Foreword

Research Skills

What's the Difference Between PubMed and MEDLINE? And How Do You Best Search MEDLINE Anyway?
MF Higgins, AP Macken, W Cullen, J Saunders, C Dunne, CS O’Gorman

Research Confuses Me: What is the Difference Between Case-Control and Cohort Studies in Quantitative Research?
MF Higgins, AP Macken, W Cullen, J Saunders, C Dunne, CS O’Gorman

What is a Randomised Controlled Trial?
CS O’Gorman, AP Macken, W Cullen, J Saunders, C Dunne, MF Higgins

What Are The Differences Between A Literature Search, A Literature Review, A Systematic Review and A Meta-Analysis? And Why Is A Systematic Review Considered To Be So Good?
CS O’Gorman, AP Macken, W Cullen, J Saunders, C Dunne, MF Higgins

What are the Differences Between Common Statistical Tests?
MF Higgins, AP Macken, W Cullen, J Saunders, C Dunne, CS O’Gorman

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, AP Macken, W Cullen, J Saunders, C Dunne, CS O’Gorman

Research Confuses Me: What is Qualitative Research & What is the Difference Between Grounded Theory and Phenomenology?
CS O’Gorman, AP Macken, W Cullen, J Saunders, C Dunne, MF Higgins

What is the Difference Between Deontological and Consequentialist Theories of Medical Ethics?
CS O’Gorman, AP Macken, W Cullen, C Dunne, MF Higgins

Practical Skills

How to Teach Practical Skills in Medicine: Out of Hospital Training
MF Higgins, AP Macken, O Coyle, W Cullen, D McGrath, CS O’Gorman

How to Teach Practical Skills in Medicine: Bridging the Gap from the Course to the Patient, and Teaching on the Job
CS O’Gorman, AP Macken, O Coyle, W Cullen, D McGrath, MF Higgins
Foreword

Students in medical school become involved in research for many reasons. Many are looking for intellectual challenges and want to benefit from the facilities or mentorship available from the school’s faculty. Some may simply have an idea that they would like to pursue as part of the structured research modules within their course. Other students may already have an eye to the future and are considering how to best position themselves for an academic career. Regardless, most student researchers find that the research experience enhances their skill-sets and that the research process of identification of a subject to study, development of methodologies, implementation of their plans, analysis and interpretation of their results, and the subsequent reporting of the findings alters their perspective when encountering clinical problems.

Research by clinicians usually begins with a clinical, or patient-related, question. Thus its objective is usually to improve patient care or the patient experience. For either conducting research studies or interpreting published studies, there are certain basic principles of medical research, which need to be mastered. The objective of this supplement is to de-code the seemingly complex and vast dictionary of medical clinical research terminology, and thus to empower the students of today and the clinicians of tomorrow.

The Graduate Entry Medical School (GEMS) at the University of Limerick was established in 2007 and, since then, medical students from a wide range of backgrounds and with diverse primary degrees have completed research projects. These have included work completed within protected time during semesters but have also involved summer research, research electives and intercalated PhD degrees. Much of this research work has been supervised by principal investigators from the University of Limerick Centre for Interventions in Infection, Inflammation & Immunity (4i), which brings together a multidisciplinary team of researchers focussed on developing studies that directly impact health outcomes. A major interest within 4i is the translation of research findings into improved patient outcomes as well as greater effectiveness, efficiency and economic returns for healthcare provision (www.4i.ie).

To support the GEMS students, and the postgraduate trainees working with our clinical faculty, we developed a number of concise monographs detailing widely used tools, terms and techniques to aid them as they begin their research career. We are delighted to have been invited to share these monographs in the Irish Medical Journal. This supplement makes these available in a single resource, which we hope proves useful to anyone engaging with or teaching medical researchers. In the words of Albert Einstein: “The important thing is not to stop questioning. Curiosity has its own reason for existing.”

CS O’Gorman1,2, C Dunne1,2, M Larkin1
1Graduate Entry Medical School and 2Centre for Interventions in Infection, Inflammation & Immunity (4i), University of Limerick
Email: clodagh.ogorman@ul.ie

What’s the Difference Between PubMed and MEDLINE? And How Do You Best Search MEDLINE Anyway?

MF Higgins1, AP Macken2-5, W Cullen2-3, J Saunders6, C Dunne2,3,5, CS O’Gorman2,5
1Maternal-Fetal Medicine, Mount Sinai Hospital, Toronto
2Graduate Entry Medical School, and 3Centre for Interventions in Infection, Inflammation & Immunity (4i), University of Limerick
4The Children’s Ark, University Hospital, Limerick
5National Children’s Research Centre, Crumlin, Dublin
6C-Star, University of Limerick

Like most medical doctors, we have searched for information on how to care for patients. As students, you may wish to search to answer a question, or to obtain background information for a project or presentation (and once you are on call, we assure you that you will be searching at 3am for evidence supporting your next option in treating a patient). Most of us with an unanswered question will ask colleagues or local experts but when they are uncertain or differ then it is appropriate to search published research for solutions. MEDLINE is most commonly used, and can be accessed via PubMed (www.pubmed.com).

What is the Difference Between PubMed and MEDLINE? PubMed is a service of the US National Library of Medicine, which provides access to over 15 million MEDLINE citations (dating as far back as the 1950s!) and to additional life science journals (so PubMed is the access point for MEDLINE). MEDLINE is the actual bibliographic database, which is one of the 40 National Library of Medicine MEDLARS databases. MEDLINE is a computerised version of the printed Index Medicus (which took up shelves and shelves of space in the library when we were students but has now been relegated to the basement).

Before we learned how to use PubMed to search MEDLINE we spent a lot of time searching through (often irrelevant) articles in order to find the research answered our question. Some of us are lucky enough to have received formal training in MEDLINE searching, so we decided to write this tutorial so that others could also know how to get the most from searching. In reading this tutorial, we suggest you also access MEDLINE (via www.PubMed.gov) (www.ncbi.nlm.nih.gov/pubmed/ or www.pubmed.com), and have a clinical query of your own. For the rest of this tutorial, we will refer specifically to PubMed, basically because it provides world-wide free access! (MEDLINE, on the other hand, requires a subscription and the search strategy differs depending on which gateway your institution subscribes to for MEDLINE access). More formalised tutorials are also available on pubmed.gov, or on youtube (see below).

Search Question

The first and most important step in any search is to be definite about what you are searching for. For example, when talking about steroids for preterm delivery: are you looking at particular gestational ages? Does it matter if steroids are given for threatened preterm labour or for early pre-eclampsia? Is your outcome of interest just perinatal mortality or are you also looking at respiratory distress syndrome, intraventricular haemorrhage and necrotising enterocolitis?

One of the easiest ways to define a question is to use a PICOT framework. This divides your question into separate components and formalises your thinking.
CCF, it is best to search under your own words (called “free text”) Congestive”. Therefore, if you are looking for articles relating to difficulty coming up with synonyms then ask your colleagues, or a mixture of both free text and MeSH terms. The amount of time it takes to find the articles which were filed under the MeSH term “Heart Failure, Congestive”. Therefore, if you are looking for articles relating to CCF, it is best to search under your own words (called “free text”) including cardiac failure, pulmonary oedema, pulmonary edema and the MeSH term “Heart Failure”, “Congestive Edema, Cardiac” or “Cardiac Output, Low” (see Figure 1, steps to finding MeSH terms). You should now have a page of search terms, with a mixture of both free text and MeSH terms. The amount of time it takes to compile these terms will depend on your needs: if you are doing a formal Systematic Review then defining your search terms may take several full time days (if not weeks), whereas a simple question for a lunch time meeting will only take a few minutes. (Next time you are reading a systematic review, check out their search strategy and see if you could do better!)

Steps to finding MeSH terms
There are two rectangular boxes at the top of the screen on MEDLINE (via PubMed). The first says “search” and the second says “for”. The search box has options for places where one can search. Most people use just PubMed, but scroll down and you will find the “MeSH” option. Type in your free text term in the search box and then press go.

You should then see a screen that gives you the various MeSH options. Click on those you wish to select, then click on the “Send to” and pick the option “search box with and”. You will be sent to another screen where your MeSH term(s) are in a “search box”. Click on the “Search PubMed” box and the articles with these MeSH terms will be retrieved.

Figure 1 How to find the relevant MeSH terms for a PubMed search of MEDLINE

MeSH TERMS
MEDLINE files articles under particular key words called MeSH terms. If you searched for articles on cardiac failure, you may miss the articles which were filed under the MeSH term “Heart Failure, Congestive”. Therefore, if you are looking for articles relating to CCF, it is best to search under your own words (called “free text”) including cardiac failure, pulmonary oedema, pulmonary edema and the MeSH term “Heart Failure”, “Congestive Edema, Cardiac” or “Cardiac Output, Low” (see Figure 1, steps to finding MeSH terms). You should now have a page of search terms, with a mixture of both free text and MeSH terms. The amount of time it takes to compile these terms will depend on your needs: if you are doing a formal Systematic Review then defining your search terms may take several full time days (if not weeks), whereas a simple question for a lunch time meeting will only take a few minutes. (Next time you are reading a systematic review, check out their search strategy and see if you could do better!)

