

Additional Research Material

Professor Alan Welsh

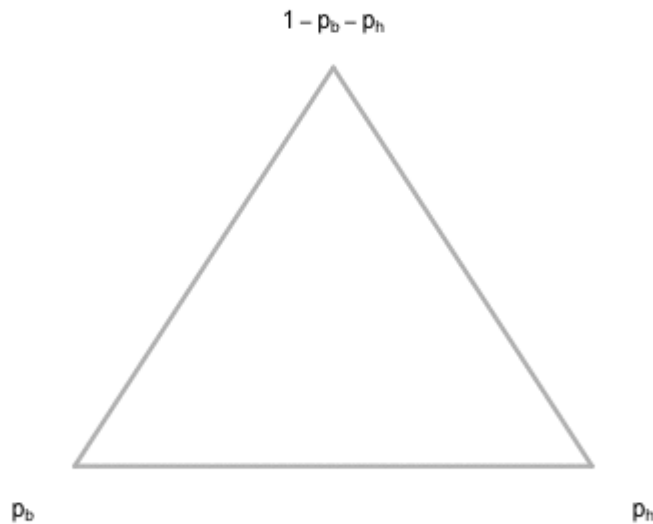
E.J. Hannan Professor of Statistics

Over the last decade, a method for analysing data called 'magnitude-based inference' has been developed and promoted in sport science as a new, improved statistical method. It has so far attracted little scrutiny by statisticians. This project considers 'magnitude-based inference' and its interpretation by examining in detail its use in the problem of comparing two means. The methodology is extracted from the spreadsheets which are provided to users of the analysis ([Sport Science Page](#)). The implemented version of the method is compared with general descriptions of it so that the method can be interpreted in familiar statistical terms.

'Magnitude-based inference' is not a progressive improvement on modern statistics. It does not replace the use of p-values by direct inference about magnitudes (as in using confidence intervals) but rather uses two different probabilities. These probabilities are not directly related to confidence intervals but rather are interpretable either as p-values for two different, nonstandard tests (for different null hypotheses) or as approximate Bayesian calculations which also lead to a type of test. This test is like using a standard test but at a very high level. This explains both how the method is 'less conservative' than the standard test and why it is not a real improvement on that test. The substantial reduction in sample sizes claimed for the method (30% of the sample size obtained from standard frequentist calculations) is not justifiable so the sample size calculations should not be used. Rather than use 'magnitude-based inference', a better solution is to be realistic about the limitations of the data and use either confidence intervals or a fully Bayesian analysis.

[Paper in Medicine and Science in sport and Exercise \(MSSE\).](#)

[Slides from talk presented at AIS, Friday 22 August 2014](#)

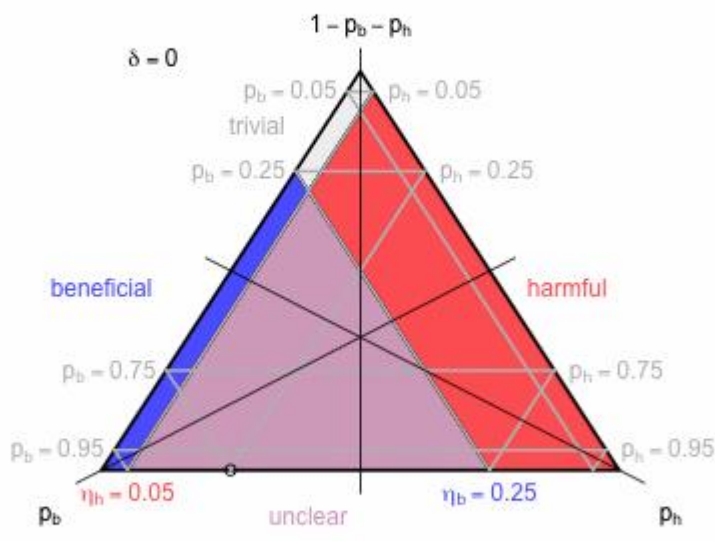


[Animation 1: Constructing the ternary diagram to interpret and show the effect of changing the thresholds \$\eta_b\$ and \$\eta_h\$](#)

The ternary plot to represent p_b , p_h and $1-p_b-p_h$ is constructed by drawing a triangle. Each probability is assigned to a vertex of the triangle. We draw in the p_b axis from the p_b vertex to the centre of the opposite side. The p_b vertex represents $p_b=1$, $p_h=0$ and $1-p_b-p_h=0$; the opposite side represents $p_b=0$ and the values of p_h and $1-p_h$. We draw in a line parallel to the opposite side at $p_b=0.05$ and label it. We repeat this for three additional values of p_b . Then we draw in the p_h axis and the lines parallel to the side opposite the p_h vertex to show the same 4 values of p_h . We complete the ternary diagram by drawing in the $1-p_b-p_h$ axis and lines parallel to the base to show 4 values of $1-p_b-p_h$. These gridlines are not labelled to prevent visual clutter.

For 'magnitude-based inference', we draw in a threshold value for p_b given by $\eta_b=0.25$. This threshold value partitions the triangle into 2 regions: a beneficial region ($p_b \geq \eta_b$) which we shade in blue and a not-beneficial region ($p_b < \eta_b$) that we shade in grey. We then draw in a threshold value for p_h given by $\eta_h=0.25$. The new threshold value partitions both the beneficial and the not-beneficial regions into 2 further regions. First, the grey not-beneficial region is partitioned into a trivial region ($p_b < \eta_b$ and $p_h < \eta_h$) which we leave shaded in grey and a harmful region ($p_b < \eta_b$ and $p_h \geq \eta_h$) that we shade in red. Second, the blue beneficial region is partitioned into a beneficial region ($p_b \geq \eta_b$ and $p_h < \eta_h$) which we leave shaded in blue and an unclear region ($p_b \geq \eta_b$ and $p_h \geq \eta_h$) that we shade in purple.

