number of times that we had to learn statistics before we understood it.

5 6 7 9 10 11 15 16 17 17 17 17 19 20 23 25 30

In this series of numbers, the mean is 13.58, the median is 17 and the mode is also 17. This illustrates one way of deciding whether information is normally distributed or not: in a bell shaped curve the median, mode and mean are all the same. Take this one more step: when describing normally distributed data, the mean is conventionally used to describe the average value (with the confidence intervals), whereas the median is used (with its range or, preferably, interquartile range) in non-parametric data. This means that if you are reading a paper, and the authors describe the data as non-parametric but use the mean and confidence intervals, then they do not know what they are talking about. (How impressive would it be to point that error out in front of your lecturer or consultant?). More usually, when a paper uses a mean and confidence interval then they are saying indirectly that the data are normally distributed.

What's the difference between a t-test and Mann-Whitney U test (and why is it important anyway?)
Once you have decided what the data are (qualitative versus quantitative, normally distributed versus non-parametric) you can decide what test to use (or when reading a paper whether they should have used that test in the first place). The simplest example is quantitative data. Often statistical tests try to compare two groups. If these groups are normally distributed a **t-test** is used, whereas when they are non-parametric a **Mann-Whitney U** test is used. If more than two groups are being compared another test is introduced, while for normally distributed data analysis of variance (ANOVA) is used. Another test that is often used in papers is the chi-squared (χ^2) test, which compares proportions (hence its full name: the χ^2 test of proportions). Essentially this compares the proportions in two groups: are there more asthmatics in group A or B? Or more women in the cases or controls?

What's the difference between an odds ratio and a relative risk?
This are another two terms that are often confused or considered to be synonymous. Let us explain these mathematically first, with reference to Table 1.

Table 1: Outcome One vs Outcome Two

	Outcome One	Outcome Two	Total
Cases	*5 (A)*	*14 (B)*	19 (E)
Controls	*12 (C)*	*5 (D)*	17 (F)
Total	17 (G)	19 (H)	36

The relative risk is also known as the risk ratio, and represents the ratio of risk in the exposed group (Cases) to the risk in the unexposed group (Controls). In Table 1, the relative risk of Outcome One is (A/E) / (C/F) or (5/19) / (12/17) = 0.37. This result means that the relative risk of Outcome One is 17% less in the exposed group to the controls, or in other words, the exposure is protective (if Outcome One is beneficial). This is usually easier to understand than an odds ratio; when the latest health scare is reported by the media (butter makes you 17% more fat!) they are usually referring to the relative risk. Results of cohort studies are most often quoted as relative risks.

The odds ratio is the ratio of odds of an outcome in the exposed group to the odds of an outcome in the unexposed group. In Table 1, the odds ratio is (A/B) / (C/D) or (5/14)/(12/5) = 0.14. Odds ratios are most often provided when reporting the results of case-control studies where the prevalence of the underlying outcome cannot be estimated. Odds ratios are slightly more difficult to understand, unless you get a kick out of maths (so why are you doing medicine?). Think of odds ratios as the odds of a greyhound winning a race (Santa's little helper at 5/1) and you've got the idea. So even though odds ratio and relative risk are often seen as being synonymous, they actually represent completely different values. (It's only when outcomes are rare that the OR and RR will be similar).

This article is really an introduction to the basics of relevant statistical tests. We have tried to show the differences between commonly used tests and terms. Most importantly, we hope that this short tutorial helps as you tackle and critically appraise the statistics section of the next paper you read.

Correspondence: CS O'Gorman[1,2]
[1]Graduate Entry Medical School, and [2]Centre for Interventions in Infection, Inflammation & Immunity (4i), University of Limerick
Email: Email: clodagh.ogorman@ul.ie

Further reading

1. **Evidence based Medicine:** How to practice and teach EBM. Sasckett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. 2nd Edition, Churchilll Livingstone 2000
2. **Essential Medical Statistics.** Kirkwood BR, Sterne JAC. 2nd Edition Blackwell Science 2003
3. **Practical Statistics for Medical Research.** Altman DG. 1st edition, Chapman and Hall,1991
4. **Interpretation and Use of Medical Statistics.** Daly L, Bourke G. 5th Edition, Blackwell Science, 2000
5. BMJ: **"Statistics Notes"** or **"Statistics for the non-statistician"**

What is the Difference Between Sensitivity and Specificity? Or Positive Predictive Value and Negative Predictive Value? And What's a ROC if It's Not a Type of Bird?

