the authors are simply ignorant of how to correctly interpret p-values despite their extensive shared expertise of quantitative data analysis. There are, of course, less charitable explanations, and I would therefore suggest that the paper is thoroughly revised to remove any doubts about the authors' intentions here."

Quote #2

- "The authors repeatedly and systematically misinterpret the p-values calculated for their statistical tests and consequently make a large number of false and misleading claims about the meaning of their results.
- The most charitable explanation for this is that the authors are simply ignorant of how to correctly interpret p-values despite their extensive shared expertise of quantitative data analysis. There are, of course, less charitable explanations, and I would therefore suggest that the paper is thoroughly revised to remove any doubts about the authors' intentions here."

Reviewer 1, May-June 2021

Quote #2

- "The authors repeatedly and systematically misinterpret the p-values calculated for their statistical tests and consequently make a large number of false and misleading claims about the meaning of their results.
- The most charitable explanation for this is that the authors are simply ignorant of how to correctly interpret p-values despite their extensive shared expertise of quantitative data analysis. There are, of course, less charitable explanations, and I would therefore suggest that the paper is thoroughly revised to remove any doubts about the authors' intentions here."

Reviewer 1, May-June 2021

Quote #2

- "The authors repeatedly and systematically misinterpret the p-values calculated for their statistical tests and consequently make a large number of false and misleading claims about the meaning of their results.
- The most charitable explanation for this is that the authors are simply ignorant of how to correctly interpret p-values despite their extensive shared expertise of quantitative data analysis. There are, of course, less charitable explanations, and I would therefore suggest that the paper is thoroughly revised to remove any doubts about the authors' intentions here."

Reviewer 1, May-June 2021

Quote #2

- "The authors repeatedly and systematically misinterpret the p-values calculated for their statistical tests and consequently make a large number of false and misleading claims about the meaning of their results.
- The most charitable explanation for this is that the authors are simply ignorant of how to correctly interpret p-values despite their extensive shared expertise of quantitative data analysis. There are, of course, less charitable explanations, and I would therefore suggest that the paper is thoroughly revised to remove any doubts about the authors' intentions here."

Reviewer 1, May-June 2021

The week started...

A ... perspective in defense of ...

Marie-Abèle Bind

QuantBio

October 30^{th} 2019

The week started...

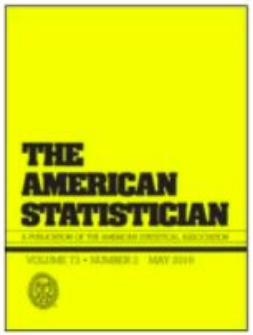
When possible, report a Fisher-exact *P* value and display its underlying null randomization distribution

M.-A. C. Bind[a,1] and D. B. Rubin[b,c]

[a]Department of Statistics, Faculty of Arts and Sciences, Harvard University, Cambridge, MA 02138; [b]Yau Center for Mathematical Sciences, Tsinghua University, Beijing 100084, China; and [c]Department of Statistical Science, Fox School of Business, Temple University, Philadelphia, PA 19122

Edited by Bin Yu, University of California, Berkeley, CA, and approved June 5, 2020 (received for review September 10, 2019)

The week started...



Taylor & Francis

The American Statistician

ISSN: 0003-1305 (Print) 1537-2731 (Online) Journal homepage: https://www.tandfonline.com/loi/utas20

The ASA Statement on *p*-Values: Context, Process, and Purpose

Ronald L. Wasserstein & Nicole A. Lazar

To cite this article: Ronald L. Wasserstein & Nicole A. Lazar (2016) The ASA Statement on *p*-Values: Context, Process, and Purpose, The American Statistician, 70:2, 129-133, DOI: 10.1080/00031305.2016.1154108

To link to this article: https://doi.org/10.1080/00031305.2016.1154108

The week started...

Taylor & Francis

The American Statistician

ISSN: 0003-1305 (Print) 1537-2731 (Online) Journal homepage: https://www.tandfonline.com/loi/utas20

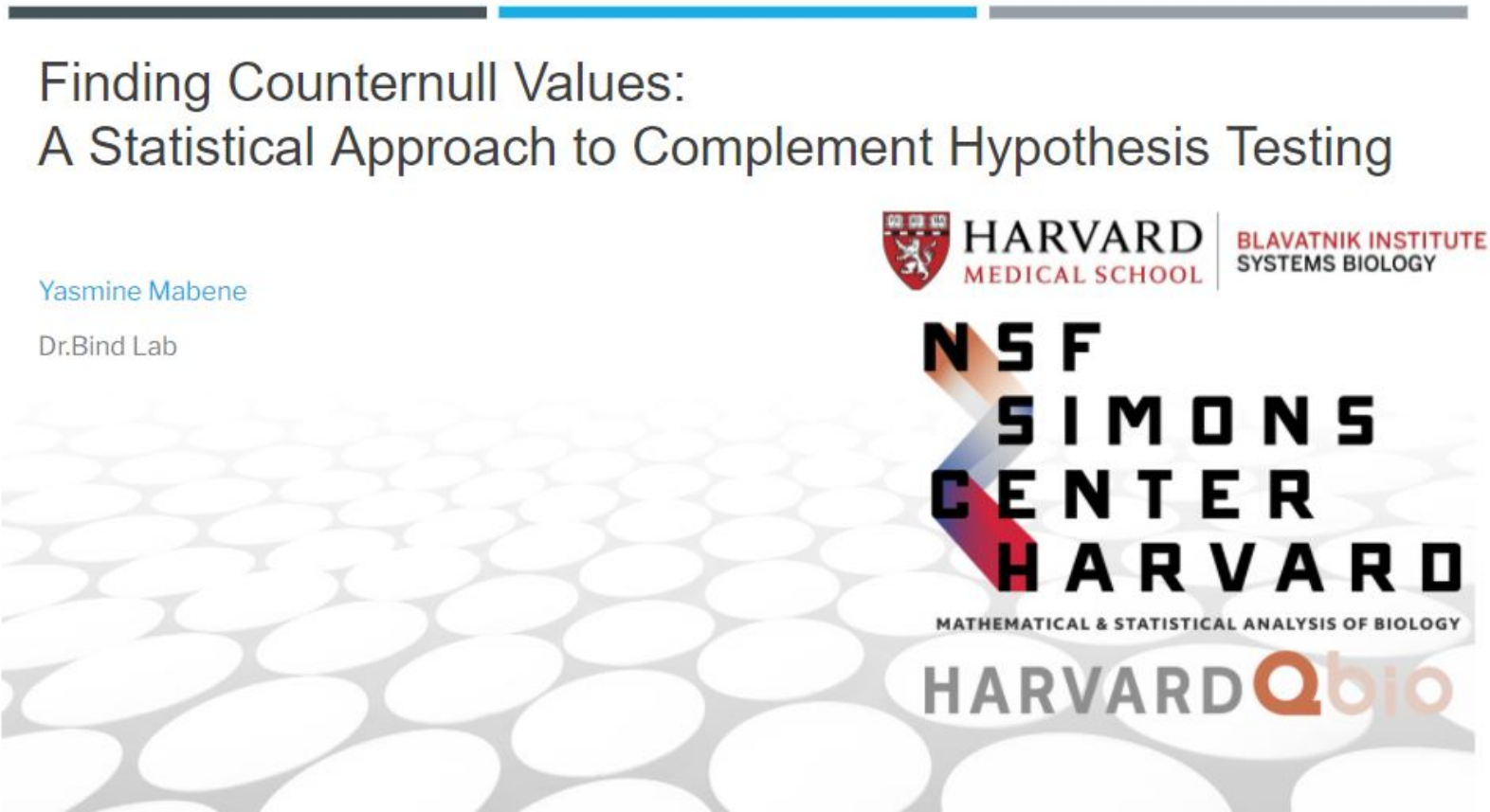
Moving to a World Beyond "$p < 0.05$"

Ronald L. Wasserstein, Allen L. Schirm & Nicole A. Lazar

To cite this article: Ronald L. Wasserstein, Allen L. Schirm & Nicole A. Lazar (2019) Moving to a World Beyond "$p<0.05$", The American Statistician, 73:sup1, 1-19, DOI: 10.1080/00031305.2019.1583913

To link to this article: https://doi.org/10.1080/00031305.2019.1583913

The week started...



Selected response to Reviewer 1

- Thank you for pointing this out.