Boolean Operators: AND, OR, NOT
Boolean operators are terms that are used to define your search. Their use seems obvious: combining “cesarean section OR cesarean sections” should result in European and North American studies looking at c(a)esarean sections. It is good to start by combining each of your categories with “OR”. For example, if you are looking at preterm labour, for participants you combine “preterm labour” OR “preterm labor” OR “preterm contractions” OR “Obstetric Labor, Premature”[MeSH]. This should then result in all the studies relating to your participants. Resist the urge to start looking through the results for the moment, and repeat this search using your terms for your intervention, comparison and outcome.

Now you need to start combining the groups. Underneath the search box is another series of boxes labelled limits, preview, history, etc. By clicking on ‘history’ you will find a record of all the searches that you have performed each labelled by a number. The next step is to do a very specific search by combining your four groups by the term “AND” (Figure 2); in this particular example, this resulted in 110 articles (search performed on 19th July 2012). This is a very specific search resulting in articles with all the groups mentioned.

Figure 2 An example of a simple search strategy based on a PICO question.

Note the use of MeSH terms and the reduction in numbers of articles obtained by using Boolean operators (AND, OR).

Sensitive Search
A second way of combining your groups is to use the term “OR”. This will result in many more articles relating to the question, but will include articles where the search term relating to, for example, the intervention, was not included. A good idea is to start with a very specific search: if this results in very few articles then you could proceed with a sensitive search. However, if your very specific search results in hundreds of articles then this is an area that has been extensively researched. You may then choose to expand the search further, but this will depend on whether you are doing a very thorough review or searching for one easily found article.

Limits
The “limits” button will lead you to a page which allows you to limit your search based on article type, year of publication and language amongst other options. A quick search is certainly easier if the results are limited to those in English (for example, as it is the language we are using for this tutorial). However, if you are searching for a systematic review, then it is considered a source of bias to search in only one language.

Clinical Queries
“Clinical queries” is a search option that allows you to search based on diagnosis, treatment and aetiology amongst other options. This also allows you to search for systematic reviews specifically. The design of “clinical queries” include search filters which aim to improve the yield of searches for clinically-relevant studies and reduce the number of “false drops”!

Take preterm labour; your general question may be: Does tocolysis (medication to stop labour) work? The formalised question now becomes: In patients with threatened preterm labour between 24 and 34 weeks gestation (participants) does atosiban (intervention) versus no tocolysis (or ritodrine, or whatever comparison you are interested in) reduce preterm labour (outcome) within one week of administration? (time).
Research Confuses Me: What is the Difference Between Case-Control and Cohort Studies in Quantitative Research?

MF Higgins1, AP Macken2-5, W Cullen2-3, J Saunders2, C Dunne2,3,5, CS O’Gorman2,5
1Maternal-Fetal Medicine, Mount Sinai Hospital, Toronto
2Graduate Entry Medical School, and 3Centre for Interventions in Infection, Inflammation & Immunity (4i), University of Limerick
4The Children’s Ark, University Hospital, Limerick
5National Children’s Research Centre, Crumlin, Dublin

What is the difference between a cohort and a case-control trial? And why is it important? As a student, it is sometimes difficult to appreciate the difference between these two study methods, and why should it matter to us anyway? After all, we study medicine to treat patients, not statistics. Study methodologies were for the scientists; we are clinicians. Fast forward to clinical practice, and the importance of research design becomes apparent. As medical doctors we treat patients, but we also look at the bigger picture: why is this happening to this patient? Why is this patient more likely to be affected than another? In order to truly care for patients it is necessary to search and query and that means doing, or being able to properly interpret, research.

The most fundamental point of both cohort and case-control trials is that they are observational trials. Unlike randomised controlled trials (RCTs), where the researchers actively divide participants into control and comparison groups, observational trials are more passive: here the researchers literally observe participants. The major drawback is the potential for bias: apparent differences may be due to known or unknown confounders. However, in emotive or ethically difficult areas (e.g., obstetrics or paediatrics) or in situations when blinded randomisation is not possible (e.g., surgical procedures) they may be the best quality evidence available.

To illustrate the differences between the two study types, a good example is the history of research into lung cancer. We all know that smokers are more likely to develop lung cancer, but where did that knowledge come from? And what if you look at this from the other direction: *How many persons with lung cancer were smokers?* These two ways of looking at a question illustrate the differences between a cohort and a case-control trial perfectly. In fact, over fifty years ago in the UK a young doctor and a statistician asked just that same question and decided to use these two methods to find an answer.

### Case-Control

**How many people with lung cancer were smokers?**

Sir Richard Doll (an epidemiologist) and Sir Austin Bradford Hill (a statistician) started off by looking at patients with lung cancer: the *“cases”*. They then picked a group of controls, patients without lung cancer but in hospital for another reason. Looking back in time (retrospectively) they tried to ascertain what the cases had been exposed to that made them more likely to develop lung cancer than the controls. The cases were divided into those exposed to smoking and those unexposed. A similar group (in this case, other hospital patients) were similarly divided into exposed and unexposed groups (Figure 1). Due to the risk of confounding (see below) the researcher then assumes (and hopes) that the cases and controls come from the same population. So the key features to a case-control trial are *retrospective* and *comparison*.  

![Case Control Trial](Figure 1 Case Control Trial)

### References

4. Correspondence: CS O’Gorman1,2
5. Graduate Entry Medical School, and Centre for Interventions in Infection, Inflammation & Immunity (4i), University of Limerick.
6. Email: clodagh.ogorman@ul.ie

**Where to Learn More**

PubMed has an online tutorial on how to search MEDLINE, which is very useful. There are also other published articles and books on how to get the most of your medline search. NCBI have uploaded tutorials on YouTube on how to use search Medline (http://www.youtube.com/user/NCBINLM),…things have come a long way since looking up Index Medicus in the library on a wet Wednesday night in November!

**SUGGESTED FURTHER READING.**


The difficulty can, sometimes, be deciding who has been exposed or unexposed: smoking might be obvious, but what do you do with ex-smokers? And what if the exposure is not obvious: how many mothers, for example, would know if they took certain medications during pregnancy? While those who have the disease may have searched for a cause, those without the disease may never be aware of the exposure they had. This can lead to “recall bias” as participants may deny having been exposed, simply because the exposure meant nothing to them.4

A second difficulty is choosing your controls: what if Doll and Hill had chosen patients with emphysema? Their smoking outcome (lung cancer versus emphysema) would have been different but their smoking habits might have been the same. These are called confounders: factors which link two groups and suggest a causal relationship. While some confounders are obvious (e.g., grey haired people have higher surgical mortality) others are not so obvious (e.g., eating eggs increases your risk of a myocardial infarction, or is this confounded by regularly eating fried bacon with these eggs?). This is another advantage of RCTs over observational trials: RCTs balance groups for confounders that have not yet been described.3

Case-control trials are usually retrospective, so the data are usually ready to be collected; therefore the study is cheaper and quicker to complete than a prospective trial. This is an advantage when studying diseases with a long latency period. Equally, if you are researching a rare disease or outcome using a cohort study, it would require huge numbers of exposed persons and many years to obtain enough people who develop the rare outcome. Therefore, a case-control trial (where the participants have already developed the rare outcome) is a more efficient use of resources and would require a smaller sample size. However, the information available is limited by what other people thought was important at the time the data were recorded and this, in itself, may lead to bias.

The key features to case-control trials, therefore, are outcome, (usually) retrospective, case-control versus cohort study, and case-control period. It is, however, also possible to have prospective case-control trials: take for example a study of serum lipoprotein as a risk factor for coronary heart disease. Here, the participants (men aged 50 years) had blood samples taken and frozen. Fast forward six years, through which the men were followed to see whether they had developed coronary heart disease or not. Those who had (the cases) were compared to randomly selected controls from the remaining participants. The blood samples of the two groups were compared to see if they differed with respect to their concentration of serum lipoprotein(a). This meant that not all the blood samples needed to be analysed, but that the cases could be accurately compared to controls for exposure to differing concentrations of lipoprotein(a) (this is an example of a nested case-control trial, that is, a case-control trial nested within a cohort trial).5

The main disadvantage of prospective case-control trials is that the ratio of cases to controls is artificially ‘created’, meaning that the prevalence of the condition cannot be estimated from the data collected and so absolute risks and, therefore, relative risks cannot be estimated either. Therefore, odds ratios are given for any risk factors.

