British Standards Institution Study Day

How do we make sense of the results?

Martin Bland

Prof. of Health Statistics
University of York

http://martinbland.co.uk

Making sense of the results

We can look at how people present results in a leading journal: *The British Medical Journal*.

Link: http://www.bmj.com/archive/online/2011/05-30

Articles published between 30 May 2011 and 5 Jun 2011.

Four research papers.

Making sense of the results

Effect of evidence based risk information on "informed choice" in colorectal cancer screening: randomised controlled trial

Results . . . 345/785 (44.0%) participants in the intervention group made an informed choice, compared with 101/792 (12.8%) in the control group (difference 31.2%, 99% confidence interval 25.7% to 36.7%; P<0.001). More intervention group participants had “good knowledge” (59.6% (n=468) v 16.2% (128); difference 43.5%, 37.8% to 49.1%; P<0.001). A “positive attitude” towards colorectal screening prevailed in both groups but was significantly lower in the intervention group (93.4% (733) v 96.5% (764); difference −3.1%, −5.9% to −0.3%; P<0.01). The intervention had no effect on the combination of actual and planned uptake (72.4% (568) v 72.9% (577); P=0.87) . . . .

Making sense of the results

Effect of evidence based risk information on "informed choice" in colorectal cancer screening: randomised controlled trial

Results . . . 345/785 (44.0%) participants in the intervention group made an informed choice, compared with 101/792 (12.8%) in the control group (difference 31.2%, 99% confidence interval 25.7% to 36.7%; P<0.001). More intervention group participants had "good knowledge" (59.6% (n=468) v 16.2% (128); difference 43.5%, 37.8% to 49.1%; P<0.001). A "positive attitude" towards colorectal screening prevailed in both groups but was significantly lower in the intervention group (93.4% (733) v 96.5% (764); difference −3.1%, −5.9% to −0.3%; P<0.01). The intervention had no effect on the combination of actual and planned uptake (72.4% (568) v 72.9% (577); P=0.87) . . . .

Making sense of the results

Association between waiting times and short term mortality and hospital admission after departure from emergency department: population based cohort study from Ontario, Canada

Results 13 934 542 patients were seen and discharged and 617 011 left without being seen. The risk of adverse events increased with the mean length of stay of similar patients in the same shift in the emergency department. For mean length of stay ≥6 v <1 hour the adjusted odds ratio (95% confidence interval) was 1.79 (1.24 to 2.59) for death and 1.95 (1.79 to 2.13) for admission in high acuity patients . . .

odds ratio = 1 for no effect.

Making sense of the results

Longer term effects of very low energy diet on obstructive sleep apnoea in cohort derived from randomised controlled trial: prospective observational follow-up study

Results . . . After the very low energy diet period, apnoea-hypopnoea index was improved by −21 events/hour (95% confidence interval −17 to −25) and weight by −18 kg (−16 to −19; both P<0.001). After one year the apnoea-hypopnoea index had improved by −17 events/hour (−13 to −21) and body weight by −12 kg (−10 to −14) compared with baseline (both P<0.001). . .

Making sense of the results

Longer term effects of very low energy diet on obstructive sleep apnoea in cohort derived from randomised controlled trial: prospective observational follow-up study

Results . . . After the very low energy diet period, apnoea-hypopnoea index was improved by −21 events/hour (95% confidence interval −17 to −25) and weight by −18 kg (−16 to −19; both P<0.001). After one year the apnoea-hypopnoea index had improved by −17 events/hour (−13 to −21) and body weight by −12 kg (−10 to −14) compared with baseline (both P<0.001). . .

Making sense of the results

Effect of weekly vitamin D supplements on mortality, morbidity, and growth of low birthweight term infants in India up to age 6 months: randomised controlled trial

Results Between group differences were not significant for death or hospital admissions (92 among 1039 infants in the vitamin D group v 99 among 1040 infants in the placebo group; adjusted rate ratio 0.93, 95% confidence interval 0.68 to 1.29; P=0.68), or referral to the outpatient clinic for moderate morbidity. . .

Making sense of the results

Effect of weekly vitamin D supplements on mortality, morbidity, and growth of low birthweight term infants in India up to age 6 months: randomised controlled trial

Results Between group differences were not significant for death or hospital admissions (92 among 1039 infants in the vitamin D group v 99 among 1040 infants in the placebo group; adjusted rate ratio 0.93, 95% confidence interval 0.68 to 1.29; P=0.68), or referral to the outpatient clinic for moderate morbidity. . .