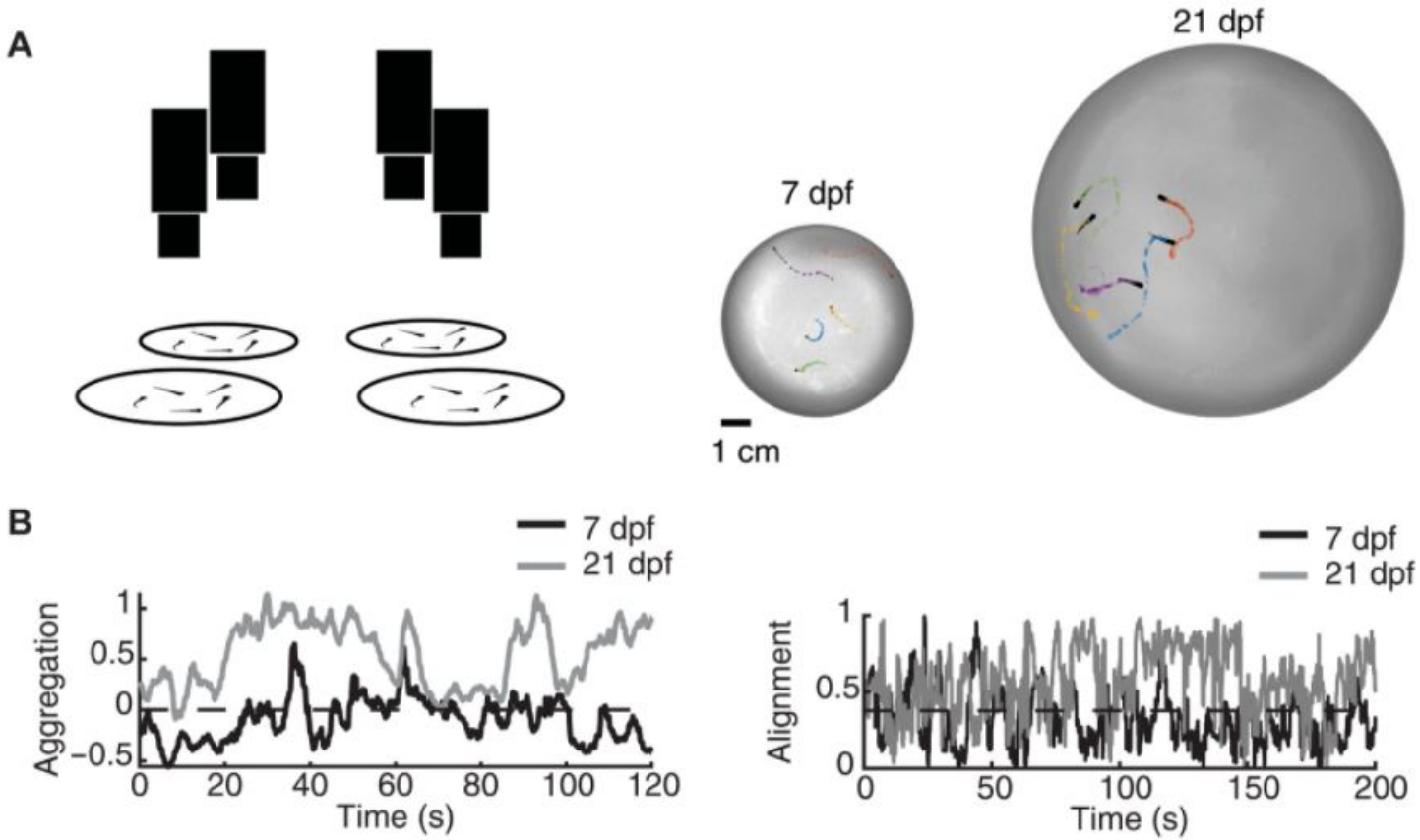
Selected response to Reviewer 1

- We have educated ourselves better around the definition and limitation of the p-value, in part by discussion with our colleague (now co-author) Dr. Bind, and by reading articles such as the American Statistical Association (ASA) Statement on p-values (Wasserstein, 2016), the article entitled "Moving to a world beyond $p<0.05$" (Wasserstein et al., 2019), and our co-author's article on Fisher-exact p-value (Bind and Rubin, 2020).

Selected response to Reviewer 1

- We have revised our empirical claims throughout the manuscript and attempted to distinguish the objective statements from the subjective ones.

Selected behavioral experiments



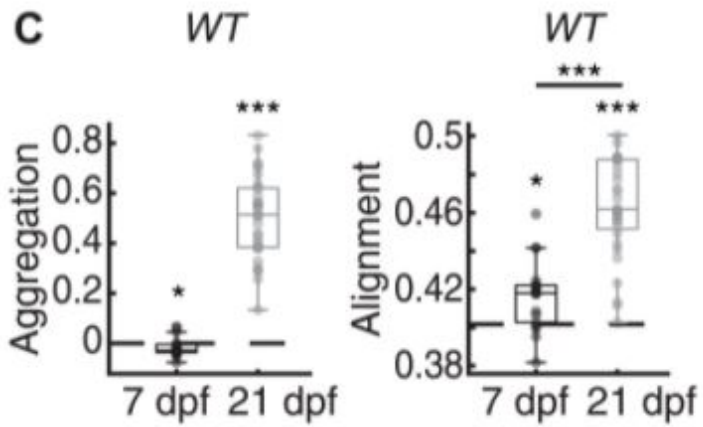
Selected behavioral experiments

Individual and group properties of free-swimming fish

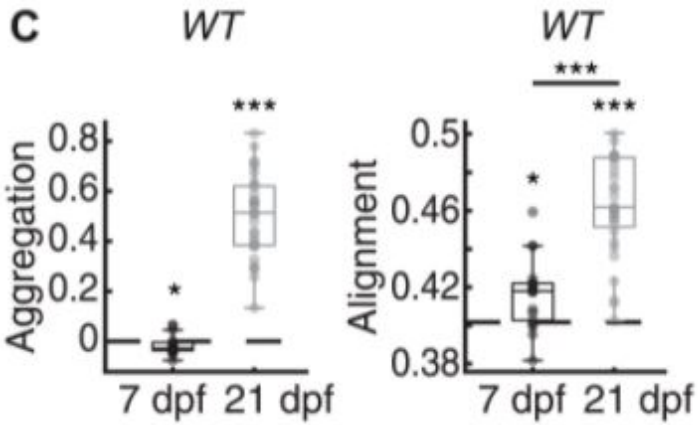
We used the extracted center of mass position of every fish (fish i;$\vec{x}_i$) to calculate the velocity of the fish $\vec{v}_i(t) = [\vec{x}_i(t + dt) - \vec{x}_i(t - dt)]/2dt$, where dt is 1 frame or 0.025 s. The speed of the fish is then $S_i(t) = |\vec{v}_i(t)|$, and the direction of motion is $\vec{d}_i(t) = \vec{v}_i(t)/|\vec{v}_i(t)|$.

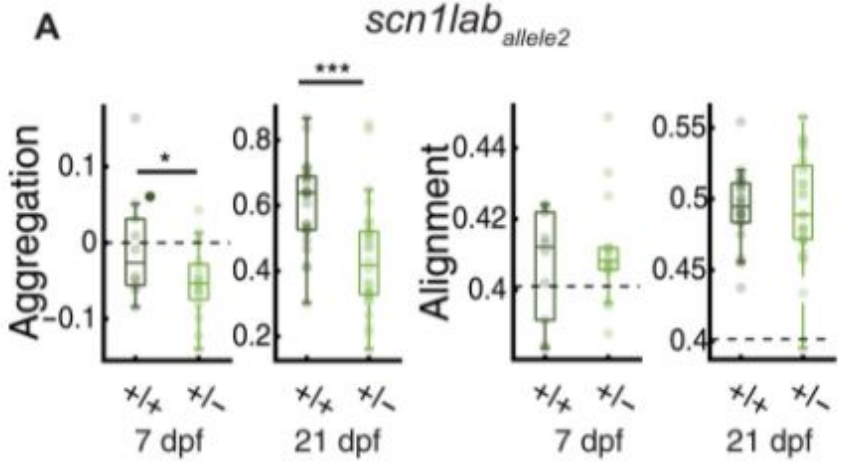
For the group, we calculate a normalized measure of group aggregation: Aggregation $= -\log(NN_1/NN_1^{\text{shuffled}})$, where NN_1 is the average nearest neighbor distance. NN_1^{shuffled} is the same distance calculated from control groups created by shuffling fish between groups such that all fish in a shuffled group were chosen from different real groups. Positive aggregation values mean that real groups are more aggregated than shuffled controls, and 0 means aggregation occurred at random. Group alignment was defined as alignment$(t) = |\sum_i^N \vec{d}_i(t)|/N$, where N is the number of fish in the group, and alignment value is bounded between 0, all fish are pointing in different directions, and 1, all fish swim in the same direction.