Cohort

Smokers are more likely to develop lung cancer

The key features to a cohort trial is to follow a group of exposed persons forward in time to see if they develop the outcome.6 For example, if people are followed prospectively to evaluate if smokers are more likely to develop cancer than non-smokers.7 There are also retrospective cohort trials but for the moment, the key features to a cohort trial are exposure and prospective. Thus a cohort trial is also termed a prospective trial. In epidemiology, a cohort is a defined population that is followed prospectively to see who develops an illness. No new additions are made, and an attempt is made to follow all those who comprised the original group.7

In Doll and Hill’s study where the outcome was lung cancer, the study would require years of follow up to see if any of the participants would ever develop the disease. (There’s no point in stopping follow up at twenty years if patients may be diagnosed after twenty-five years). Due to the time and effort required, it is first important to at least have a basis for your hypothesis: this is why many researchers start with a case-control trial (as Doll and Hill did). Another disadvantage of the time required for the study is the loss to follow up: how do you keep in contact with your original study group? Doll and Hill very sensibly decided to have as their cohort doctors registered with the General Medical Council (the GMC, the professional organisation for medical doctors in the UK). So they would have a group of participants who would be easier to follow (as they needed to remain registered with the GMC) and would also, appreciate the importance of participation in such a study.

The cohort trial of British medical doctors has been published every five years since 1954, the last update being published in 2004.8 The original trial gained worldwide notice, and resulted in similar trials starting in other countries, including the Nurses Health study in the US.9 Information obtained on multiple outcomes in smokers was obtained which has significantly changed medical treatment. However, this is a very long-term study and involved significant work from the authors and their support teams.

The advantage of cohort studies over case-control studies is that as long as a representative cohort has been recruited, then prevalence, absolute and relative risks can be estimated readily for any risk factors.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Key features comparing case control vs. Cohort studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case Control Studies</td>
<td>Cohort Studies</td>
</tr>
<tr>
<td>Starts with the disease</td>
<td>Starts with the exposure</td>
</tr>
<tr>
<td>Usually retrospective</td>
<td>Usually prospective</td>
</tr>
<tr>
<td>Odds ratio</td>
<td>Relative risk</td>
</tr>
<tr>
<td>Usually quicker to perform</td>
<td>Usually longer to perform</td>
</tr>
<tr>
<td>Good to study rare disease</td>
<td>Good to study rare exposures</td>
</tr>
</tbody>
</table>

In summary, case-control trials and cohort trials are easily confused, so the key features are shown in Table 1. Both are observational trials. However, case-control trials compare cases with controls to investigate what each group was exposed to such that cases become cases and the controls remain controls. So, the key features to case-control trials are retrospective (usually) and outcome. In contrast, cohort trials only investigate one group and usually follow them forward in time to see if they develop a disease. Therefore, the key features to cohort trials are exposure and prospective (usually).

Correspondence: CS O’Gorman1,2

1Graduate Entry Medical School, and 2Centre for Interventions in Infection, Inflammation & Immunity (4i), University of Limerick

Email: clovesogorman@ul.ie

References


**What is a Randomised Controlled Trial?**

CS O’Gorman1,4, AP Macken1-4, W Cullen 1-4, J Saunders1-4, C Dunne1-4, MF Higgins6
1Graduate Entry Medical School, and 2Centre for Interventions in Infection, Inflammation & Immunity (4i), University of Limerick
3The Children’s Ark, University Hospital, Limerick
4National Children’s Research Centre, Crumlin, Dublin
5C-Star, University of Limerick
6Maternal-Fetal Medicine, Mount Sinai Hospital, Toronto

**The basics**

A randomised controlled trial (RCT), also known as a randomised controlled clinical trial, is a study in which participants are assigned randomly to one of two or more arms (groups with different interventions) of a clinical trial. Occasionally, a placebo is used as one of the interventions, but, generally, if there is a recognised and accepted intervention that works (the “gold standard”), then a new drug, device or intervention is tested against this gold standard rather than against placebo. Where a gold standard drug or intervention exists, it would be unethical to randomise to a placebo and, by doing so, make an effective treatment unavailable to some participants. Generally, RCTs are conducted because there is equipoise (or uncertainty) about whether a new intervention is potentially better than an existing one. The trialists (the team of people that plan, conduct, supervise and analyse the results of the trial) start with the hypothesis that there is no difference between the two interventions (this is “the null hypothesis”). The purpose of the RCT is to reject or accept the null hypothesis. If they manage to reject the null hypothesis, they can accept the “alternative hypothesis”, i.e., that there is a difference between the two interventions.

**Blinding**

RCTs can be, but are not always, blinded. Blinding means that someone who plays an active part in the trial does not know what treatment (new intervention, gold standard or placebo) has been assigned to each participant. Trials can be single-blinded, double-blinded or even triple-blinded depending on how many types of people involved in the trial are blinded. For example, the participant could be blinded and not know what intervention they are assigned to. Or the medical doctor who deals with all the participants might not know which intervention each participant is assigned. Or the statistician who reviews all the datasets and performs the statistical analysis might not know which group of participants has been assigned to which intervention. As you can imagine, if the participants are blinded, then there is less likelihood that they will complain of symptoms or side-effects that are known to be associated with either the new intervention, or the gold standard, or placebo. Similarly, a blinded doctor is less likely to assess patients in a biased way. To avoid the bias the statistical tests should be chosen prior to starting the RCT along with the rationale given for choosing them. Everybody has biases – even you!

**Random Allocation**

RCTs by definition, randomly allocate participants to the different arms. This is designed to mimic chance, and to ensure that there is no difference between groups. A good trial published in a journal will show the characteristics of the various intervention groups summarised (usually in a table) and compared (often with p values and confidence intervals, though not always) to prove to the reader that there are no differences at baseline.

Randomisation does not mean assigning alternate treatments to every second patient, nor assigning intervention A to patients who present on Mondays, Wednesdays and Fridays and intervention B to all others; if a well-meaning, but biased, physician wants his favourite patient to be assigned to intervention A he can tell that patient to come in on the day that intervention will be assigned; this is known as selection bias. Randomisation is designed to prevent biases, as well as to ensure “same-ness” between the assignment groups. The best method of randomisation is to use computer software to generate a sequence of random numbers, where each number refers to one of the interventions.

**Sample Size Calculations**

Calculating the number of patients needed for a trial is important. If you can show that a new intervention is statistically significantly better than the old intervention by randomising fifty patients, you can avoid the expense of randomising and treating one hundred patients (which would also be unethical if the new treatment was beneficial). On the other hand twenty-five patients might not be enough to detect a statistically significant difference; even if a difference truly does exist (this is a type II error). In other words the ‘power’ of the study was too low to show the difference. To avoid costly errors when planning a trial, trialists use a nomogram (e.g., Altman’s nomogram) inputting three pieces of information: the required level of statistically significant difference (usually to 0.05 or 0.01 level); a pre-determined difference between the interventions that would be clinically relevant; and the power of the study (the risk of making a type 2 error), which the trialists choose to set at a pre-determined level (often around 80%). This nomogram then calculates how many patients are needed to show this predetermined clinical difference, at the predetermined power and level of statistical significance. This is the total number of participants needed to complete the trial. Most trialists will try to recruit more than this, to allow for drop-outs during the study period. Sample sizing software can also be used to calculate the
optimal sample sizes, and simple algorithms can be used to produce sample size estimates.

The number of participants assigned to each individual intervention will also depend on the study design. The most common trial uses approximately the same number of participants in each arm, a 1:1 trial. But, if the trialists decide that they want two participants to receive intervention A for every one participant that receives intervention B; this is a 2:1 trial – then, advanced statistical input is required and your nomogram calculations are not appropriate. Interestingly, most large trials will not have exactly the same number of participants in each arm; this is due to the use of random number sequences. However, there are methods of randomisation available to statisticians that can be used to ensure equal numbers per arm of the trial, e.g., the use of ‘blocking’.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Glossary of terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocking</td>
<td>Randomisation occurs in groups (blocks) rather than based on individuals</td>
</tr>
<tr>
<td>Intention to treat (ITT)</td>
<td>All participants are analysed in their original groups even if drop out or cross over occurs</td>
</tr>
<tr>
<td>Nomogram</td>
<td>A graph used to calculate sample size</td>
</tr>
<tr>
<td>Sample Size</td>
<td>The number of participants required to detect a true difference between two interventions, if such a difference exists in the population</td>
</tr>
<tr>
<td>Type I error</td>
<td>A statistically significant difference is found between groups which is not true for the population</td>
</tr>
<tr>
<td>Type II error</td>
<td>No statistically significant difference is found between the groups though there is a real difference in the population; this is because the sample sizes were too small for the outcome being studied.</td>
</tr>
</tbody>
</table>

