Introduction to Scientific Writing: Magnitude of Effects

Dr Cameron Hurst
cphurst@gmail.com

CEU and DAMASAC, Khon Kaen University

26th August, 2557

Preamble

- Almost always the main objective of a quantitative research article involves a statement about the magnitude of an effect
- Suppose you get a correlation coeffcient of $r = 0.35$ or an odds ratio of $OR = 1.5$
- Irres[ective] of whether these values are statistically significant, are they **clinically** significant
- Today we will talk about magnitude of effects and how we need to be careful with their interpretation

Conventions

Usual conventions:

Note:.....

Things to note given in a green box

Pitfalls:.....

Common mistakes and things to watch out for given in a red box

What we cover today (this session)

1. Introduction
2. Comparing groups
   - Continuous outcomes
   - Differences based categorical outcomes
3. Association among continuous variables
4. Concluding remarks

Magnitude of effects: Definition

The **magnitude of an effect** represents the **level of association** between two variables (or equivelently, the **level of difference** between two groups)

The magnitude of effect is also often called the **Effect size**, and it is by using the magnitude of the effect that we gauge the **clinical** importance of a result

Clinical vs Statistical significance

Before we continue, I would like to clarify and differentiate two distinct concepts used in clinical research:

- **Clinical Significance** is the whether our risk factor (or intervention) makes a **REAL** and tangible difference to the patients. Statistical hypothesis testing CANNOT TELL US THIS
- **Statistical Significance** is more about BOTH the level (magnitude) AND precision of our estimate. For example:

    - Even though our odds ratio may be high ($OR = 2.5$) we can't exclude 1 from the confidence interval ($p \not< 0.05$)
    - This quite large *magnitude of effect* might be an artifact of the sample we collected on the day; we can't rule out sampling variablility (chance) to explain its high value

Power, sample size and magnitude of effect

The fact is that formal statistical testing (using p-values) can even be misleading:

- If our sample size is TOO LARGE, a trivial value of our measure of association (e.g. $OR = 1.05$) can be highly statistically significant
- Conversely, if our sample size is TOO LOW, a seemingly important effect (e.g. $OR = 5$) cannot be considered important (it could be that high due to chance alone)

It is for this reason that clinical research reviewers (examiners, grant reviewers etc) are so obsessed with the approriate sample size (more later)

Magnitude of effects and statistical tests

For the the rest of this lecture, I will couch magnitude of effects in terms of specific statistical tests. In particular:

- Those comparing groups:
  1. Comparing two groups using a continuous outcome: **t-tests**
  2. Comparing groups using a categorical outcome: **Binary Logistic regression**
- Those concerning associations between continous variables: **Correlation** and **Linear regression**

A quick revision of the t-tests

I will assume that you are familar with the basic *Two independant samples t-tests*:

Hypothesis:
$H_0 : \mu_1 = \mu_2$ (On average, groups are equal)
$H_A : \mu_1 \neq \mu_2$ (On average, groups differ)

So we use this test to determine whether there is a **statistically significancant** difference between the groups (i.e. a difference unlikely to be by chance)

Test statistic for the t-test

The test statistic for the independent t-test is:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2))}{S_{\bar{x}_1 - \bar{x}_2}}$$

$\mu_1 - \mu_2$ represents the hypothesized difference (under H_0) and if we are testing $H_0 : \mu_{Exer} = \mu_{Contr}$ then $\mu_1 - \mu_2 = 0$

Note $S_{\bar{x}_1 - \bar{x}_2}$, the standard error of the mean difference: how much does $\bar{x}_1 - \bar{x}_2$ vary around from sample to sample, and:

$$S^2_{\bar{x}_1 - \bar{x}_2} = S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Where S_p is the pooled estimate of the standard deviation:

$$S_p = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

Revision of independent t-tests

In a test of (no) difference, we should have:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2))}{S_{\bar{x}_1 - \bar{x}_2}}$$

where $\mu_1 - \mu_2 = 0$, this simplifies to:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S_{\bar{x}_1 - \bar{x}_2}}$$

which tells us:

How many standard errors apart are our two sample means? Can conclude a 'statistically significant' difference between the two groups?