Selected empirical results

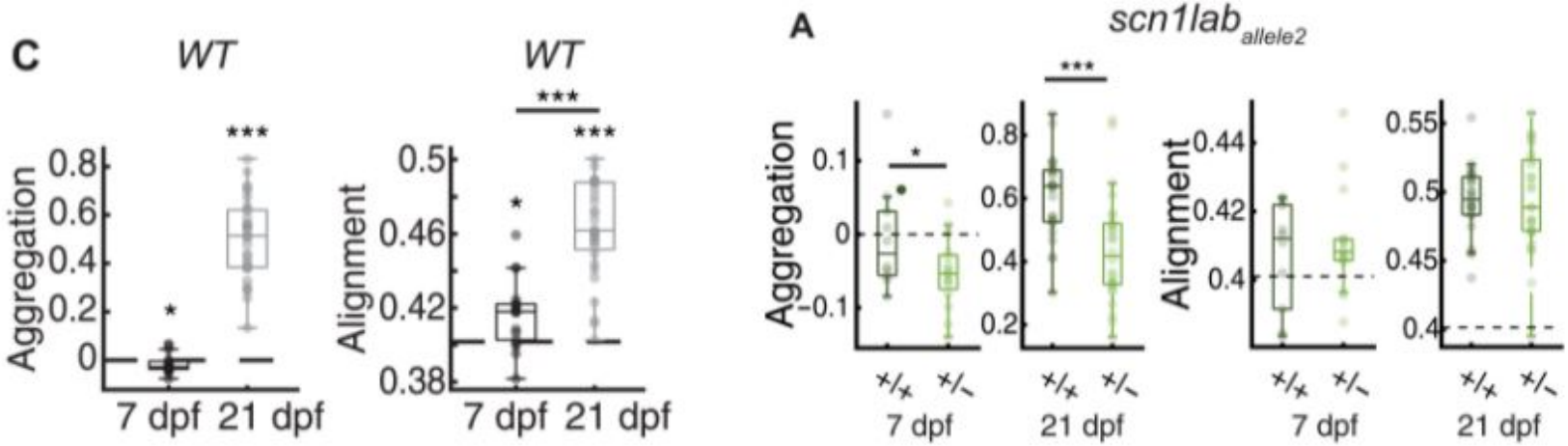


Selected empirical results



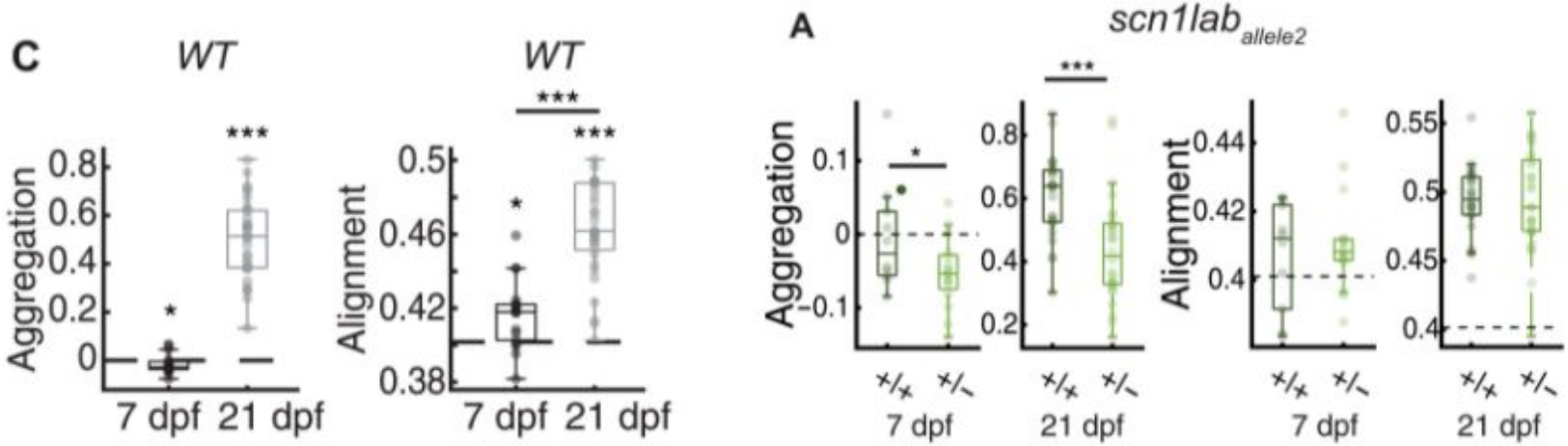


Selected empirical results



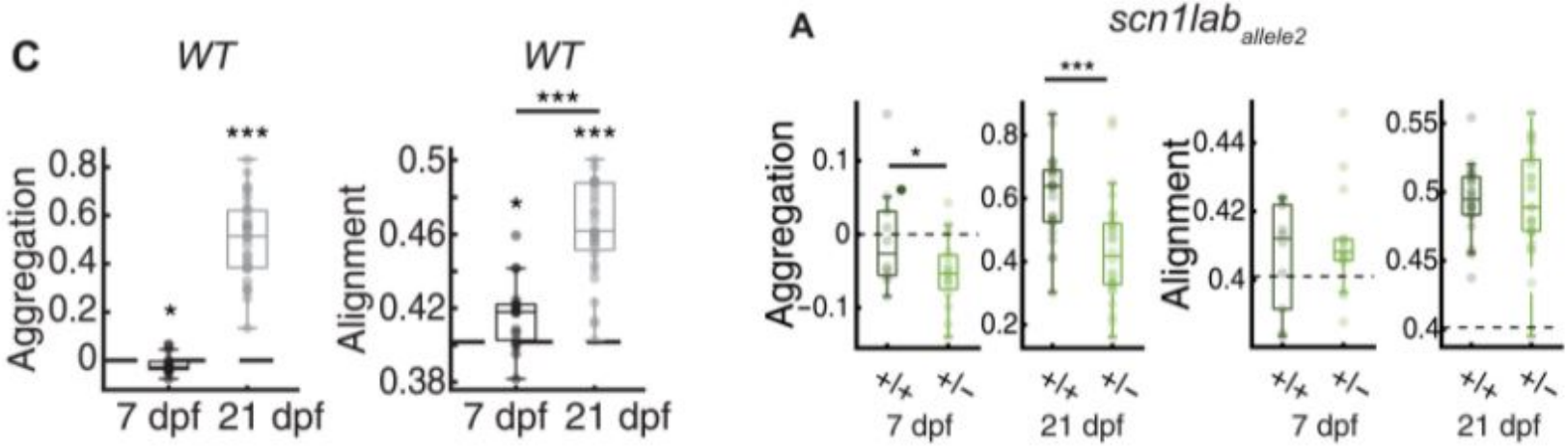
Florian : "We don't really need statistics to see an effect in neuroscience."

Selected empirical results



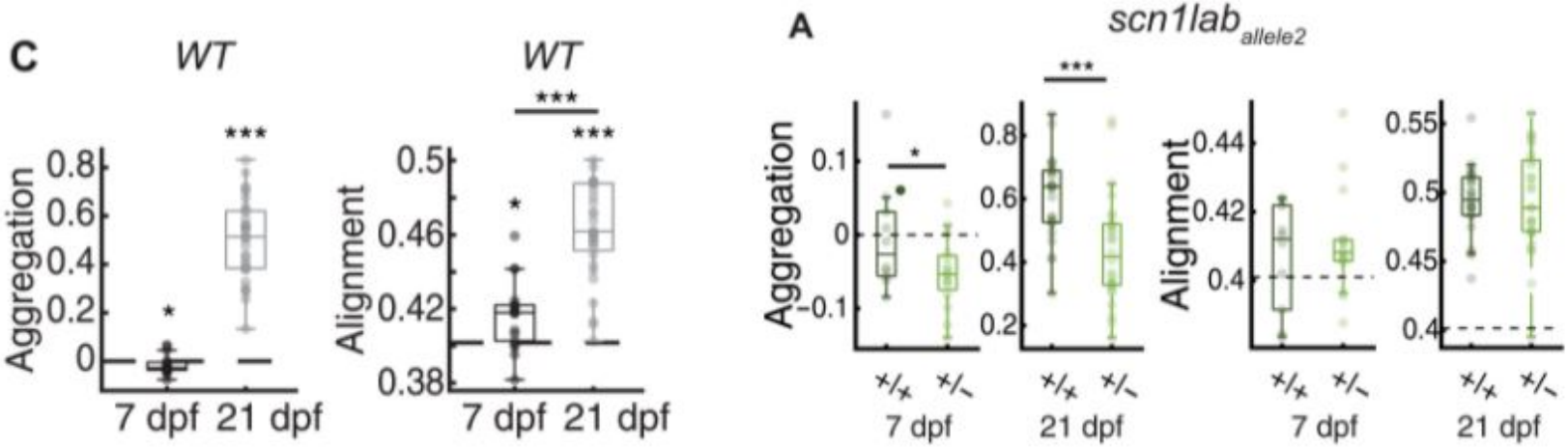
Reviewer #1 : "The authors obscure or simply do not include important statistical information for their results. For example, the authors present many of their results as p-value thresholds, but the actual p-values need to be reported."

Selected empirical results



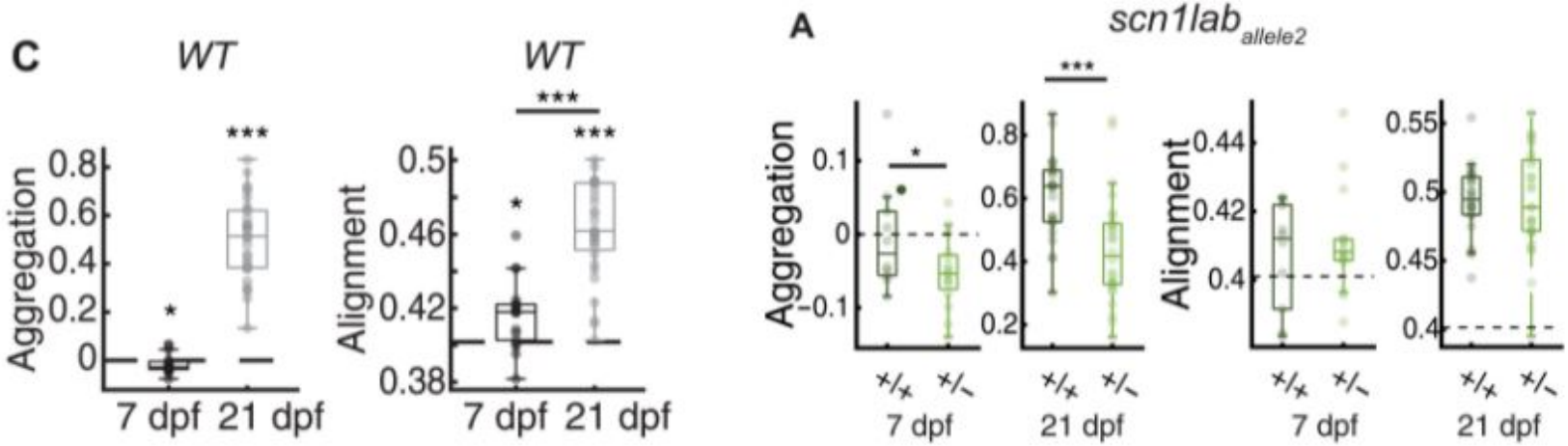
Reviewer #1 : "If the authors wish to claim their results with p-values of 0.063 and 0.065 are "marginally significant" then they must also be willing to claim that their results with p-values <0.05 are marginally *insignificant*."