CONSORT
The CONSORT statement1,2 (www.consort-statement.org, where CONSORT stands for CONsolidated Standards of Reporting Trials) was originally published in 2001 simultaneously in The Journal of the American Medical Association (JAMA), Annals of Internal Medicine, the Lancet and BMC BioMed Central, and an additional “Explanation and Elaboration Document”3 explaining how to use CONSORT. The aim of CONSORT is to standardise reporting of the methods and results of RCTs. The rationale is that if the study is reported properly, it was probably conducted properly. Currently, most journals require that manuscripts reporting RCTs follow CONSORT guidelines. These guidelines include 22 points and a flow-diagram summarising the participation of patients during enrolment, intervention allocation, follow-up, and analysis in the RCT. One of the benefits of the flow-diagram is that readers can immediately identify if there was a significant drop-out in any of the arms of the trial (maybe in the new intervention arm, as participants found the side-effects intolerable, for example). Also, it helps the reader to identify if intention-to-treat analyses were really used to analyse both the efficacy and the side-effects profiles of the interventions being studied. Intention-to-treat analyses mean that the effects of the intervention are studied in the group that was randomised to that intervention; so even drop-outs are still included in the analyses. If drop-outs are excluded from analyses, it could potentially over-estimate the treatment effect of the intervention as well as under-report the side-effects experienced.

RCT Registration
Currently, anyone hoping to conduct a RCT is encouraged to register it. This helps to ensure that the results of “negative” as well as “positive” trials are disseminated – the publication bias for “positive” trials is well recognised although, of course, it is unethical to conduct a trial and fail to disseminate the information, even if its results are “negative”. Central registration of RCTs also helps to prevent duplication of trials studying a particular intervention and helps comparison of results, in the event of duplication. As a resource for identifying both “positive” and “negative” trials, it also facilitates authors of systematic reviews to identify all trials and to produce unbiased systematic reviews. Registration of trials when they begin is a requirement for publication of the results in certain journals, e.g., The British Medical Journal (BMJ). There are two large on-line sites for RCT registration; one is run by the National Institutes of Health in the USA (www.clinicaltrials.gov) and the other is the Cochrane Central Register of Controlled Trials, run by the Cochrane Collaboration in the UK (www.cochrane.org).

A properly-conducted trial is one of the highest levels of evidence, second only to systematic reviews, in the Hierarchy of Evidence available to medical decision-makers today (i.e., you). We hope that this tutorial helps you to evaluate RCTs and to make informed decisions. Maybe in the future, you will partake in, design or conduct your own RCTs! For further information on RCTs, see the sources below.

Correspondence: CS O’Gorman1,2
1Graduate Entry Medical School, and 2Centre for Interventions in Infection, Inflammation & Immunity (4I), University of Limerick
Email: Email: clodagh.ogorman@ul.ie

References

SUGGESTED FURTHER READING.
Websites: www.clinicaltrials.gov
www.cochrane.org
www.consort-statement.org

Books:
What Are The Differences Between A Literature Search, A Literature Review, A Systematic Review and A Meta-Analysis? And Why Is A Systematic Review Considered To Be So Good?

It takes time to recognise the differences between a literature search (LS), a literature review (LR), a systematic review (SR) and a meta-analysis (MA), especially as these terms are often used interchangeably by many authors. For example, a colleague said recently that she planned to do SR as part of her background for her post-graduate research thesis. She planned to have it completed within five days. After talking to her, it was clear that she did not understand the concept (or the workload!) involved in a SR. On the other hand, we all do so-called “quick and dirty” LSs every day! Those are the kind of search where you have a question, you open up your favourite search engine (PubMed, EMBASE, etc.), plug in a few key words and press “search”. Usually, with this type of search, you only put more effort into the search strategy if the “quick and dirty” approach does not yield enough (or any) relevant articles or if you are doing the LR for your thesis, or research project.

So, what are the differences between LS, LR, SR and MA?

Actually, they can be described along a continuum, where literature search is the most basic and SR the most complex, with a MA the statistical extension of a systematic review where appropriate. A narrative (literature) review is a review of various articles, but generally lacks explicit descriptions of systematic methods to identify articles and/or failure to critically appraise them. Let us start by discussing each term.

A literature search means exactly what it implies: you search the literature to answer a question, e.g., “What is Giardiasis?” You can open the book on infectious diseases, check MEDLINE or search Google in order to come across one summary article that gives you a general idea of what Giardiasis is. In experienced hands, this will take only a few minutes and you gain some superficial knowledge. (Alternatively you could spend months searching for every article written in the medical and non-medical literature on Giardia, but you will have still only searched the literature; you may not have any more information on Giardia than when you started).

Generally, when doing a literature review, a search strategy is drawn up and one or more medical databases is used to implement the search strategy. (A search strategy incorporates the issues, such as PICOT, previously discussed in our “What’s the difference between PubMed and MEDLINE?” tutorial.) As part of a literature review, it is expected that the retrieved articles are reviewed (i.e., critically appraised) and not just superficially read.

A systematic review is a scientific investigation, a research article with pre-defined systematic methods that identify systematically articles relevant to the research question, appraise their quality, extract data and then synthesise the results of these articles. The original studies which make up a SR (including published and unpublished data, conference proceedings, abstracts, etc.) are the “subjects” of the scientific systematic review. A SR employs methods to limit bias and random error. In a SR, the results of primary studies will be summarized, but may or may not be statistically combined to give a final figure. When statistical methods are used to combine the results of two or more studies, this is called a meta-analysis.

Let’s summarise to date:

1. **Literature search:** searching the literature for some studies.
2. **Literature review:** reviewing the studies which have been identified.
3. **Systematic review:** systematically searching the literature to identify all relevant published and unpublished data in order to appraise their quality and summarise the overall findings. If the studies are homogeneous (similar), and of sufficient quality, then their results can be amalgamated into a meta-analysis in order to obtain one final result summarising all the included studies.

**Why is a systematic review considered to be the highest level of evidence?**

Simply because a well-designed SR will summarise good quality randomised controlled trials (or, increasingly these days, observational studies). Let’s go through how you could perform a SR. If you understand the concepts and steps behind performing a SR, it will help when trying to read and appraise them.

**Planning**

Firstly, you have to have a question that you want to answer. And it helps if it is your question rather than your supervisor’s question, as it is likely that you will read and appraise all the articles, while your supervisor supervises! Then, the question must be framed as a research question, in the PICOT format. For this, you should try to come up with every synonym for all the elements of your PICOT. After this, some people will design their own way of searching the literature systematically, and others will use some of the pre-validated search strategies available (e.g., search strategies for retrieving randomised controlled trials from MEDLINE® and PubMed®). When planning the search, be careful not to make it too wide (or you will have a huge number of false-negative articles: imagine searching for all the treatments of a heart attack) or too narrow (where you are likely to miss many articles that would be relevant to your question). This is the planning stage, and the more you plan, the more time you can save later (if your search is not too big) or the more likely you are to find the articles that really answer your question (if your search is not too small).

Next, you should define your inclusion and exclusion criteria. For example, do you want to include only randomised controlled double blind clinical trials, or does blinding not matter? Or, if it is unlikely that you will find RCTs, will you accept cohort or case control studies? Is there an age limit appropriate to your participants for included studies? What about exclusion criteria? Predefining your inclusion and exclusion criteria helps to prevent against bias when reviewing your articles for inclusion in your SR.

**Searching for evidence**

Once you have planned your search and inclusion and exclusion criteria, now you implement your search strategy. Different databases will be more relevant to different searches. There is overlap between MEDLINE (www.ncbi.nlm.nih.gov/pubmed) and EMBASE (www.embase.com), so performing your search in both might not add a significant number of articles. However, some
Data extraction
At this stage, all articles that have been retrieved by the search are analysed for the preliminary inclusion and exclusion criteria. In some cases, review of the title and abstract may be sufficient to exclude some articles. In other cases, the article will need to be reviewed in detail (the number of studies excluded at each of these stages are generally shown as the first figure in good reviews). Next, data extraction is performed on all studies that you think should be included. Data retrieved by the search are analysed for the search-premier). In some cases, review of the title and abstract may be performed independently by two reviewers, in order to reduce bias. Care should be taken to note the study design when extracting the data as results from different study designs are not always directly comparable.

Quality appraisal
The next step is to critically appraise the studies by comparing and contrasting studies that are included in the final systematic review. Comments should also be made on sources of (potential) bias and specific areas of interest such as publication bias, ethics, quality analysis of included studies and comments on specific areas of your results, as required— for example, analyses performed according to years of publication of studies might be appropriate if, say, a new treatment was released during that period. (Basically, you are looking for good quality studies in order to form a good quality SR).