Making sense of the results

Effect of weekly vitamin D supplements on mortality, morbidity, and growth of low birthweight term infants in India up to age 6 months: randomised controlled trial

Results Between group differences were not significant for death or hospital admissions (92 among 1039 infants in the vitamin D group v 99 among 1040 infants in the placebo group; adjusted rate ratio 0.93, 95% confidence interval 0.68 to 1.29; P=0.68), or referral to the outpatient clinic for moderate morbidity. . .

Making sense of the results

Effect of weekly vitamin D supplements on mortality, morbidity, and growth of low birthweight term infants in India up to age 6 months: randomised controlled trial

Results Between group differences were not significant for death or hospital admissions (92 among 1039 infants in the vitamin D group v 99 among 1040 infants in the placebo group; adjusted rate ratio 0.93, 95% confidence interval 0.68 to 1.29; P=0.68), or referral to the outpatient clinic for moderate morbidity. . .

rate ratio = 1.0 if there is no difference.

Making sense of the results

What do these things mean?

- 95% confidence interval?
- P<0.001?

Two methods of “statistical inference”:

- confidence interval estimate,
- P value for a significance test.

Samples and populations

The data we have are from a sample from a much larger population.

Sample: people we have in the study.

Population: all the other people like them, including people who will be like them in the future.

We want to use the sample to tell us about the population.

Samples and populations

Problem: we have this sample.

Would another sample give us a different answer?

". . . 345/785 (44.0%) participants in the intervention group made an informed choice, compared with 101/792 (12.8%) in the control group (difference 31.2%, 99% confidence interval 25.7% to 36.7%; P<0.001) . . . "

Would another sample give difference = 31.2 percentage points?

Might it give 30.0 percentage points? Or 40 percentage points? Or even −31.2 percentage points?

The confidence interval enables us to deal with this problem.

Samples and populations

Problem: we have this sample.

Would another sample give us a different answer?

". . . 345/785 (44.0%) participants in the intervention group made an informed choice, compared with 101/792 (12.8%) in the control group (difference 31.2%, 99% confidence interval 25.7% to 36.7%; P<0.001) . . . "

How strong is the evidence that informed choice increases following evidence-based information?

Would it increase in another sample?

Would it increase in the population?

The significance test P value enables us to deal with this problem.

Confidence intervals

". . . 345/785 (44.0%) participants in the intervention group made an informed choice, compared with 101/792 (12.8%) in the control group (difference 31.2%, 99% confidence interval 25.7% to 36.7% . . . "

In the sample the difference is 31.2 percentage points.

We want to know the difference ***in the population***.

We cannot know exactly what it is.

The sample difference, 31.2 percentage points, is only an estimate of the difference in the population.

Confidence intervals

". . . 345/785 (44.0%) participants in the intervention group made an informed choice, compared with 101/792 (12.8%) in the control group (difference 31.2%, 99% confidence interval 25.7% to 36.7% . . . "

25.7% to 36.7% is also an estimate.

We estimate that, ***in the population***, the difference in percentage making an informed choice is somewhere between 25.7 to 36.7.

In mathematics, the set of values between 25.7 and 36.7 is called an interval.

"Between 25.7 and 36.7 percentage points" is called an interval estimate.

Confidence intervals

". . . 345/785 (44.0%) participants in the intervention group made an informed choice, compared with 101/792 (12.8%) in the control group (difference 31.2%, 99% confidence interval 25.7% to 36.7% . . . "

We can think of "25.7 and 36.7 percentage points" as a range of values within which we estimate the difference in the population to be.

Confidence intervals

". . . 345/785 (44.0%) participants in the intervention group made an informed choice, compared with 101/792 (12.8%) in the control group (difference 31.2%, 99% confidence interval 25.7% to 36.7% . . . "

Why "99%" confidence interval?

This is the difficult bit.

We choose the interval so that for 99% of the possible samples which we could take, of which this is just one, the interval would include the population reduction.

99% of confidence intervals include their population value.

1% do not.

More often use 95% rather than 99%.

Confidence intervals

". . . 345/785 (44.0%) participants in the intervention group made an informed choice, compared with 101/792 (12.8%) in the control group (difference 31.2%, 99% confidence interval 25.7% to 36.7% . . . "

How is the 99% confidence interval calculated?

Many methods for different situations.

What a lot of statistical method is about.

Best left for another day.