MF Higgins[1], AP Macken[2-5], W Cullen[2,3], J Saunders[6], C Dunne[2,3,5], CS O'Gorman[2-5]
[1]Maternal-Fetal Medicine, Mount Sinai Hospital, Toronto
[2]Graduate Entry Medical School, and [3]Centre for Interventions in Infection, Inflammation & Immunity (4i), University of Limerick
[4]The Children's Ark, University Hospital, Limerick
[5]National Children's Research Centre, Crumlin, Dublin
[6]C-Star, University of Limerick

Let's start honestly: sensitivity and specificity are two terms that confuse nearly everyone. As a medical student, they are something that you learn for an exam and then forget, until you meet them again at a journal club and the consultant starts talking about the sensitivity of the test and you frantically try to remind yourself where the false positives went and are the denominator false negatives or positives. Worse yet, when you are the consultant and are faced with a group of bright eyed trainees who

studied the new curriculum in medicine and who have excellent statistics skills… It's even worse again when you are doing research, or trying to decide whether to use a new test. Hopefully, if we understand the first principles, we will remember the more complicated concepts. So, let's start at the very beginning.

Here are some important abbreviations used in this tutorial:

TP	True Positive
FN	False Negative
TN	True Negative
FP	False Positive
PPV	Positive Predictive Value
NPV	Negative Predictive Value
ROC	Receiver Operator Curve

A concept that most medical students realise at some stage in their medical training is that tests could be wrong. For example, we all would have thought at one stage in our training that once a CT pelvis suggested someone had an ovarian tumour then that woman had to have an ovarian tumour. In university, you suddenly realise that every result is based on numerous factors. In the case of the potential ovarian tumour on CT, these factors will include the presence or absence of scar tissue in the area, the quality of the films and the experience of the person interpreting the results, amongst others. So, one of the fundamentals of evaluating the usefulness of a test is to factor in that fact that some tests may be wrong: either the test says someone has the disease when they don't (false positive [FP]) or the test says that someone does not have the disease when they actually do (false negative [FN]). A test is useful only if it has very few FP or FN: otherwise, why do it? There are four different ways of describing a test, each giving different pieces of information. Let's go through them one by one. We will start with sensitivity and specificity, which focus on the patients, and move on to predictive values, both positive and negative, which focus on the tests themselves.

Sensitivity

The textbooks will tell you that "sensitivity = TP / TP + FN"[1] or, in other words, the number of people correctly identified with the disease (true positives [TP]) divided by the total number of people with the disease (TP and FN). sensitivity gives an idea of how good a test is at correctly identifying those with the disease (alternatively, sensitivity is the risk of sending a guilty man to jail). Another way of remembering this is to use the mnemonic "SnNout": a test with high sensitivity (Sn-) with a negative result (-N-) will rule "out" the diagnosis.[3]

Specificity

Again, the textbooks will tell you that "specificity = TN / TN + FP",[2] or in other words the number of people correctly identified as being disease free (true negatives [TN]) divided by the total number of people who are truly disease free (TN and FP). Another way of looking at this is that specificity gives an idea of how good a test is at correctly identifying those who are well (in the context of the disease under investigation), or, the risk of setting an innocent man free. If you think of the mnemonic "SpPin"[3]: a test with high specificity (Sp-) with a positive result (-P-) will rule the diagnosis "in".

Clinical examples of sensitivity and specificity would be the diagnosis of ventricular fibrillation (VF) using a defibrillator. Obviously if a patient is in cardiac arrest due to VF, one needs to be pretty certain that the patient definitely has a VF before shocking them (high specificity) and not something else (high sensitivity). In such life threatening emergencies, when we say "pretty certain" we mean REALLY certain. Therefore the sensitivity and specificity of a defibrillator in diagnosing VF are 98.6% and 97.7% respectively[4].

Positive Predictive Value

Positive predictive value refers to the likelihood that a positive test result is correct. The textbooks will tell you that "PPV = TP / TP + FP", thus using only the positive results. A clinical example of a high positive predictive value would be in the diagnosis of liver fibrosis associated with Hepatitis C, which traditionally is diagnosed by a liver biopsy. An alternative test, the FIB-4 index, which combines aspartate aminotransferase (AST), alanine aminotransferase (ALT), platelets and age, has a positive predictive value of 82% (at a FIB4 index level greater than 3.25) in the prediction of liver fibrosis[5] compared with liver biopsy and fibrotest. The researchers in this study conclude that FIB-4 is a simple, accurate and non-invasive test for the assessment of liver fibrosis in Hepatitis C.