Is it appropriate to do a meta-analysis?
MA is a statistical method of combing the results of studies included in a SR so that an overall treatment effect can be ascertained. Care should be taken with MA—strict criteria should be observed, and consultation with experts may be required, to prevent heterogeneity of included studies (i.e., to prevent comparing ‘apples’ with ‘oranges’). One of the results of MA is the generation of a “forest plot”. When interpreting a forest plot, look for the following (Figures 1 and 2).

There are several types of meta-analysis. For example, random effects models are used when there is thought to be significant heterogeneity (differences) between studies, but fixed effects models are used when there is less heterogeneity, and Poisson models are used for meta-analyses with very small sample sizes. Using the wrong method in the wrong circumstances can lead to different, and therefore incorrect, conclusions. It has also been pointed out by some statisticians that it is safer to assume ‘random’ effects in all meta-analyses, as the tests for homogeneity are of low power. Also, the effects of publication bias—where mainly positive results were published—needs to be assessed when interpreting the results.

No discussion of systematic reviews is complete without acknowledging the Cochrane Collaboration, named after Archie Cochrane, a Scottish medical researcher and pioneer of evidence-based medicine (www.cochrane.org). This is an international organization that publishes rigorous SRs and MAs, and it encourages regular updates of its published SRs and MAs. (In fact, the symbol of the Cochrane Collaboration is the forest plot from one of the first MAs; this showed that antenatal steroids improved perinatal outcome and, thus, profoundly changed obstetrical management and reduced neonatal mortality and morbidity worldwide). This is the real strength of a SR and MA: to show an overall treatment effect which is stronger than the individual treatment effects of any included trials, even when the individual included trials show no overall effect. A further cautionary note is required here, remembering that in the past there was a certain publication bias for positive results. Some MAs, conducted using studies published during periods when this publication bias was common, may have produced inaccurate results and treatments may have been changed to the detriment of patients. Large trials (RCTs) are always to be preferred, when possible. And, in the future, it is hoped that efforts to counteract publication bias (with the consequent publication of negative trials) will mean that future MAs will include inputs from trials with both positive and negative results.

We hope that this paper helps you to understand the methodology of SR and MA, and to appraise published data. This paper alone is not sufficient to train you to perform SR and MA, but if you are now inspired and would like to learn more, there are plenty of courses that would be very helpful. However, we do hope that in explaining the complexity of performing a systematic review we have helped you to appreciate why it is valued so as a research tool.

A more complicated MA using six studies to analyse the effect of an intervention (experiment) compared to placebo (control) showing no difference between the groups. Note however that two studies (O’Gorman and Macken) cross the line of “no effect” suggesting that there was no difference, two favour the intervention (Aiden and Eve) and two favour the control (Higgins and Saunders).

Box 1: Listed are some resources that you could access for more information. However, we learned...
about SRs and MAs by reading, helping to design them and doing them! An excellent resource is The Cochrane Collaboration, which regularly runs seminars on how to do a systematic review.

RESOURCES BOX
Systematic reviews in health care: meta-analysis in context\(^7\) (book)
Systematic reviews to support evidence-based medicine: how to review and apply findings of healthcare research\(^8\) (book)
Systematic reviews in healthcare: a practical guide\(^9\) (book)
www.cochrane.org/resources/handbook/\(^{10}\) (web-site)
The QUORUM statement, The Lancet 1999\(^{11}\) (instructions for reporting on SR and MA)

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

What are the Differences Between Common Statistical Tests?

MF Higgins\(^1\), AP MacKen\(^2,5\), W Cullum\(^2,3\), 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 Ar, University Hospital Limerick
\(^5\)National Children’s Research Centre, Crumlin, Dublin
\(^6\)C-Star, University of Limerick

Statistics are essential components of quantitative (and qualitative) research that we all should know. We have 12 tools through numerous lectures on the subject, and we know that we need to understand it and really should by now, even if only to not appear un-knowledgeable at journal club meetings. Still many people skip the section on statistical analysis when reading papers. Sometimes it feels that only those who have performed full time research really understand statistics - and not even then if they were lucky enough to have a statistician on their team. As part of our series of research tutorials we would now like to remove some of the mystery surrounding the art of statistics. Let’s start with the raw data….

What’s the difference between qualitative and quantitative data?
As you begin your analysis you will always have a body of raw data which you can then use to reject or accept a null hypothesis. (Remember the Null Hypothesis? It is the chance that there is no difference between the groups being compared.) Before deciding which test you are going to use, you need to first decide what kind of data you are going to collect. Data are either qualitative (e.g., colour of hair, type of job, place of birth, “qualitative” information) or quantitative (e.g., BP readings, serum bilirubin levels, birth weight: quantities, numbers). While that seems relatively easy, some people will try and confuse you by referring to qualitative data as categorical or to quantitative data as numerical. We are going to keep it simple, and we suggest that you stick with the simple subtypes and then take it from there.

What’s the difference between parametric and non-parametric data?
Remember the famous Gaussian curve of the normal distribution? If not, look at Figure 1, and it will immediately spring to mind again. A normal distribution is symmetrically distributed around the mean with a bell-shaped curve. If your data are normally distributed, then you can use tests based on the normal distribution (such as the t-test: more on this later); if the data are not normally distributed (i.e., non-parametric, or skewed) then you can either transform to normal (which is not that hard) or use non-parametric tests. Transformation means using specific statistical tools to convert “not-normally” distributed data to normally distributed data, e.g., data that are positively skewed (i.e., skewed to the right) might be transformed by getting the logarithmic of each individual data in the dataset. (However this is risky as the hypothesis being tested will also change).

What’s the difference between average, mean, median and mode?
Primary level maths taught us all the meaning of the term “average”. The mean, median and mode may be different numbers but all represent the average value of data. Essentially, the mean is the arithmetic mean (the sum of all the values divided by the number of values), the median is the middle number in a series of numbers (thus, dividing the distribution in half), and the mode is the value that occurs most often (I think of it as being fashionable, or “in mode”, so it is repeated most often). Here are a couple of examples, from this group of numbers, or raw data. This could be ages, or grammes of medication required to get an effect, or

Figure 1: Gaussian curve: normally distributed population
number of times that we had to learn statistics before we understood it.

5 6 7 9 10 11 15 16 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 ($\chi^2$) test, which compares proportions (hence its full name: the $\chi^2$ 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 is another two terms that are often confused or considered to be synonymous. Let us explain these mathematically first, with reference to Table 1.

<table>
<thead>
<tr>
<th>Table 1: Outcome One vs Outcome Two</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome One</td>
</tr>
<tr>
<td>-------------</td>
</tr>
<tr>
<td>Cases</td>
</tr>
<tr>
<td>Controls</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

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
1Graduate Entry Medical School, and 2Centre for Interventions in Infection, Inflammation & Immunity (4i), University of Limerick
Email: Email: clodagh.o'gorman@ul.ie

Further reading
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 Higgins1, AP Macken2,3, W Cullen2,3, J Saunders6, C Dunne3,5, CS O’Gorman2,5
1Maternal-Fetal Medicine, Mount Sinai Hospital, Toronto
2Graduate Entry Medical School, and 3Centre for Interventions in Infection, Inflammation & Immunity (4i), University of Limerick
4The Children’s Ark, University Hospital, Limerick
5National Children’s Research Centre, Crumlin, Dublin
6C-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:

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP</td>
<td>True Positive</td>
</tr>
<tr>
<td>FN</td>
<td>False Negative</td>
</tr>
<tr>
<td>TN</td>
<td>True Negative</td>
</tr>
<tr>
<td>FP</td>
<td>False Positive</td>
</tr>
<tr>
<td>PPV</td>
<td>Positive Predictive Value</td>
</tr>
<tr>
<td>NPV</td>
<td>Negative Predictive Value</td>
</tr>
<tr>
<td>ROC</td>
<td>Receiver Operator Curve</td>
</tr>
</tbody>
</table>

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” 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.²

**Specificity**

Again, the textbooks will tell you that “specificity = TN / TN + FP”; 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”: 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.³

**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 fibrosis⁴ 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%⁵ 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

**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 and 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/~eng/javarad/roc/ROCCFITI.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 Graduate Entry Medical School, and 3Centre for Interventions in Infection, Inflammation & Immunity (4i), University of Limerick
Email: Email: clodagh.oorman@ul.ie