Selected empirical results



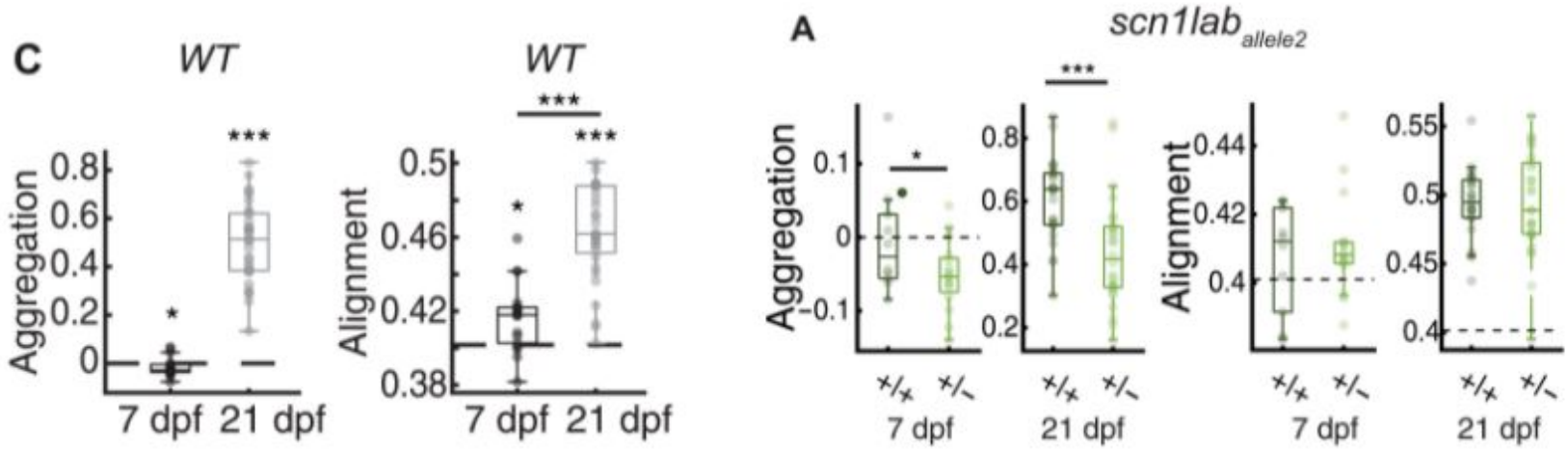
Reviewer #1 : "The authors do not discuss effect sizes in any meaningful way. If the authors do not discuss effect sizes and what they mean biologically (some of which are very small and probably not meaningful) then they are abdicating their central role as scientists. For example, they claim a specific effect is “small but significant” without providing any further explanation of what they mean by this. I assume they mean statistically significant (with a small p-value) but with only a small difference between treatment groups. In which case, the more reasonable interpretation is “statistically detectable, but too small to be biologically important” rather than implying there is a biologically meaningful effect solely due to a statistically significant p-value."

Selected empirical results



Reviewer #1 : "The reasoning for the choice of tests as well as the assumptions of each test and how these assumptions were validated for the data to which the tests were applied are not explained."

Selected empirical results



Reviewer #1 : "The reasoning for the choice of tests as well as the assumptions of each test and how these assumptions were validated for the data to which the tests were applied are not explained. [...]

Were these tests chosen *a priori* or in an *ad-hoc* manner after data collection ? If chosen *ad-hoc* then the results are exploratory, and the p-values calculated for the tests cannot be used to reliably make claims within the null hypothesis significance testing framework as they are subject to a substantially higher rate of false positives. [...] Statistical tests can certainly be used as quantitative descriptions of the data, but the fact that these are exploratory results would then need to be more explicitly stated in the text and the results interpreted more conservatively."

With the help of Roy, statistics and interpretations changed !

Quote #3

"Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise".

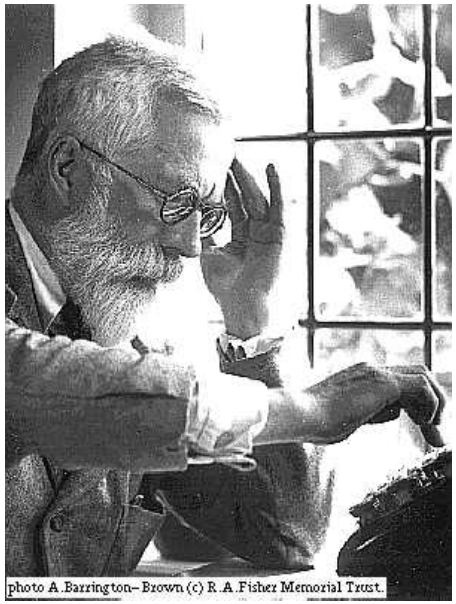
Tukey, 1962

Randomization-based inference

- Calculating a Fisher-exact p-value is superior to the current, more common, practice of calculating an approximating asymptotic (i.e., large sample) p-value.

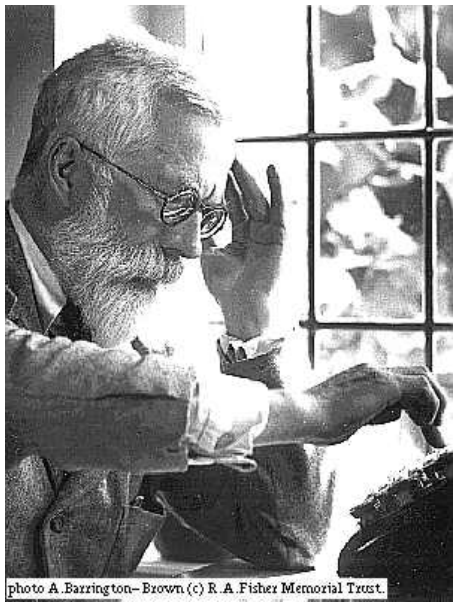
Randomization-based inference

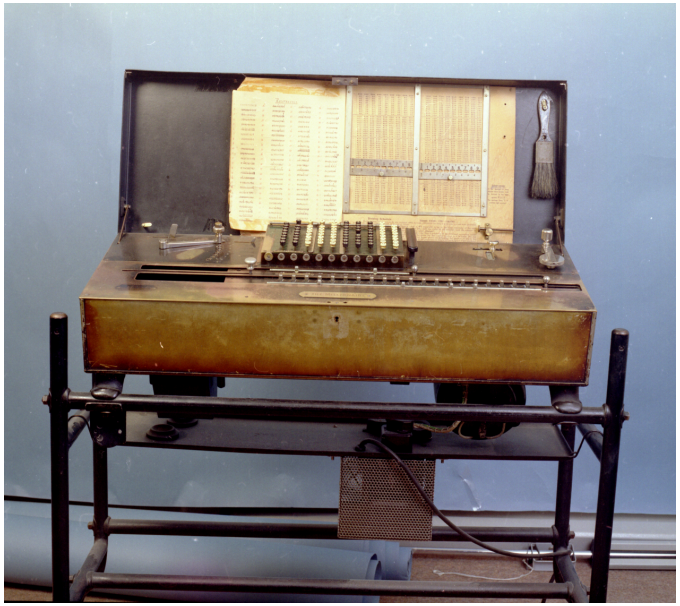
- Calculating a Fisher-exact p-value is superior to the current, more common, practice of calculating an approximating asymptotic (i.e., large sample) p-value.



Randomization-based inference

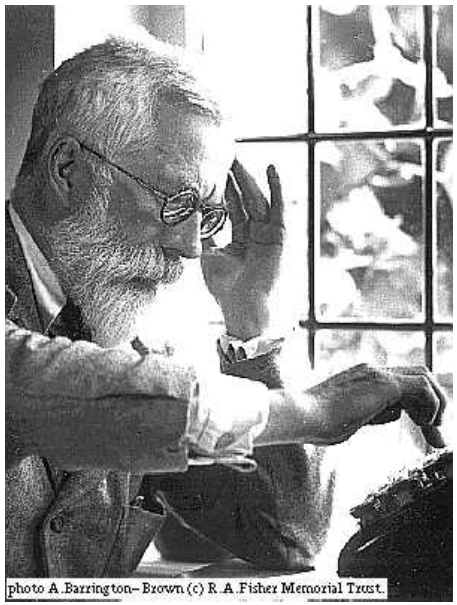
- Calculating a Fisher-exact p-value is superior to the current, more common, practice of calculating an approximating asymptotic (i.e., large sample) p-value.