The animation then decreases the two thresholds together in a sequence of steps from $\eta_b = \eta_h = 0.25$ to $\eta_b = \eta_h = 0.05$ to present the effect of changing the threshold in 'mechanistic magnitude-based inference'. Finally, the animation holds $\eta_h = 0.05$ fixed and increases η_b in a sequence of steps from $\eta_b = 0.05$ to $\eta_b = 0.25$ to present the effect of changing the η_b threshold in 'clinical magnitude-based inference'.



[Animation 2: The effect of changing \$\delta\$; on \$p_b\$ and \$p_h\$ in the ternary diagram and the probabilities of finding an effect when there is none](#)

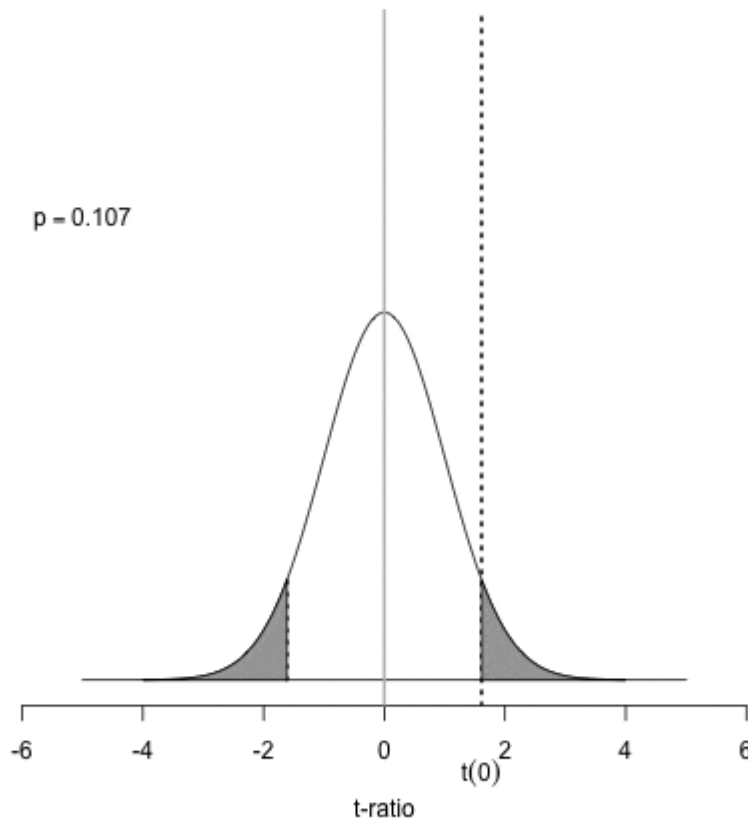
The one-sided p-value for a single sample is shown on the base of the triangle. The animation shows the path the triple p_b , p_h and $1-p_b-p_h$ traces through the ternary diagram as δ increases. This moves through the unclear to the beneficial and finally into the trivial region.

The animation is repeated for six different samples. The same pattern holds for samples with initial p_b sufficiently large ($p/2$ sufficiently small); once the initial p_b is smaller than 0.75, the path moves either through the unclear to the harmful and finally into the trivial region or through the harmful into the trivial region.

The next animation follows the triples p_b , p_h and $1-p_b-p_h$ for 200 random samples generated under the null hypothesis of no effect as δ increases. The empirical distribution is curved in the simplex and all the points eventually move up into the trivial region. The next sequence shows the same animation with plots of the empirical probabilities (based

on 10,000 samples) of p_b , p_h and $1-p_b-p_h$ falling in the beneficial, trivial and harmful regions as δ increases. The horizontal lines correspond to probabilities of 0.05, 0.25, 0.75 and 0.95; the dashed vertical line corresponds to $\delta=4.41$, a recommended default value of δ

for this example. Finally, we show larger versions of these empirical probability plots to show the details.



[Animation 3: The effect of changing \$\delta\$ on \$p_b\$ and \$p_h\$, showing both the Frequentist and the Bayesian interpretations of these probabilities](#)

The animation starts by showing the sampling distribution and the p-value for testing the null hypothesis of no effect. The figure shifts up so we can add a second figure to show the effect of changing δ . We first show the one-sided p-value which corresponds to p_h when $\delta=0$. The initial value of $p_h=0.054$ so the value for $p_b=1-p_h=0.946$. The value of p_h is shaded red because it is greater than the harm threshold $\eta_h=0.05$. We then add p_b when $\delta=0$; p_b is much greater than the beneficial threshold $\eta_b=0.25$ so is shaded light blue (so we can distinguish it later). The combined conclusion of red and blue is Uncertain (which was represented in purple).

As δ increases, p_b and p_h decrease. As p_h decreases to below $\eta_h=0.05$, the shading of the area representing p_h switches to blue. With both values shaded blue, the conclusion is that there is a Beneficial effect. As δ continues to increase, p_b eventually decreases to less

than $\eta_b=0.25$, at which point the shading becomes grey and the conclusion is that the effect is Trivial. The figure shifts up and simultaneously reduces δ to return to the start of the previous sequence so we can add an additional figure. This figure represents the posterior distribution of the difference in means. The final sequence shows the effect of increasing δ on the posterior probabilities that the difference in means is less than $-\delta$ and greater than δ respectively. Compared to the p-values, the areas are in the opposite tails and there is a single posterior distribution rather than two sampling distributions.