References

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

CS O’Gorman1,2, AP Macken1,4, W Cullen1,2, J Saunders1,5, C Dunne2,4, MF Higgins5
1Graduate Entry Medical School, and 2Centre for Interventions in Infection, Inflammation & Immunity (4i), University of Limerick
3The Children’s Ark, University Hospital, Limerick
4National Children’s Research Centre, Crumlin, Dublin
5C-Star, University of Limerick
6Maternal-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 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 research, using two research questions for illustration: firstly, what are the attitudes of pregnant women with breech presentation to external cephalic version (ECV)? 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 1 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; less than 70% would recommend ECV, and only 78% would have an ECV themselves.9 So what do women want? This is an ideal area for qualitative research – a complex, emotive issue. Does 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 procedure5…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,5,6 haemoglobin5 and small intestinal mucosal histology8. 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 diagnosis10 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 1960s12, 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...
with IDDM, after being diagnosed with CD also? Grounded theory is both inductive and deductive\textsuperscript{12,13} – as the theory develops, hypotheses are tested and re-tested. Grounded theory starts with each patient (or “participant”) discussing how they feel about their diagnoses and dietary restrictions. The researcher records and analyses all conversations with all participants, looking for similar ideas (or “themes”). These themes are systematically and individually coded. Participants are recruited until no new themes are developed. As the theory develops, researchers might select participants, to try to follow particular elements of the emerging theory (“theoretical sampling”). Participants must be willing and able to participate, able to express their own views, be aware of their diagnoses and aware of the implications to them and of their feelings about this.\textsuperscript{15} Following the coding of data, participants are asked if they agree with the results of the coding process (“triangulation”). As themes are extracted, new themes are created. The literature is then reviewed, to help develop the theory.\textsuperscript{16} This is different to quantitative research where the literature review occurs prior to the study. The participants’ social situation is also important. Researchers might include participants’ diaries or daily dietary schedules, for example, school meals.

**Ethnography**

Ethnography evolved from anthropology and can be defined as any full or partial description of a group; “ethno” meaning folk and “graphy” meaning description.\textsuperscript{17} Its basis is the assumption that humans share enough characteristics to develop social relationships.\textsuperscript{18} Let’s rephrase the coeliac disease research question: “How might the episode of a co-diagnosis of celiac disease and diabetes congregate? Of course, adolescents with IDDM and coeliac disease attending specialist clinics all share characteristics and, so, are a society, but they might not communicate in this setting. A focus group of these individuals, prior to clinic or at a different time and location, might encourage communication. “Gate-keepers”, all those involved with the group of participants (including parents, the paediatric gastroenterology and diabetes teams) should be included in ethnography, in order to develop a better understanding of the social situation of participants.

Ideally, ethnographic researchers should be immersed in the field: this might involve living with a relevant subject for a period of time. This is called being a “complete participant”. It is obviously very difficult to do, but, when successful, it provides invaluable information. Alternatively, the researcher might be a “complete observer”, recording interviews and taking field notes. In the focus group, the researcher must be a facilitator (encouraging the input of participants) while a co-researcher records non-verbal communications and other group interactions (description of situation and participants, personal reactions, etc.). To reduce bias, pre-conceived ideas of the study should be disregarded before entering the field. As data are analysed, theories are developed and again tested against observations and participants are again asked to triangulate.

**Phenomenology**

A phenomenological study describes the “meaning of the lived experiences” for individuals relevant to a concept or the phenomenon.\textsuperscript{19} Phenomenography is a type of phenomenology where lived experience is described through writing (“graphy”). Let’s ask the ECV research question: “How might the lived experiences of individual women with a breech presentation who are offered ECV explain their perceptions, feelings and behaviour when making a decision about ECV?” As a subgroup of pregnant women, women with a breech presentation have a unique view. Using interviews or written accounts as a way of presenting the women’s own views, describing their real life experiences, exploring their perceptions and opinions and presenting it in their own words: these are the essentials of phenomenology.\textsuperscript{10} Trust is an essential component of phenomenology (and all qualitative research) as findings should reflect the reality of the experience.

To explore the ‘lived experience’, there are several possible methods of data collection. Triangulation of data collected from a combination of semi-structured interview with open questions, the opportunity to express feelings on paper, diaries, and discussion with other family members would be appropriate. Extensive interviewing is an important feature of phenomenology. Interviews require “freedom”, i.e., the researcher’s willingness to digress from the question list and explore topics introduced by the participant. Extensive and accurate field notes are required. Bracketing, i.e., disregarding any previous knowledge of the study topic, is important so that the interviewers’ views do not bring biases to their data interpretation. The interviews are transcribed and non-verbal information added to the transcripts. For data analysis and theory development, the researcher describes his own experience of the phenomenon, even if this is limited to medical information and pre-conceptions. (This is one of the differences between ethnography and phenomenology.) Then, he finds statements in the participants’ interviews about ECV experiences, develops a list of nonrepetitive, nonoverlapping data, and groups these into “meaning units” – where similar inferences are classified together (e.g., anger / rage, shame / embarrassment). A description of experiencing the phenomenon starts with each patient (or “participant”) discussing how they feel about their diagnoses and dietary restrictions. The researcher might be a “complete participant”. It is obvious that this is very difficult to do, but, when successful, it provides invaluable information. Alternatively, the researcher might be a “complete observer”, recording interviews and taking field notes. In the focus group, the researcher must be a facilitator (encouraging the input of participants) while a co-researcher records non-verbal communications and other group interactions (description of situation and participants, personal reactions, etc.). To reduce bias, pre-conceived ideas of the study should be disregarded before entering the field. As data are analysed, theories are developed and again tested against observations and participants are again asked to triangulate.

**Case Study**

In medicine, we use case studies regularly, whether discussing an interesting case at rounds or publishing in the literature. Conventionally, a quantitative case study calls for an explanation of the condition, a thorough description of the context of the problem (or case), a discussion of important elements of the case and a summary of “take-home messages”.\textsuperscript{19} In qualitative research, as opposed to a “medical” case study, multiple cases can be studied together in a single case study. In the case of the ECV question, data are collected from the woman in the form of transcribed interviews, observations, audio-visual recordings of conversations, as well as supporting documents and reports, both from the pregnant woman and those surrounding her – doctors, midwives, nurses, family, friends, etc. A wide range of expectant mothers, with different viewpoints, might be interviewed. The most important part of conducting a case study is the collection of enough information to present a detailed description of the problem under study, whether those are the attitudes of the woman to the ECV, or the attitudes of many women to ECV.

A final point is that, in some cases, a full investigation into a particular area of research can involve both qualitative and quantitative approaches. Designs which incorporate both methodologies are termed ‘mixed’ designs and it is perfectly acceptable to plan studies involving both. These designs can incorporate qualitative and quantitative research occurring simultaneously or one followed by the other in any order, and can often result in very rich and rewarding holistic research.

**Summary**

To answer either of the research questions that we posed at the start of this paper, we could choose any of the summarised methodologies, but our personal preference would be phenomenology, because it defines the “lived experiences” of a group of participants. There is generally an exhaustive amount of work involved in all qualitative research, but certainly this benefits topics where there is very little already known. In fact, this is one of the great strengths of qualitative research. This could direct future research – both qualitative and quantitative. ‘Mixed’ designs incorporating qualitative and quantitative approaches are also possible.
What is the Difference Between Deontological and Consequentialist Theories of Medical Ethics?

CS O’Gorman1-4, AP Macken1-4, W Cullen1,2, C Dunne1,2,4, MF Higgins5

1Graduate Entry Medical School, and 2Centre for Interventions in Infection, Inflammation & Immunity (4i), University of Limerick
3The Children’s Ark, University Hospital, Limerick
4National Children’s Research Centre, Crumlin, Dublin
5Maternal-Fetal Medicine, Mount Sinai Hospital, Toronto

Every day, every doctor, senior or junior, faces ethical decisions. From the moment you start seeing patients on the wards, there are decisions to make. And as every decision you make can/will have a significant impact on the lives of so many people – your patient, your patient’s family, you, other healthcare professionals - medical ethics can give you a framework to help you to approach some of these decisions. The aim of this brief tutorial is to introduce some theory to frame your practice.

Why is patient confidentiality so important? Do all patients, even children, have rights? What about those with psychiatric disease? If we only had limited funds, would we treat diabetes or cancer, or fund in vitro fertilisation for infertile prospective parents? Illustration of some ethical principles will help us try to answer these questions. Ethics is a branch of philosophy. It is based in morality and it allows us to try to differentiate right from wrong, in the framework of rules or standards of good or moral behaviour.

What is the difference between “right” and “wrong”? Should be easy to answer, right? Wrong! Some people believe that ethics is not about differentiating right from wrong, but that ethics is a matter of opinions. Some people use information from their backgrounds to differentiate right from wrong; these backgrounds can be based on the beliefs of the family with whom they grew up, or on religious, cultural or other societal beliefs.

Some base their beliefs on what they have been taught about specific ethical theories or on what they believe the majority would choose (or “what would others do in this situation?”) Regardless, each individual’s beliefs or choices are equally relevant and important. This is called ethical relativism; it allows us to tolerate other people’s beliefs and choices, without losing track of our own.