Randomization-based inference

- Calculating a Fisher-exact p-value is superior to the current, more common, practice of calculating an approximating asymptotic (i.e., large sample) p-value.



Simple randomized experiment

- Consider a simple treatment-control experiment with a completely randomized assignment mechanism.

Simple randomized experiment

- Consider a simple treatment-control experiment with a completely randomized assignment mechanism.
- Consider N units, indexed by i, in an experiment randomized to a binary treatment W, either active (i.e., $W_i=1$) or control (i.e., $W_i=0$).

Simple randomized experiment

- Consider a simple treatment-control experiment with a completely randomized assignment mechanism.
- Consider N units, indexed by i, in an experiment randomized to a binary treatment W, either active (i.e., $W_i=1$) or control (i.e., $W_i=0$).
- An outcome variable Y is measured after exposure to the treatment.

Simple randomized experiment

- Consider a simple treatment-control experiment with a completely randomized assignment mechanism.
- Consider N units, indexed by i, in an experiment randomized to a binary treatment W, either active (i.e., W_i=1) or control (i.e., W_i=0).
- An outcome variable Y is measured after exposure to the treatment.
- No hidden treatment and no interference between units (SUTVA). $Y_i(W_i$=w) is a function of unit i and treatment w.

Simple randomized experiment

- Consider a simple treatment-control experiment with a completely randomized assignment mechanism.
- Consider N units, indexed by i, in an experiment randomized to a binary treatment W, either active (i.e., W_i=1) or control (i.e., W_i=0).
- An outcome variable Y is measured after exposure to the treatment.
- No hidden treatment and no interference between units (SUTVA). $Y_i(W_i$=w) is a function of unit i and treatment w.
- For each unit i=1,..., N :
  - $Y_i(W_i$=1)=value of Y when i is exposed to the active treatment
  - $Y_i(W_i$=0)=value of Y when i is exposed to the control treatment.

Simple randomized experiment

- Consider a simple treatment-control experiment with a completely randomized assignment mechanism.
- Consider N units, indexed by i, in an experiment randomized to a binary treatment W, either active (i.e., W_i=1) or control (i.e., W_i=0).
- An outcome variable Y is measured after exposure to the treatment.
- No hidden treatment and no interference between units (SUTVA). $Y_i(W_i$=w) is a function of unit i and treatment w.
- For each unit i=1,..., N :
  - $Y_i(W_i$=1)=value of Y when i is exposed to the active treatment
  - $Y_i(W_i$=0)=value of Y when i is exposed to the control treatment.
- Because of SUTVA, the science table simplifies to two columns and N rows.

Florian's experiment

- Consider a simple mutant-wildtype zebrafish experiment with an hypothetical completely randomized assignment mechanism.

Florian's experiment

- Consider a simple mutant-wildtype zebrafish experiment with an hypothetical completely randomized assignment mechanism.
- Consider N groups of 5 zebrafish, indexed by i, in an experiment hypothetically randomized to a binary genotype G, either mutant (i.e., G_i=1) or wildtype (i.e., G_i=0).

Florian's experiment

- Consider a simple mutant-wildtype zebrafish experiment with an hypothetical completely randomized assignment mechanism.
- Consider N groups of 5 zebrafish, indexed by i, in an experiment hypothetically randomized to a binary genotype G, either mutant (i.e., G_i=1) or wildtype (i.e., G_i=0).
- An outcome variable Y (e.g., aggregation, alignment at 7 dpf) is measured for mutant and wildtype zebrafish.

Florian's experiment

- Consider a simple mutant-wildtype zebrafish experiment with an hypothetical completely randomized assignment mechanism.
- Consider N groups of 5 zebrafish, indexed by i, in an experiment hypothetically randomized to a binary genotype G, either mutant (i.e., G_i=1) or wildtype (i.e., G_i=0).
- An outcome variable Y (e.g., aggregation, alignment at 7 dpf) is measured for mutant and wildtype zebrafish.
- No hidden mutation and no interference between groups of zebrafish (SUTVA). $Y_i(G_i$=g) is a function of unit i and genotype g.

Florian's experiment

- Consider a simple mutant-wildtype zebrafish experiment with an hypothetical completely randomized assignment mechanism.
- Consider N groups of 5 zebrafish, indexed by i, in an experiment hypothetically randomized to a binary genotype G, either mutant (i.e., G_i=1) or wildtype (i.e., G_i=0).
- An outcome variable Y (e.g., aggregation, alignment at 7 dpf) is measured for mutant and wildtype zebrafish.
- No hidden mutation and no interference between groups of zebrafish (SUTVA). $Y_i(G_i$=g) is a function of unit i and genotype g.
- For each group of 5 larval zebrafish i=1,..., N :
  - $Y_i(G_i$=1)=value of Y when i is mutant
  - $Y_i(G_i$=0)=value of Y when i is wildtype.

Florian's experiment

- Consider a simple mutant-wildtype zebrafish experiment with an hypothetical completely randomized assignment mechanism.
- Consider N groups of 5 zebrafish, indexed by i, in an experiment hypothetically randomized to a binary genotype G, either mutant (i.e., G_i=1) or wildtype (i.e., G_i=0).
- An outcome variable Y (e.g., aggregation, alignment at 7 dpf) is measured for mutant and wildtype zebrafish.
- No hidden mutation and no interference between groups of zebrafish (SUTVA). $Y_i(G_i$=g) is a function of unit i and genotype g.
- For each group of 5 larval zebrafish i=1,..., N :
  - $Y_i(G_i$=1)=value of Y when i is mutant
  - $Y_i(G_i$=0)=value of Y when i is wildtype.
- Because of SUTVA, the science table simplifies to two columns and N rows.

Science and observed tables

- Unobservable science table :

i	$Y_i(G_i=0)$	$Y_i(G_i=1)$
1	$Y_1(G_1=0)$	$Y_1(G_1=1)$
...	...	...
N	$Y_N(G_N=0)$	$Y_N(G_N=1)$

Science and observed tables

- Unobservable science table :

i	$Y_i(G_i=0)$	$Y_i(G_i=1)$
1	$Y_1(G_1=0)$	$Y_1(G_1=1)$
...	...	...
N	$Y_N(G_N=0)$	$Y_N(G_N=1)$

- Observed data table :

i	G_i^{obs}	$Y_i(G_i=0)$	$Y_i(G_i=1)$
1	0	Y_1^{obs}	?
...	...	...	...
N	1	?	Y_N^{obs}

Science and observed tables

- Unobservable science table :

i	$Y_i(G_i=0)$	$Y_i(G_i=1)$
1	$Y_1(G_1=0)$	$Y_1(G_1=1)$
...	...	...
N	$Y_N(G_N=0)$	$Y_N(G_N=1)$

- Observed data table :

i	G_i^{obs}	$Y_i(G_i=0)$	$Y_i(G_i=1)$
1	0	Y_1^{obs}	?
...	...	...	...
N	1	?	Y_N^{obs}

$$Y_i^{obs} = G_i^{obs} Y_i(G_i^{obs} = 1) + (1 - G_i^{obs}) Y_i(G_i^{obs} = 0)$$

Causal inference : a missing data problem

Observed data table :

i	G_i^{obs}	$Y_i(G_i{=}0)$	$Y_i(G_i{=}1)$
1	0	Y_1^{obs}	?
...	...	...	...
N	1	?	Y_N^{obs}

- We can never get an exact measurement of a causal effect (e.g., $\tau_i = Y_i(G_i = 1) - Y_i(G_i = 0)$.

Causal inference : a missing data problem

Observed data table :

i	G_i^{obs}	$Y_i(G_i{=}0)$	$Y_i(G_i{=}1)$
1	0	Y_1^{obs}	?
...	...	...	...
N	1	?	Y_N^{obs}

- We can never get an exact measurement of a causal effect (e.g., $\tau_i = Y_i(G_i = 1) - Y_i(G_i = 0)$.
- Therefore, we have to turn into probabilistic inference.