EXACTLY, we use a t-distribution, but for a decent sized sample ($n_1 = n_2 > 30$), the answer is about **2 standard errors**

Test statistics and sample size

Now let's have a close look at the test statistic:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

What is the first thing we should note??? The presence of n_1 and n_2 in the denominator.

The upshot is:

Sample size test statistics

All else being equal, the larger the sample size$\Rightarrow$ the larger a test statistic $\Rightarrow$ the more likely we are to identify a **statistically significant** difference

Effect of sample size on the test statistic

OK. Let's look at an example, to demonstrate the effect of sample size. For this example we will keep the mean difference and standard deviation constant:

$\bar{x}_1 - \bar{x}_2 = 10$; and
$S_{pooled} = 20$

Now we will consider three sample sizes:

1. $n_1 = n_2 = 5$: $t = \frac{10}{20\sqrt{\frac{1}{5}+\frac{1}{5}}} = 0.8$ and $p \not< 0.05$
2. $n_1 = n_2 = 50$: $t = \frac{10}{20\sqrt{\frac{1}{50}+\frac{1}{50}}} = 2.5$ and $p < 0.05$
3. $n_1 = n_2 = 500$: $t = \frac{10}{20\sqrt{\frac{1}{500}+\frac{1}{500}}} = 7.9$ and $p < 0.001$

We can see sample size has a profound effect....

Cohens' Δ

We need a measure of effect size that is indepedant of sample size. One such measure is called **Cohens' Δ**:

$$\Delta = \frac{\bar{x}_1 - \bar{x}_2}{S_p}$$

if we contrast this to the t-test test statistics:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

How do they differ?

Cohens' Δ

So how do we interpret Cohens' Δ??

Ans: The number of 'standard deviations' seperating the two means. Generally:

- Small: $\Delta = 0.2$
- Moderate: $\Delta = 0.5$
- Large: $\Delta = 0.8$

So for continuous outcomes, we now have a measure of the **Magnitude of effect**

Absolute vs Relative difference

Now what does this mean for *Clinically meaningful difference*? Not much.

Instead people will often use **Percentage difference** when trying to set a clinically meaningful effect of a continuous outcome.

For example, if in a control group the average cholesterol level is 120 units, we may want to reduce this by 20 % before we could conclude a new treatment works (i.e. 96 units). Such approaches are particularly common in sample size calculation.

Categorical outcomes: Comparing groups

Understanding effect sizes for continuous outcomes (between groups) can be a little painful. **We need to understand the scale of the instrument itself**. For example, Blood pressure (mmHg) etc. We need some sort of clinical knowledge to understand whether a differences is clinically important, or not.

Fortunely for categorical outcomes, we have a more standard (context-free) appraoch we can use to gauge effect size: The measure of association itself. Statistics like Odds ratios (OR), Relative risk (RR), Rate ratios (IRR) and Hazard ratios (HR) are all on a familar scale.

However, there are a few little details we still have to note.

Revision: Relative Risk vs Odds Ratio

- You should remember (from your Epi101) that the difference between the magnitude RRs and ORs is dependant on the prevalence of a disease
- For **rare** diseases, the **RR and OR are almost equivelent**
- BUT for more **common** diseases the **OR will always be more extreme**
- If we don't rememebr this, we may over estimate the impact of an exposure (or treatment) when we interpret the OR

To illustrate this, let's consider an example

Revision: Relative Risk vs Odds Ratio

In scenerio 1, let's consider a rare disaese:

	Outcome	
Exposure	+	-
+	2	998
-	1	999

Now:

$$OR = \frac{\frac{2}{998}}{\frac{1}{999}} = 2.002; \text{ and}$$

$$RR = \frac{\frac{2}{1000}}{\frac{1}{1000}} = 2$$

Revision: Relative Risk vs Odds Ratio

In scenerio 2, a more common condition (BUT with the exact same effect size):

	Outcome	
Exposure	+	-
+	200	800
-	100	900

$$OR = \frac{\frac{200}{800}}{\frac{100}{900}} = 2.25; \text{ and}$$

$$RR = \frac{\frac{200}{1000}}{\frac{100}{1000}} = 2$$

We can see that the RR has remained unchanged, whereas the OR indicates a stronger magnitude of effect.

Relative Risk vs Odds Ratio

Why is this?The answer is in what these two measures of association represent.

- Relative risk (or the Risk ratio) indicates the relative **chance** or **probablitiy** of the outcome in one group relative to the other
- In contrast, Odds ratios represent the relative **odds** of an outcome in one group relative to the other

Uphsot: RRs vs ORs

- Although ORs and RRs move in the same direction (e.g. $RR > 1 \Rightarrow OR > 1$),;
- Odds ratio **DO NOT** represent the relative chance (risk) of an outcome in one group relative to another
- Be very careful in your interpretaions of ORs.

Rate ratios (RRs), Rate ratios (IRRs) and Hazard ratios(HRs)

For the other two measures of association I mention-Incident rate ratios (IRRs) and Hazard ratios (HRs) we can interpret them in very much the same way as the RR. For example:

- If the indicence rate ratio (IRR) $= 2$, then the exposed group has twice the **rate** of the disease, relative to the unexposed group
- If the Hazard ratio is 2 (such that we might obtain from a cox regression when conducting survival analysis), then the chance of the hazard (e.g. death) is twice as likely in the exposed group, relative to the unexposed group.

A general guide: Effect sizes for binary outcomes

Now a guide you can use to gauge whether an effect size is large is:

Statistic	Small	Moderate	Large
RR, IRR and HR	1.2	1.5	3.0
OR	1.5	1.9	9.0
Percentage difference	10%	30%	50%

See final slide for reference

Special case: Continuous predictors and categorical outcomes

It is **VERY IMPORTANT** to note, that the above guides only work for categorical predictors (i.e. comparing groups)

WHY doesn't it apply to continuous predictors???????

Example: Changing the scale of predictors

Because the scale of many continuous predictors is abritraty. Let's consider an example using a predictor we all understand well: **AGE**

We could consider:

1. Age in months
2. Age in years
3. Age in decades

The magnitude of the effect will be different depending on the scale of the predictor

Scale of continuous predicotrs and statistical significance

Even though the **magnitude** of an effect can change with our scaling of a variable, it's **statistical significance WILL NOT**

Example of changing the scale of predictors: DMHT data

Let's consider an example:

I ran a standard bivariate logistic regression to determine the effect of age on achieving the Haemoglobin A1C clinical target in Type 2 diabetics. Now:

1. $OR_{Months} = 0.997,\ 95\%CI : 0.996, 0.997;\ p < 0.0001$
2. $OR_{Years} = 0.965,\ 95\%CI : 0.959, 0.972;\ p < 0.0001$
3. $OR_{Decades} = 0.697,\ 95\%CI : 0.648, 0.751;\ p < 0.0001$

So? Which has the largest effect size????

Correlation vs Linear Regression

Now we come to the final set of methods: Those that consider (only) continuous variables. Namely, **Correlation analysis** and **Linear Rgerssion Analysis**.

REVISION: Why use one approach over the other???