Negative Predictive Value

Negative predictive value (NPV) tells you the likelihood that a negative result is correct and, again, the textbooks will tell you that "NPV = TN / TN + FN". An example of this is the use of fetal fibronectin in the detection of preterm labour. When a woman presents with symptoms suggestive of preterm labour, but is not obviously in labour, a vaginal swab for fetal fibronectin has a high negative predictive value- 98%- for prediction of preterm labour within the next seven days.[6] What this means clinically is that women with a negative test are sometimes discharged home with advice, whereas those with a positive result may be retained in hospital for observation.[7]

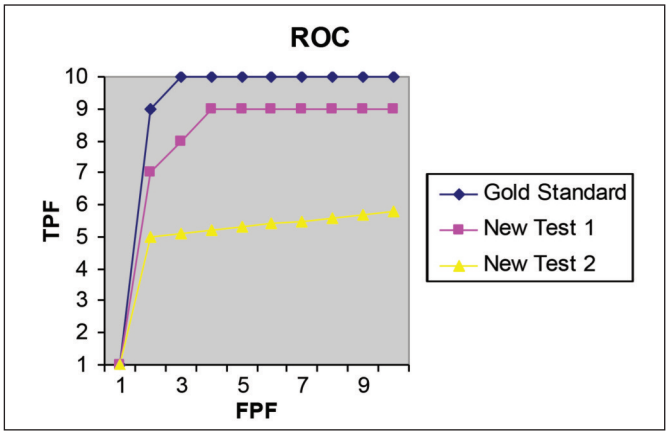

Figure 1 ROC curve comparing the perfect "gold standard" test to two new ones: the first compares reasonably well whereas the second is a poor test for diagnosing the condition compared to gold standard.

ROC (Receiver Operator Characteristic) Curve

Unless you go to a very research-orientated medical school, you might not learn about a ROC until you come to write your first paper and your supervisor suggests it. Rather than looking blankly at them, learn the basics of it now. ROC says a lot about the usefulness of a test in a graphic visual form.[8] Most of us are primarily visual, so while a sensitivity of 80% may sound great on paper, the ROC allows us to see it on a graph and also to compare it to other tests. The "gold standard" test is represented in Figure 1, which also shows two other tests that we shall deal with in a moment. If the x axis represents the FP fraction (FP/ total or [1-specificity]), and the y axis represents the TP fraction (TP / total or sensitivity), then the perfect test is shown as a line which slopes steeply, reaches the maximum y value quickly and then stays there. If this represents the "gold standard" test, then all other new tests can be compared to it on the same graph. Referring back to Figure 1, the first new test is good, but not as good as the gold standard test, and the second one is not reasonable at all. The best cut-off point for any test can be found by picking the point nearest the left-hand corner of the graph as long as the sensitivity scale goes to 1 or 100% and both axes start at 0. Several tests can be compared by calculating the area under the curve (AUC) for each test. This is only an introduction to ROC curves, but if you are interested in knowing more then we

recommend a website (http://www.rad.jhmi.edu/jeng/javarad/roc/JROCFITi.html) which allows you to change data and watch the resulting effects on the ROC.[8]

Like the other tutorials in this series, we aim to give you a working knowledge of sensitivity and specificity in order to use it clinically. For further reading, refer to any of the books or articles in the reference list.

Correspondence: CS O'Gorman[1,2]
[1]Graduate Entry Medical School, and [2]Centre for Interventions in Infection, Inflammation & Immunity (4i), University of Limerick
Email: Email: clodagh.ogorman@ul.ie

References

1. Diagnosis and Screening, Chapter 4 in Evidence-Based Medicine. How to practice and Teach EBM Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB 2nd Edition, Churchill Livingstone page 67-94.
2. Understanding sensitivity and specificity with the right side of the brain, Loong TW, BMJ 327 716-9.
3. Appraising diagnosis articles, Heneghan C, Badenoch D, Evidence-based Medicine Toolkit; 2nd Edition, BMJ Books, Blackwell publishing.
4. Ge et al. BioMedical Engineering OnLine 2002;1:5 doi 10.1186/1475-925X-1-5.
5. FIB-4: An inexpensive and accurate marker of fibrosis in HCV infection. Vallet-Pichard A, Mallet V, Nalpas B, Verkarre V, Nalpas A, Dhalluin-Venier V, Fontaine H, Pol S. Hepatology 2007.
6. Assessment of rapid fetal fibronectin in predicting preterm delivery. Obstetrics and Gynaecology. Tekesin I, Marer S, Hellmeyer L, Reitz D, Schmidt S. 2005 Feb 105: 280-4.
7. The impact of fetal fibronectin testing for women with symptoms of preterm labour in routine clinical practice within a New Zealand population. Groom KM, Liu e, Allenby K, ANZJOG 2006 Oct 46(5) 440-5.
8. Hanley JA, McNeil BJ. The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve. Radiology 1982;143:29-36.
9. www.anaesthetist.com/mnm/stats/roc/Findex.htm.