Causal inference : a missing data problem

Observed data table :

i	G_i^{obs}	$Y_i(G_i{=}0)$	$Y_i(G_i{=}1)$
1	0	Y_1^{obs}	?
...	...	...	...
N	1	?	Y_N^{obs}

- We can never get an exact measurement of a causal effect (e.g., $\tau_i = Y_i(G_i = 1) - Y_i(G_i = 0)$.
- Therefore, we have to turn into probabilistic inference.
- Three major modes :

Causal inference : a missing data problem

Observed data table :

i	G_i^{obs}	$Y_i(G_i{=}0)$	$Y_i(G_i{=}1)$
1	0	Y_1^{obs}	?
...	...	...	...
N	1	?	Y_N^{obs}

- We can never get an exact measurement of a causal effect (e.g., $\tau_i = Y_i(G_i = 1) - Y_i(G_i = 0)$.
- Therefore, we have to turn into probabilistic inference.
- Three major modes :
  - Fisher / Fiducial (i.e., stochastic proof by contradiction)

Causal inference : a missing data problem

Observed data table :

i	G_i^{obs}	$Y_i(G_i{=}0)$	$Y_i(G_i{=}1)$
1	0	Y_1^{obs}	?
...	...	...	...
N	1	?	Y_N^{obs}

- We can never get an exact measurement of a causal effect (e.g., $\tau_i = Y_i(G_i = 1) - Y_i(G_i = 0)$.
- Therefore, we have to turn into probabilistic inference.
- Three major modes :
  - Fisher / Fiducial (i.e., stochastic proof by contradiction)
  - Neyman / Frequentist (i.e., sampling variance of an estimator, large-sample argument)

Causal inference : a missing data problem

Observed data table :

i	G_i^{obs}	$Y_i(G_i{=}0)$	$Y_i(G_i{=}1)$
1	0	Y_1^{obs}	?
...	...	...	...
N	1	?	Y_N^{obs}

- We can never get an exact measurement of a causal effect (e.g., $\tau_i = Y_i(G_i = 1) - Y_i(G_i = 0)$.
- Therefore, we have to turn into probabilistic inference.
- Three major modes :
  - Fisher / Fiducial (i.e., stochastic proof by contradiction)
  - Neyman / Frequentist (i.e., sampling variance of an estimator, large-sample argument)
  - Rubin / Bayesian (i.e., $P(Y_i^{mis}|Y_i^{obs}, G, X)$)

Stochastic proof by contradiction

- We followed three steps to assess the plausibility of a Fisher sharp null hypothesis (H_0) :

Stochastic proof by contradiction

- We followed three steps to assess the plausibility of a Fisher sharp null hypothesis (H_0) :

$$\forall i \in \{1, ..., N\} \quad Y_i(G_i{=}1) = Y_i(G_i{=}0).$$

Stochastic proof by contradiction

- We followed three steps to assess the plausibility of a Fisher sharp null hypothesis (H_0) :
  1. We chose an appropriate scalar test statistic, T, and defined more extreme.

Stochastic proof by contradiction

- We followed three steps to assess the plausibility of a Fisher sharp null hypothesis (H_0) :
  1. We chose an appropriate scalar test statistic, T, and defined more extreme.

$$\underline{\text{Ex. of T}} : \quad T(G, Y^{obs}) = \frac{\frac{1}{N_t}\sum_{i:G_i=1} Y_i^{obs} - \frac{1}{N_c}\sum_{i:G_i=0} Y_i^{obs}}{\sqrt{\frac{s_t^2}{N_t} + \frac{s_c^2}{N_c}}}$$

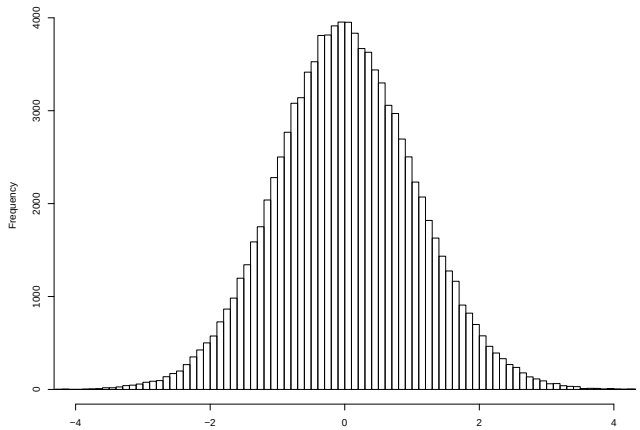
  - Statistic : function of the observed data.
  - Sensitive to expected departures from the null hypothesis.
  - Example of more extreme : $T(W, Y^{obs}) < T^{obs}$

Stochastic proof by contradiction

- We followed three steps to assess the plausibility of a Fisher sharp null hypothesis (H_0) :
  1. We chose an appropriate scalar test statistic, T, and defined more extreme.
  2. Assuming H_0, we calculated the value of T for all possible randomized allocations to obtain the null randomization distribution of T.

Stochastic proof by contradiction

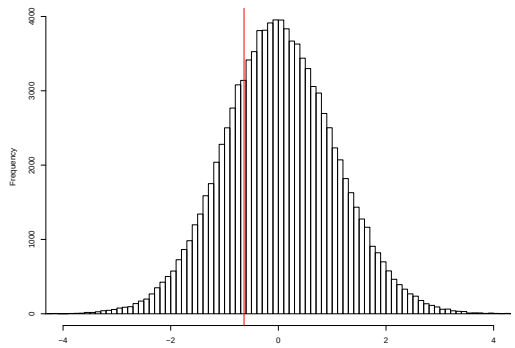
- We followed three steps to assess the plausibility of a Fisher sharp null hypothesis (H_0) :
  1. We chose an appropriate scalar test statistic, T, and defined more extreme.
  2. Assuming H_0, we calculated the value of T for all possible randomized allocations to obtain the null randomization distribution of T.



Each point in this histogram uses the observed data, what changes is whether each group of zebrafish has been "randomized" to wildtype vs. mutant.

Stochastic proof by contradiction

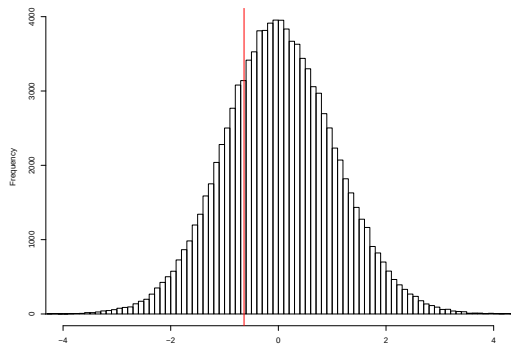
- We followed three steps to assess the plausibility of a Fisher sharp null hypothesis (H_0) :
  1. We chose an appropriate scalar test statistic, T, and defined more extreme.
  2. Assuming H_0, we calculated the value of T for all possible randomized allocations to obtain the null randomization distribution of T.



  3. We located the observed value of the test statistic, T^{obs}, in the null randomization distribution constructed in Step 2.

Stochastic proof by contradiction

- We followed three steps to assess the plausibility of a Fisher sharp null hypothesis (H_0) :
  1. We chose an appropriate scalar test statistic, T, and defined more extreme.
  2. Assuming H_0, we calculated the value of T for all possible randomized allocations to obtain the null randomization distribution of T.



  3. We located the observed value of the test statistic, T^{obs}, in the null randomization distribution constructed in Step 2.

The Fisher-exact p-value corresponds to the proportion of values of the test statistic that are as extreme or more extreme than the observed value of that test statistic.

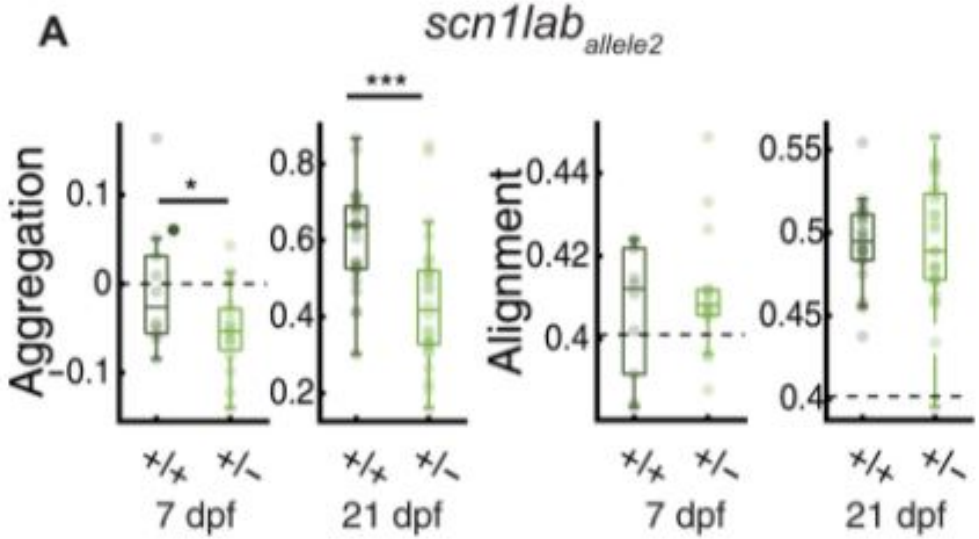
Standard Student's t test approximation

- Null randomization distribution historically approximated because of limited computing power, but used even in the late 20^{th} century.
- Capitalizes on the approximating null distribution of T_{Welch}
  - a Student's t distribution with df $\approx \frac{(\frac{s_t^2}{N_t}+\frac{s_c^2}{N_c})^2}{\frac{s_t^4}{N_t^2(N_t-1)}+\frac{s_c^4}{N_c^2(N_c-1)}}$.
- Nowadays, why not report randomization-based p-values ?