Significance tests

". . . 345/785 (44.0%) participants in the intervention group made an informed choice, compared with 101/792 (12.8%) in the control group (difference 31.2%, 99% confidence interval 25.7% to 36.7%; P<0.001) . . . "

The P value is a indicator of the strength of the evidence which this sample provides that, ***in the population***, the percentage making an informed choice increases when people are given evidence based information.

P is a probability, between 0 and 1.

Small P ➔ strong evidence = statistically significant.

Large P ➔ weak evidence = not statistically significant.

Usual cut-off for decision: P= 0.05.

In this study, cut-off for decision: P= 0.01.

Significance tests

". . . 345/785 (44.0%) participants in the intervention group made an informed choice, compared with 101/792 (12.8%) in the control group (difference 31.2%, 99% confidence interval 25.7% to 36.7%; P<0.001) . . . "

What is a P value, exactly?

Suppose information did not affect informed choice.

Would many possible samples of this size produce a difference as big as 31.2 percentage points?

The proportion of possible samples which would produce a difference as big as 31.2 percentage points is P.

Significance tests

". . . 345/785 (44.0%) participants in the intervention group made an informed choice, compared with 101/792 (12.8%) in the control group (difference 31.2%, 99% confidence interval 25.7% to 36.7%; P<0.001) . . . "

What is a P value, exactly?

In the example, P<0.001.

The proportion of samples which would have a difference as big as 31.2, or bigger, if the difference in the population were zero, is less than 0.001, or less than 1 in 1000.

This is small, so ***either*** we have a sample which is very unusual ***or*** the difference in population is not zero.

Here the evidence for a non-zero difference is strong.

The difference is (very highly) significant.

Significance tests

What is a P value, exactly?

Sometimes people say that the P value is the probability that the null hypothesis, e.g. that there is no difference in the population from which our sample was taken, is true.

Sometimes people say that the P value is the probability that the observed difference arose by chance.

Neither of these is quite correct.

The P value is the probability of a sample being as far from what you would expect, if the null hypothesis were true.

If the data would be unlikely if the null hypothesis were true, we take this as evidence that it is not.

Significance tests

". . . A "positive attitude" towards colorectal screening prevailed in both groups but was significantly lower in the intervention group (93.4% (733) v 96.5% (764); difference −3.1%, −5.9% to −0.3%; P<0.01) . . . "

Usual choices:

- P>0.05 ➔ not significant.
- P≤0.05 ➔ significant.
- P<0.01 ➔ highly significant.
- P<0.001 ➔ very highly significant.

We can vary these cut-offs sometimes.

In this study they used P<0.01 as significant.

Significance tests

". . . 345/785 (44.0%) participants in the intervention group made an informed choice, compared with 101/792 (12.8%) in the control group (difference 31.2%, 99% confidence interval 25.7% to 36.7%; P<0.001) . . . "

Why say "P<0.001" rather than give the exact P value?

Most statistical computer programs give P to only 3 or to 4 decimal places.

They do this because for many tests of significance, the calculation of P becomes imprecise when P is very small.

Significance tests

". . . 345/785 (44.0%) participants in the intervention group made an informed choice, compared with 101/792 (12.8%) in the control group (difference 31.2%, 99% confidence interval 25.7% to 36.7%; P<0.001) . . . "

How is the P value calculated?

Many methods for different situations.

What a lot of statistical method is about.

Best left for another day.

Significance tests

The most important thing about significance tests:

If P>0.05, we say that the difference is not significant.

This does not mean that there is no difference!

We have not found evidence for a difference.

Absence of evidence is ***not*** the same as evidence of absence.

Significance tests

The most important thing about significance tests:

Results . . . The intervention ***had no effect*** on the combination of actual and planned uptake (72.4% (568) v 72.9% (577); P=0.87) . . . .

99% confidence interval = −6.3 to +5.3 percentage points.

A difference of this size is compatible with the data.

We cannot conclude that the treatment had no effect.

We should give the confidence interval.

Significance test or confidence interval?

Significance tests and confidence intervals are two ways to make the link between sample and population.

Why have two approaches?

- Historical reasons, including some very argumentative statisticians.
- Computation problems: sometimes we cannot find a confidence interval without modern computing power.
- Computation problems: sometimes we cannot find a confidence interval in a straightforward way.
- Sometimes there is no meaningful estimate to find.
- Sometimes we are concerned with existence more than how big.

Significance test or confidence interval?

Which approach is better?

If a confidence interval can be found, it conveys more information.

It is the approach recommended for clinical trials by all major journals.

It is the approach recommended by CONSORT, the consolidated standards of reporting trials.

But, we cannot always find a confidence interval; we can almost always do a significance test.