One possible definition of right and wrong could be the balance of benefits and harms – the right action is likely to lead to more benefits than harms, and the wrong action is likely to lead to more harms than benefits. This is commonly used in medicine when weighing up options.

Are “right and wrong” a bit like “pleasure and pain”? Only sort of. But there are ethical theories that can be explained using the concepts of pleasure and pain – the consequentialist and utilitarian theories.

Consequentialism (“The end justifies the means”) Here the end or the consequence is more important than the means used to achieve that end, or that an action is “right” if it leads to the “best” outcome. Of course, that depends on who and how defines “best”! This is a problems with consequentialism – it does not define which consequences are morally most important.

References

Utilitarianism, as an example, suggests that the best consequences are those in which human happiness (utility) is maximised. One of the fathers of consequentialism defined human happiness as "the balance of pleasure over pain".

For example, if we have a limited amount of money, do we choose to fund several months of chemotherapy for an adult with lung cancer or perform an elective caesarean section on a woman who has chosen this? Consequentialism tells us that the best course of action is the one in which happiness is maximised. But it is difficult to decide which of these individuals would have the maximum happiness from their individual outcomes. Each individual would value their happiness differently, as would those that have an interest in the individual's outcome.

Deontological (or "duty-based") theory ("The measures must be just")
Here the best choice is defined by the methods that must be followed to achieve an outcome – not by the outcomes themselves. This is where deontology and consequentialism differ. The deontological theory believes that, in any given situation, some acts are ethically and morally "wrong" and not acceptable, even if they are supposed to lead to the desired outcome. So, euthanasia would be considered wrong, as it is an active killing, even though the aim is to ultimately relieve suffering. Or, even if we know that giving nightly growth hormone injections to a child with short stature will help that child’s growth, and even if the child’s parents want the treatment, if the child objects to the treatment, then deontology tells us that to proceed with treatment is ethically morally "wrong". But consequentialism tells us that if that child achieves an acceptably normal height, then the action of injections to which the child objects, is justified because we achieve the desired outcome. (Of course, this does not touch on the ethical question of whether or not we should agree to "standardisation" of children’s heights.)

The four principles of medical ethics
Autonomy: This is respect for individuals, their rights and requests. This is why doctors are obliged to maintain confidentiality – because the information belongs to the individual patient. Autonomy also tells us that patients have to be allowed to come to their own conclusions; and doctors can support this by providing relevant important information.

Beneficence: This is the pursuit of the outcome that is best for the patient. This principle deals with doing good to others, or doing good for your patient. Generally, the patient and the doctor both have the best interest of the patient as their desired outcome. Problems arise when the expressed desired outcome of the patient and the doctor are not the same.

Non-maleficence: This is the pursuit of not doing harm to the patient. This is not always the same as beneficence. Particularly if the desired outcome is only achieved by a method that causes some harm or distress to the patient.

Justice: This relates to the allocation of limited resources. Justice makes two particular points: firstly, each individual is entitled to the same resources; secondly each time a patient accesses resources, this impacts on other patients, to whom this resource is no longer available. Justice is how we distribute limited resources in an ethical and moral manner.

What is an ethical problem?
Because ethical decisions can have so many potential outcomes, and because it is difficult for us to guess which outcome is "most right", ethical decisions can lead to "moral distress". And following making a difficult ethical decision, any remaining uncertainty is called "moral residue".

Do all individuals have the same rights – children, adults, elderly, those with psychiatric disorders, those with or without access to education or medical insurance? Some ethical theories believe that all individuals have the same rights; some ethical theories do not.

How does this help us to approach an ethical problem?
We suggest starting with an attempt to define the ethical problem and the ethical principles affected by the problem. Then, gather the background to the problem: consider any person that might be affected by the problem – the patient, their family, parents or guardians. Consent may be required before divulging confidential information. Then discuss the problem, looking for advice based on opinion or experience – with your peers, your seniors, doctors, nurses. It may be helpful to enlist the help of a bio-ethicist or a legal advisor also. After identifying potential solutions, consider the various action, choose one and implement it. We discuss and review the progress and outcome regularly with the patient (and/or family members or guardians). The details of an ethical problem may change and evolve constantly: sometimes, the decision will need to be reversed and a new action chosen and implemented. We also realise that we frequently need help with ethical decisions – to decrease our own moral distress and moral residue! And to feel reassured that we are helping to choose the best course of action for our individual patient and individual problem.

Sounds hard? Every hour of every day medical staff go through this process. Think of the ethics of resuscitation at the edges of viability. Obstetricians have to discuss whether to monitor the heart rate of a 22/23 week foetus whose mother is in threatened labour. Simultaneously, neonatologists must consider whether this baby should undergo full resuscitation, which may lead to a prolonged NICU stay with high likelihood of mortality or profound morbidity. If this baby remains in utero for even 1 week, then the potential clinical outcomes change significantly (with the advancing gestation), and both the obstetric and paediatric clinical teams need to re-evaluate their ethical decisions and processes.

Every day of your medical career, from very junior to very senior, you are likely to encounter ethical problems that affect many people. We hope that this paper helps you to create a framework that you can use to try to answer some of these questions. So often, it seems that there are more questions than answers!

Correspondence: CS O’Gorman1,2
1 Graduate Entry Medical School, and 2 Centre for Interventions in Infection, Inflammation & Immunity (4i), University of Limerick
Email: Email: clodagh ogorman @ ul.ie.
How to Teach Practical Skills in Medicine: Out of Hospital Training

MF Higgins¹, AP Macken²,⁵, O Coyle²,⁴, W Cullen²,³, D McGrath², CS O’Gorman²,⁵

¹Maternal-Fetal Medicine, Mount Sinai Hospital, Toronto
²Graduate Entry Medical School, and ³Centre for Interventions in Infection, Inflammation & Immunity (4i), University of Limerick
⁴The Children’s Ark, University Hospital, Limerick
⁵National Children’s Research Centre, Crumlin, Dublin

*See one, do one, teach one* is the traditional paradigm for teaching medicine while working, the apprenticeship model. This paradigm is based on training during long working hours and with evaluation by mentors¹. More recently, medical education is turning towards more structured programmes of teaching skills, where formal training can be objectively assessed using competency-based assessment². At an undergraduate level this is driven by the requirement of a newly-qualified doctor to be familiar with basic competencies required for clinical work; these competencies are often assumed by other members of the healthcare team and are desired by the undergraduate students themselves. In fact, students themselves have requested training in particular practical techniques such as venepuncture, catheterisation and suturing in order to better prepare themselves for the practicalities of working life³.

Changes in methods to achieve competency in practical skills in postgraduate medical education have been driven by several factors. Firstly, the introduction of the European Working Time Directive (EWTD) has reduced the working hours of junior doctors and thus the number of procedures performed by trainees and thus decreased the emphasis on the apprenticeship model⁴. Secondly, there are increased requirements to assess skills based on competency⁵. Thirdly, new procedures (such as laparoscopy) have been introduced so quickly that all grades of doctors have needed to be trained at the same time¹. Many new techniques for teaching practical clinical skills have just been introduced recently. Therefore trainers who themselves were taught using the “see one, do one, teach one” paradigm are now the postgraduate teachers of students who have used — and are therefore familiar with — the newer methods of teaching.

This paper is a literature review of the evidence in the area of teaching practical techniques in medicine. The first article reviews the research on skills laboratories and simulation and the second concentrates on training in direct contact with patients.

Skills labs
The aim of a clinical skills laboratory is to allow students the opportunity to practise practical procedures in a safe, non-stressful environment, where procedures can be broken into a number of steps in order to improve understanding. From its beginnings in eighteenth century France where Madame Du Coudray used fetus and pelvis models to train midwives, clinical skills laboratories have expanded to utilise many varieties of media. Animals models provide living simulations⁶ but may raise concerns about moral issues, cost and infections⁷. Virtual reality techniques raise interesting possibilities⁸ but are expensive and may not be accessible to all. More commonly, manikins, synthetic tissues, trainers or skill stations have been used to teach both basic and more advanced practical skills.

Manikins
Manikins are of value in training a large number of students in a variety of skills procedures. Studies have assessed the use of manikins in endotracheal intubation⁹ and in the assessment and treatment of an acutely ill patient using a Laerdal SimMan™. Other simulators may be used to teach uncomfortable procedures such as otoscopy, vaginal delivery, catheterisation, colonoscopy, bronchoscopy and digital rectal examination. The generic components of these practical sessions with manikins include: breaking the skill down into individual steps, learning on simple scenarios before moving onto more challenging clinical scenarios (cognitive based learning), limitations of time to allow for maximum concentration and low teacher to student ratios. In all these scenarios, students can use the manikins to practice team work in assessing and treating these acute emergencies. In the study on intubation, after a single session, 93% of the 115 students reached the required standard to attempt intubation on a patient and feedback from students was very positive.