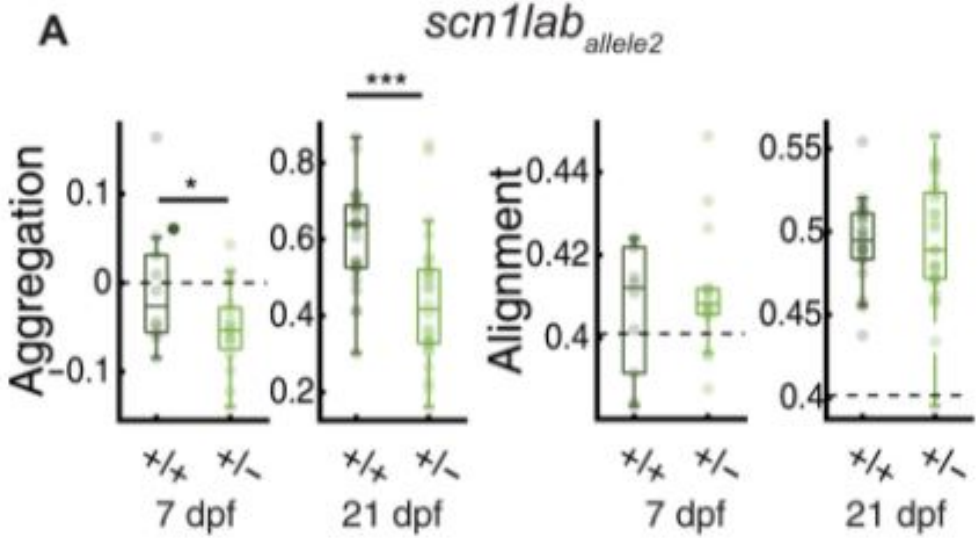
Return to Reviewer #1

- Recall Reviewer #1 : "The authors repeatedly and systematically misinterpret the p-values calculated for their statistical tests and consequently make a large number of false and misleading claims about the meaning of their results.

Selected results' interpretation - Aggregation

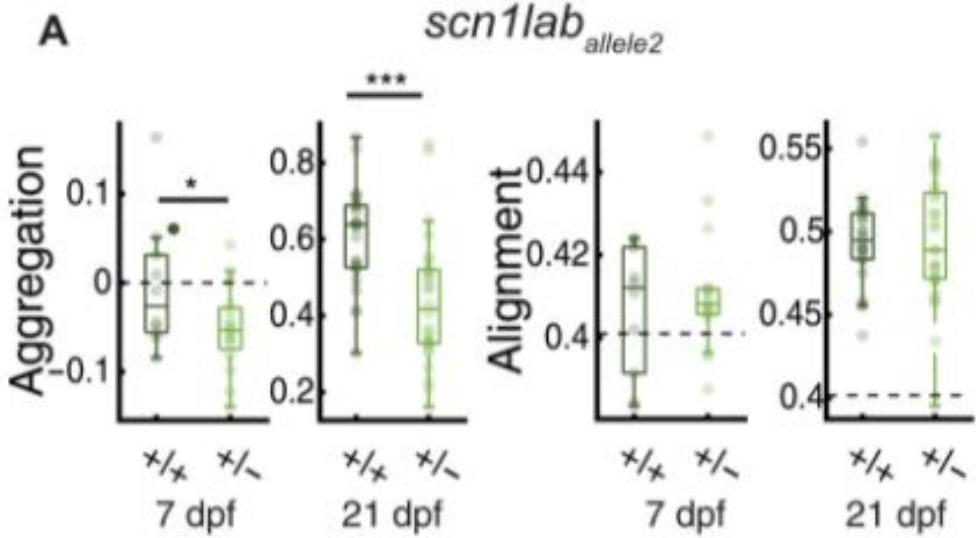


Selected results' interpretation - Aggregation



- "At 7 dpf, groups of mutant *scn1lab* fish are significantly less aggregated than wild-type siblings (P<0.05, N=26 groups ; ttest) and are also less aggregated than expected by chance (or, more dispersed) (P<0.001, N=16 groups ; ttest)."
- "At 7 dpf, *scn1lab*$^{+/-}$ fish are more dispersed than wild-type siblings ($P_{Fisher} \approx 0.036$, $N_{+/+} = 10$, $N_{+/-} = 16$, Cohen's d = 0.8). Dashed lines represent values of shuffled groups. At 21 dpf, fish are more aggregated. *scn1lab*$^{+/-}$ aggregate less than *scn1lab*$^{+/+}$ ($P_{Fisher} \approx 0.0001$; $N_{+/+} = 25$, $N_{+/-} = 26$ groups, Cohen's d = 1.09)."

Selected results' interpretation - Alignment



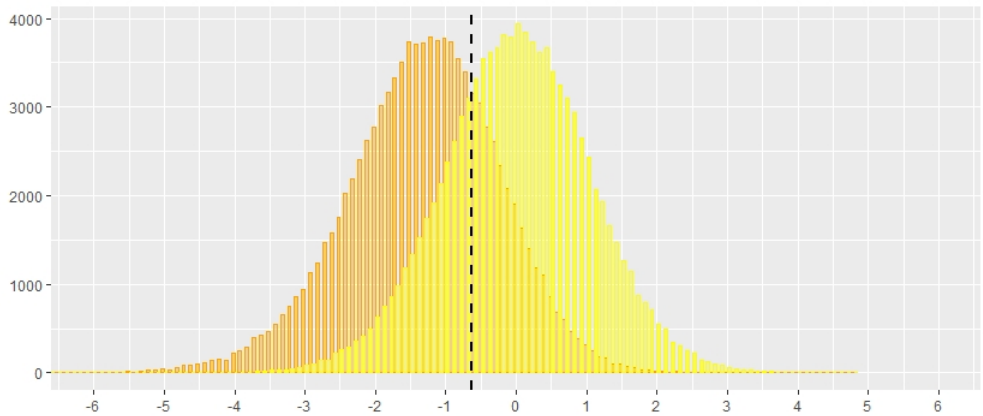
- "Groups of 7 dpf scn1lab$^{+/-}$ are more aligned than expected by chance ($P<0.05$, $N = 16$ groups ; ttest) and groups of 21 dpf fish are more highly aligned than 7 dpf fish. No direct effects of the mutations were observed at either age."
- "Group alignment increases with age ; however, we could not detect an effect of the *scn1lab* mutation ($P_{7dpf\ Fisher} = 0.26$ and $P_{21dpf\ Fisher} = 0.26$)."

Beyond our statistical analyses

- For the alignment experiment, Florian could have reported a counternull value, which was introduced by Rosenthal and Rubin (1994) as a nonnull value of an effect (e.g., difference in alignment between *scn1lab*$^{+/+}$ and *scn1lab*$^{+/-}$) that is supported by exactly the same amount of evidence as the null value of the effect.

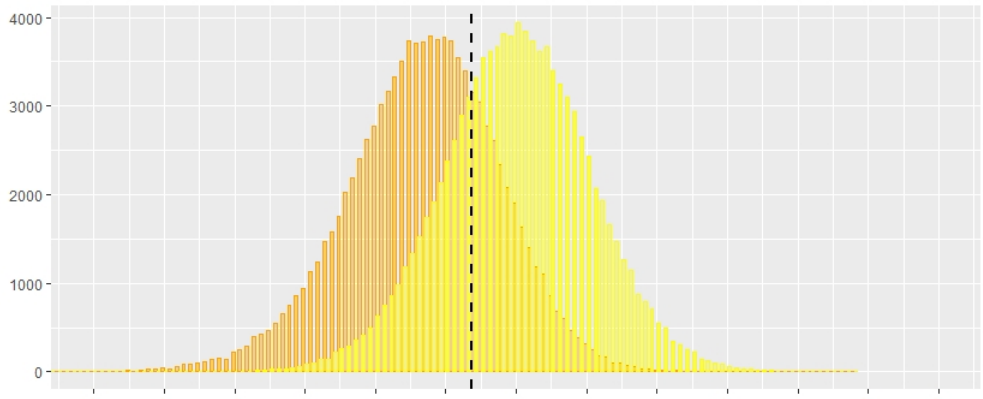
Beyond our statistical analyses

- For the alignment experiment, Florian could have reported a counternull value, which was introduced by Rosenthal and Rubin (1994) as a nonnull value of an effect (e.g., difference in alignment between *scn1lab*$^{+/+}$ and *scn1lab*$^{+/-}$) that is supported by exactly the same amount of evidence as the null value of the effect.



Counternull sets in randomized experiment

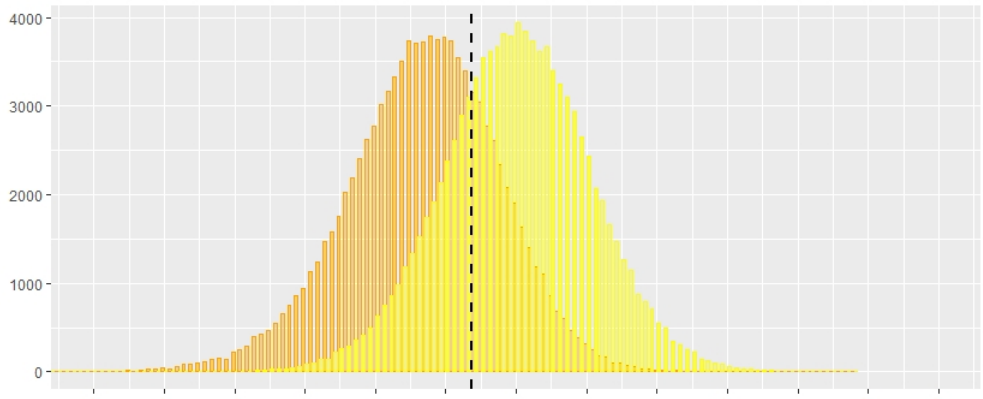
- Here, we can equate the randomization-based null and counternull p-values.