- We use Linear regression (only) when we have CONTEXTUAL (scientific) evidence that one variable explains the other (i.e a causal relationship)
- In contrast, correlation analysis only requires two variables to be associated

Pearson's correlation coeffcient, r

The formula for pearsons correlation coeffient is:

$$r = \frac{cov(x_1, x_2)}{\sqrt{var(x_1)var(x_2)}}$$

The important thing to note is the denominator. The covariance is standardized by the two variables' individual variances. This gives us a coeffcient:

$$r \in (-1, 1)$$

In other words, r is scale-independant and the closer that r is to 1 (-1), the stronger the positive (negative) **linear** association between x_1 and x_2

Raw β vs Standardized β (β_Z)

Like our problem I outlined above, in linear regression the size of the β coeffcient depends on two thing:

1. The level of association: The larger the association between our outcome and our predictor, the larger the magnitude of the effect will be.
2. **BUT** the scale of the variable also has a part to play. So again, we would expect $\beta_{Decades}$ to be larger than β_{Months}

Fortunately, we have a soulation to this problem: The standardized beta (β_Z)

Standardized β (β_Z)

Very simple idea: Rescale our outcome (y) and our predictors (x's) and then re rerun the regression. Now, let

$$Z_y = \frac{Y_i - \mu_y}{S_y}$$

and,

$$Z_x = \frac{X_i - \mu_x}{S_x}$$

Now refit the regression line:

$$Z_y = \beta_0 + \beta_1 Z_x$$

The resulting β_1 will be standardized (which we denote β_Z) and $\beta_Z \in (-1, 1)$ and we can interpret it in exactly the same way as Pearson's coerrelation coeffcient

Guide: Effect sizes for r and β_Z

So now the question remains:

When is a correlation coeffcient (or β_Z) high, and when is it low?

Generally speaking:

- r $\approx$ 0: no **linear** association
- r $\approx$ 0.1 (or -0.1): small linear association
- r $\approx$ 0.3 (or -0.3): moderate linear association
- r $\approx$ 0.5 (or -0.5): high linear association
- r $>$ 0.7 (or $<$-0.7): SUPER-DUPER linear association

Recap: Clinical vs Statistical significance

- These *Rules of Thumb* that I have given you are NOT WRITTEN IN STONE (they are a guide), we should, whenever it is possible, be contextually guided.
- We should not consider an OR (for example) as unimportant because it is 1.49999 rather than 1.5
- Especially for continuous predictors we should remember that there is sometimes a **threshold effect**.
  - For example, a 10 unit increase in cholesterol level for someone with a level of 100, is MUCH LESS PROFOUND then someone who has a base level of 150.
- Finally, we should also note that sample size plays a very imporant part in statements of **statistical** significance

Effect size and prospective powering:
Sample size calculation and Minimal clinical difference

The final point I would like to make is about **Minimal Clinical Difference**.

MCD represents the **Smallest possible IMPORTANT effect** we are trying to detect when we prospectively power a study (i.e. calculate sample size)

For example, we might ask questions like:

- What sample size do I need to show a **10 unit reduction in cholesterol** is statistically significant; or
- What sample size do I need to show a **25% reduction in cholesterol** is statistically significant; or
- What sample size do I need to show an **odds ratio of 1.5** is statistically significant; or

Magnitude of effect and sample size calculation

This *a priori* choice of effect size should be based 100 % on clinical science and **IMPORTANTLY** 0% on statistics.

What do I mean by this? It means I should **NEVER, NEVER, NEVER** base my minimal clinical difference on the 'findings' of another paper (i.e. Empirically). For example, I should **NOT** ever hear you say something like:
...We based our MCD on the findings of Hurst et al. (2013) who showed a 23.8 % in the cholesterol levels using drug XYZ.

Important point: Purpose of sample size calculation

Sample size claculation is to make a statement of **statistical significance** coincide with a statement of **clinically important difference**. So when I can say $p < 0.05$ I am ALSO saying we detected a clinically imporant effect.

A VERY important reference

All of the effect sizes (guides) I have quoted come from the refrence

McGraw, K. O., and Wong, S. P. (1992). A common language effect-size statistic. Psychological Bulletin, 111, 361-365.

Any questions??????

Thank-you!!!!!!

QUESTIONS???