Simulators
Simulators permit practice to achieve mastery both techniques and instruments used in laparoscopic surgery within a controlled environment. The user-friendliness of such models for novice surgeons is evidenced by the continuing popularity of courses using these models in teaching surgical skills.

Skill stations
Skill stations have been successfully used in undergraduate and postgraduate medical education to teach practical skills. One Canadian study assessed skills training of fourth year medical students in teaching basic surgical practice at the start of a surgical rotation. Here 69 students were taught such practices as scrubbing, gowning, gloving, aseptic technique, suture cutting and instrument handling. Students taught using skill stations within a one-hour station by a surgeon and nurse educator scored higher in assessment than those who were not exposed to the teaching module. Student feedback ranked the teaching module highly in terms of value, contribution to knowledge and increased confidence in technical skills¹⁰. In addition, students in the teaching group had improved post-test scores compared to both their own pre test scores and to the non-teaching group. A similar study performed in the Netherlands used a short (2 hours) course to teach skills to qualified general practitioners. These skills included shoulder injection, cervical smear taking and ophthalmic evaluation in diabetes. Competence in the skills was measured by a knowledge test. After the course, a significant positive effect on performance in practice was found for both cervical smear taking and shoulder injection¹¹.

Interaction with other teaching methods
Skills labs may also be used in conjunction with other methods of teaching: one example in teaching neonatal examinations and procedures (Apger scoring, assessment of gestational age, oxygen therapy) used a combination of an interactive multimedia programme on CD-ROM followed by practice in a skills laboratory¹². Students also found the CD to be useful for revision purposes. On their own, computer assisted learning tools are not as useful as they lack immediate feedback; one study showed that medical students learning how to tie a knot were more effectively taught using a lecture and feedback seminar than by a CD alone¹³. Notwithstanding, a definite advantage of interactive computer programs is the ability to deliver such training in many different languages.

Multidisciplinary team skills labs
Interestingly, a multidisciplinary skills lab has also been developed to allow medical and nursing students to learn how to work both individually and as a team, in order to integrate their learning at an early stage. In one example, a simulated patient takes the role of a patient being admitted for hemicolectomy with four students (medical and nursing) sharing the tasks required to admit the patient to the ward, assess him for surgery, consent him and discuss postoperative recovery while being observed by a general surgeon and stoma therapist. After the shared exercise, the
How to Teach Practical Skills in Medicine: Bridging the Gap from the Course to the Patient, and Teaching on the Job

CS O’Gorman,1-5 AP Macken,1-5 O Coyle,1-5 W Cullen,1-5 D McGrath,1-5 MF Higgins

1Graduate Entry Medical School, and 2Centre for Interventions in Infection, Inflammation & Immunity (4), University of Limerick
2The Children’s Ark, University Hospital, Limerick
3National Children’s Research Centre, Crumlin, Dublin
4Maternal-Fetal Medicine, Mount Sinai Hospital, Toronto

One of the disadvantages of clinical skills laboratories is the lack of “real life” scenarios which might bridge the gap between the simulated laboratory and clinical settings. While technical skills are important in learning a practical procedure, effective communication with a patient is essential in order to competently complete the procedure. Taking “blood” from an orange is one thing; drawing blood from an 80 year old woman with dementia at 9am is another experience entirely. Various techniques have been developed which bridge the gap between the clinical skills laboratory and clinical settings, usually using simulated patients (SPs) or a simulated environment.

Simulated patients

Simulated Patients (SPs) are defined as “actors trained to provide a consistent performance of a clinical role and to offer structured, learner-centred feedback”. SPs can be used in a variety of settings. Firstly, SPs may be used in skills labs, and there are now many examples of these. One purpose-built suite in St Mary’s Hospital, London is described in published literature. Here ceiling mounted video recorders and playback equipment explored how both teachers and learners responded to the use of simulated patients in the teaching of practical procedures such as suturing or catheterisation. In one example, simulation of wound closure used a pad of simulated skin attached to the “patients” arm and covered with a drape in order to simulate a real wound 2-3.

Simulated environment

Alternatively, skills labs may be developed to simulate a clinical environment: simulating a domestic environment to mimic a home visit by a general practitioner, a simulated accident and emergency to run resuscitation procedures (with the addition of appropriate equipment and tape recordings to further add reality) or a theatre “scrub area” to teach students how to scrub, gown and glove. The disadvantage of using simulated patients or environments is that, although students perceive the procedure to be more realistic, they are still aware that the “procedures” are carried out in a non-clinical environment.

In response, one group developed a quasi clinical scenario using a portable recording device to record students performing procedures on SPs in the Minor Procedures Room in the Accident and Emergency Department of a hospital1. The recording device was linked to two miniature cameras mounted on a drip stand, providing different views of the procedure. Student assessment...
was positive ("you always know it is simulation, but much better than just models" 'it's an intermediate step, bridging the gap between clinical skills lab and seeing a patient") and feasible (preparation time 20 minutes, time to remove equipment 10 minutes, though in case of clinical need equipment could be relocated in one minute).  

**Further simulation**  
An alternative method to bridge the gap between the clinical skills lab and real life is to use actors more directly as the patients themselves. Pelvic examination is traditionally taught on pelvic models or, with consent, on patients in clinics or under anaesthesia. Teaching associates, in contrast, are "women trained to teach pelvic examination while themselves being examined" usually "working in pairs with one acting as patient and the other as instructor". Teaching associates have been used in the US, Canada and Australia, are acceptable to medical students, and are an effective method of teaching. A UK study showed that students using this method to teach speculum and bimanual pelvic examination scored higher in objective assessment than those taught traditionally, both on communication and technical skills.  

While some might question the ethics of recruiting and supporting (or paying) teaching associates, another variation of this theme has emerged with the concept of the "expert patient". The National Health Service (NHS) in the UK has encouraged the development of the expert patient scheme in the management of chronic illnesses. Here patients, with the diagnosis of a chronic illness such as diabetes or arthritis, can run courses in management of the chronic disease following training as lay leaders. The hypothesis is that the lived experience of the disease is as valuable, if not more so, than the knowledge of a medical professional in some aspects of the disease. (Ref). In medical education the expert patient has been evaluated in the education of students of physiotherapy. One theme that emerged was that the students were anxious about their role in this session and what to take from a relatively unconventional educational session. Education on the lived experience of a disease would provide valuable exposure to the realities of the disease that may not be appreciated from routine learning opportunities.  

**Teaching of practical skills within clinical practice**  
Sir William Osler taught to "have no teaching without a patient for an instructor". This holds true even into the 21st century. At a certain point all students will have to leave the safety of the teaching laboratories (simulated or not) and venture into the wider world of clinical medicine. The learning environment is one of the most important features to facilitate learning and dictates how teaching is perceived within the clinical area.  

**Figure 1 Peynons four steps to teaching a practical skill**  

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Demonstration</td>
<td>Instructor demonstrates the skill at normal speed.</td>
</tr>
<tr>
<td>2. Deconstruction</td>
<td>Instructor demonstrates the skill by breaking it down into simple steps.</td>
</tr>
<tr>
<td>3. Formulation</td>
<td>Instructor demonstrates the skill while being &quot;talked through&quot; the steps by the student.</td>
</tr>
<tr>
<td>4. Performance</td>
<td>Student performs the skill and describes the steps.</td>
</tr>
</tbody>
</table>

**A culture of training**  
It is important when teaching practical skills in clinical medicine that a "training culture" exists which can support both trainers and trainees in their aims. Clinical work has many barriers to teaching. The development of a training culture creates an educational environment, which can motivate both the trainer and trainee and maximise their educational opportunities. One example is that of the cardiovascular unit in Papworth hospital in the UK, where the working paradigm is "whenever there are opportunities to train, you train". Practically, this means that on the job training occurs actively on every level from junior to senior trainee, and safe, graduated practice leads trainers through progressive steps of learning. These learning steps are based on the theory that the "best performance is achieved by the combination of an objective a little further away than one thinks one can achieve combined with a relentless expectation from above that one will achieve it". This is a variation of Peynons theory of how to teach practical skills (Figure 1).  

**Multi-disciplinary team learning in a clinical setting**  
Since 1996 the Faculty of Health Science in Linkoping (Sweden) has run a training ward in an eight bedded Orthopaedic surgical unit which is permanently staffed by one doctor and one nurse. A multidisciplinary team of students (1-2 medical, 2-3 nursing, 1 physiotherapy, 1 occupational therapy and either 1 community care or 1 medical laboratory technology student) is assigned to the ward every 2 weeks, and this team works in shifts throughout the day and night. The aims of this mandatory assignment are to simulate and develop multidisciplinary co-operation, team work and knowledge of different professional competencies and skills.  

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

**References**  