Counternull sets in randomized experiment

- Here, we can equate the randomization-based null and counternull p-values.
  - H_{Null} states that for each unit i,

$$Y_i(W_i = 1) - Y_i(W_i = 0) = 0$$



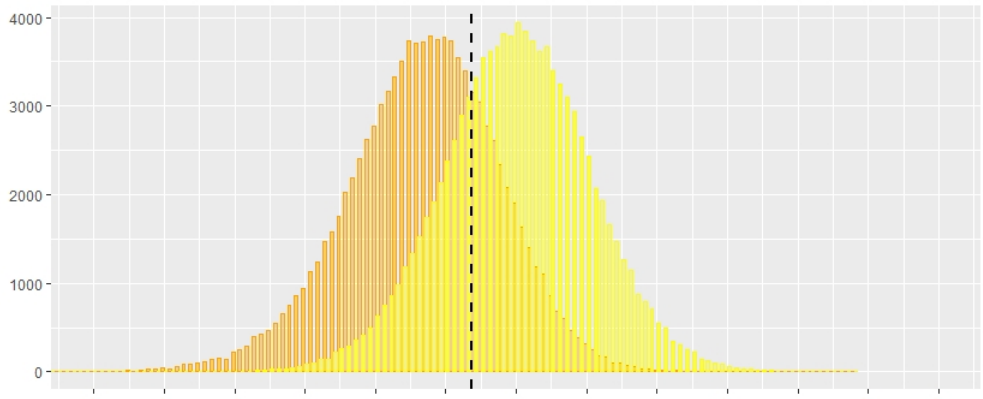
Counternull sets in randomized experiment

- Here, we can equate the randomization-based null and counternull p-values.
  - H_{Null} states that for each unit i,

$$Y_i(W_i = 1) - Y_i(W_i = 0) = 0$$

  - $H_{Counternull}$ states that for each unit i,

$$Y_i(W_i = 1) - Y_i(W_i = 0) = a \quad (a \neq 0).$$



Counternull sets in randomized experiment

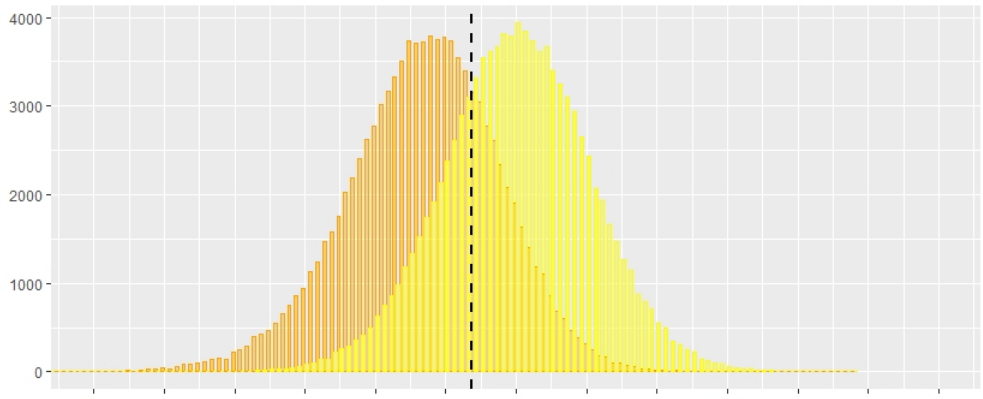
- Here, we can equate the randomization-based null and counternull p-values.
  - H_{Null} states that for each unit i,

$$Y_i(W_i = 1) - Y_i(W_i = 0) = 0$$

  - $H_{Counternull}$ states that for each unit i,

$$Y_i(W_i = 1) - Y_i(W_i = 0) = a \quad (a \neq 0).$$

- Equate proportion of values of the statistic that are as "unusual or more unusual" than T_{obs}.



Counternull sets in randomized experiment

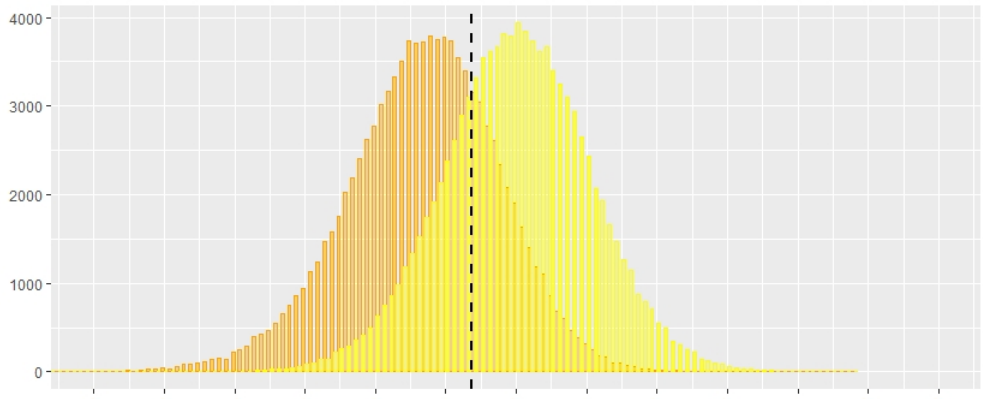
- Here, we can equate the randomization-based null and counternull p-values.
  - H_{Null} states that for each unit i,

$$Y_i(W_i = 1) - Y_i(W_i = 0) = 0$$

  - $H_{Counternull}$ states that for each unit i,

$$Y_i(W_i = 1) - Y_i(W_i = 0) = a \quad (a \neq 0).$$

- Equate proportion of values of the statistic that are as "unusual or more unusual" than T_{obs}.
- Counternull set =[-0.00762456 ; -0.00762451]



Counternull sets in randomized experiment

- Here, we can equate the randomization-based null and counternull p-values.
  - H_{Null} states that for each unit i,

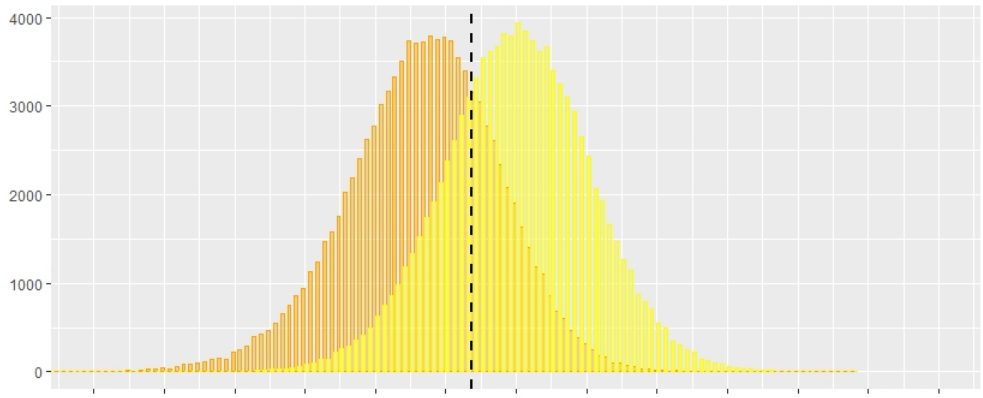
$$Y_i(W_i = 1) - Y_i(W_i = 0) = 0$$

  - $H_{Counternull}$ states that for each unit i,

$$Y_i(W_i = 1) - Y_i(W_i = 0) = a \quad (a \neq 0).$$

- Equate proportion of values of the statistic that are as "unusual or more unusual" than T_{obs}.

- Counternull set (21-day agg, WT vs. SCN) =[-0.01228 ; -0.01230]



Advantage of reporting counternull values

- The counternull approach helps avoid misinterpretations when testing a null hypothesis.

Advantage of reporting counternull values

- First, one of the reported p-values ≈ 0.27. Florian did not reject the null, and subsequently was tempted to accept the null hypothesis. In this situation, reporting the counternull forced a discussion on accepting the null value, but also the counternull!

Advantage of reporting counternull values

- Now consider a counternull value of an effect that is associated with a low p-value, but with low magnitude. In this situation, the counternull value is also worth reporting, because, if small, it would indicate that Florian's intervention is not necessarily biologically relevant.

Happy ending

Science Advances

Current Issue First release papers Archive About Submit manuscript

HOME > SCIENCE ADVANCES > VOL. 7, NO. 41 > COLLECTIVE BEHAVIOR EMERGES FROM GENETICALLY CONTROLLED SIMPLE BEHAVIORAL MOTIFS IN ZEBRAFISH

RESEARCH ARTICLE | NEUROSCIENCE

Collective behavior emerges from genetically controlled simple behavioral motifs in zebrafish

ROY HARPAZ, ARIEL C. ASPIRAS, SYDNEY CHAMBULE, SIERRA TSENG, MARIE-ABÈLE BIND, FLORIAN ENGERT, MARK C. FISHMAN, AND ARMIN BAHL Authors Info & Affiliations

SCIENCE ADVANCES • 6 Oct 2021 • Vol 7, Issue 41 • DOI: 10.1126/sciadv.abi7460

PDF Help

Happy ending

Taylor & Francis
Taylor & Francis Group

The American Statistician

ISSN: 0003-1305 (Print) 1537-2731 (Online) Journal homepage: www.tandfonline.com/journals/utas20

Counternull Sets in Randomized Experiments

M.-A. C. Bind & D. B. Rubin

To cite this article: M.-A. C. Bind & D. B. Rubin (2025) Counternull Sets in Randomized Experiments, The American Statistician, 79:2, 275-285, DOI: 10.1080/00031305.2024.2432884

To link to this article: https://doi.org/10.1080/00031305.2024.2432884

Published online: 17 Jan 2025.


thank
you