Research Confuses Me: What is Qualitative Research & What is the Difference Between Grounded Theory and Phenomenology?

CS O'Gorman[1-4], AP Macken[1-4], W Cullen[1,2], J Saunders[1,5], C Dunne[1,2,4], MF Higgins[6]
[1]Graduate Entry Medical School, and [2]Centre for Interventions in Infection, Inflammation & Immunity (4i), University of Limerick
[3]The Children's Ark, University Hospital, Limerick
[4]National Children's Research Centre, Crumlin, Dublin
[5]C-Star, University of Limerick
[6]Maternal-Fetal Medicine, Mount Sinai Hospital, Toronto

What is qualitative research? Are there different types of qualitative research? Is qualitative research important in medicine? Is it not enough for me to understand quantitative research and its methodology? It is unsurprising that medical students might feel overwhelmed studying many different research study designs. Some might even do as some of us initially did: we "skim-read" the chapters on qualitative analyses. We reasoned that so much medical teaching and medical literature relies on numbers (p-values, confidence intervals, odds ratios, relative risks, etc), the "hard data". By comparison, qualitative research has more words and fewer numbers. The outcomes can sometimes appear "obvious", and we wondered if so much effort was needed or justified. However, the benefits of qualitative research in medicine are widely recognised and accepted. This extends especially to the Cochrane Database of Systematic Reviews, where systematic reviews of *qualitative* studies are conducted.

So what is qualitative research? In this tutorial, we will explain the methodologies and terminologies of common types of qualitative studies, using two research questions for illustration: firstly, what are the attitudes of pregnant women with breech presentation to external cephalic version (ECV) and why might one woman opt for ECV while another refuses it? Secondly, what is the impact/emotional burden of an extra diagnosis of celiac disease in an adolescent with type I diabetes mellitus (T1DM)? In different types of qualitative studies, the phrasing of the research question will need to vary in order to try to adapt the question to the strengths of each individual study design. Many qualitative research terms (although defined in this paper under one study type) can actually be applied to qualitative methodologies.

External Cephalic Version (ECV) research question

Management of breech presentation is controversial, but common: 3% of term babies are breech.[1] The Term Breech trial showed increased safety for mother and baby if delivered electively by Caesarean Section[2]. ECV has been used for many years as a way of changing the position of the baby, making it easier to have a normal delivery. However, less than 60% of obstetricians routinely perform ECV,[3] less than 70% would recommend ECV, and only 78% would have an ECV themselves[4]. So what do women want? This is an ideal area for qualitative research – a complex, emotive question, with cultural and personal factors that involve not only the woman but her partner, her family, her perceptions of pregnancy and delivery…As medical professionals, we are aware of some factors - concerns regarding safety and pain - but what else is hidden, unexplored? It has been shown that ECV is more successful if a woman undergoes either clinical hypnosis or neurolinguistic programming prior to the procedure[5]…but why is that? Qualitative studies empower us to explore these hidden issues and concerns.

Coeliac Disease research question

The dietary limitations of having both coeliac disease and type 1 (insulin dependent) diabetes mellitus (T1DM) are significant. In children with coeliac disease, with or without T1DM, the introduction of a gluten-free diet has been shown to improve their qualitative sense of well-being and vitality,[6,7] as well as quantitative growth,[6,7] haemoglobin[8] and small intestinal mucosal histology.[9] The impact on the child with diabetes of a co-diagnosis of coeliac disease is an ideal question for a qualitative study. So is the impact of recommending a gluten-free diet to a child who is already trying to adhere to a dietary plan for their T1DM. Children may be asymptomatic prior to diagnosis[10] and this may lead to suboptimal dietary adherence.[11] Qualitative studies could highlight how health professionals can help to encourage dietary adherence.

Grounded Theory Approach

The grounded theory approach was developed in the 1960s[12], when sociologists studied the communication of health professionals with dying patients. Their results changed this communication forever from a culture of subterfuge to open discussion. Grounded theory is defined as "a way of thinking and conceptualizing the data"[13] (in other words, forming new theories).

Let us re-phrase our diabetes-coeliac research question: "What theory might explain the feelings and perceptions of adolescents