Critical Appraisal

from Papers to Patient

A Practical Guide

DUNCAN BOOTLAND • EVAN COUGHLAN
ROBERT GALLOWAY • STEPHANIE GOUBET
EMILY MCWHIRTER

CRC Press
Taylor & Francis Group

Critical Appraisal
from Papers to Patient
A Practical Guide

Critical Appraisal
from Papers to Patient
A Practical Guide

DUNCAN BOOTLAND • EVAN COUGHLAN
ROBERT GALLOWAY • STEPHANIE GOUBET
EMILY MCWHIRTER


CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Printed on acid-free paper
Version Date: 20161010

International Standard Book Number-13: 978-1-4822-3045-1 (Paperback)

Library of Congress Cataloging-in-Publication Data

Names: Bootland, Duncan, author. | Coughlan, Evan, author. | Galloway, Robert, author. | Goubet, Stephanie, author. | McWhirter, Emily, author. | Title: Critical appraisal from papers to patient : a practical guide / Duncan Bootland, Evan Coughlan, Robert Galloway, Stephanie Goubet, and Emily McWhirter. Description: Boca Raton : CRC Press, [2017] | Includes bibliographical references. Identifiers: LCCN 2016026644 (print) | LCCN 2016027455 (ebook) | ISBN 9781482230451 (pbk. : alk. paper) | ISBN 9781482230475 () Subjects: | MESH: Evidence-Based Medicine | Test Taking Skills | Statistics as Topic | Examination Questions Classification: LCC R723.7 (print) | LCC R723.7 (ebook) | NLM WB 18.2 | DDC 616--dc23 LC record available at https://lccn.loc.gov/2016026644

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Contents

Foreword

Professor Sir Cyril Chantler, a paediatrician and former chair of the Kings Fund, once told a parliamentary patient safety committee that "Medicine used to be simple, ineffective and relatively safe." Today of course the reverse is true. As it has grown in capability, medicine has also become more complex, expensive and dangerous.

A century ago, medicine was a thing of fine craft, delivered by individuals guided as much by dogma as anything else. The efficacy of interventions was determined largely by the response of individual patients to discrete interventions and an overvalued sense of the causal connection between the two. Today we know better. We understand our treatments and their consequences as a thing of statistical probability.

We improve our outcomes largely through marginal gains, finding our survival advantage in small fractions has spread across large populations. And as a result, for the inhabitants of high-income countries, life has never been safer or longer lived. This is a far better world but, consequently, one in which the efficacy of our interventions is rarely obvious at the end of the bed. In medicine today, the benefits we deliver and the harm we inflict are largely invisible.

In this world we save lives by knowing the evidence offered to us in published research, by not being seduced by poorly or falsely constructed arguments, by designing studies that can tease apart the differences in our outcomes and by understanding the limits of the foundation of knowledge on which we base our decisions.

Today, a working knowledge of statistics and the ability to appraise the medical literature intelligently is as essential to your practice as a knowledge of gross anatomy. This text represents a brilliant primer and the tools and skills it equips the clinician with are designed to help them navigate the uncertainties of medical practice today.

Kevin J. Fong

Authors

Duncan Bootland, MBBS, BSc, FRCEM (@dbootland)
Emergency Medicine Consultant, Brighton and Sussex University Hospitals NHS Trust
Duncan Bootland trained at University College London and has worked in south-east England since then. His specialist training was in Emergency Medicine and Intensive Care, as well as working for Kent, Surrey, Sussex Helicopter Emergency Medicine Service. He was appointed an emergency medicine consultant in 2012. Bootland has an excellent reputation in teaching, especially for the Royal College of Emergency Medicine examinations and in critical appraisal courses.

Evan Coughlan, MB Bch NUI, MRCSEd, DSEM, FRCEM, DipTox (@Evpco)
Emergency Medicine Consultant, Brighton and Sussex University Hospitals NHS Trust
Evan Coughlan qualified in Ireland and finished his registrar training in the United Kingdom. He was appointed as a consultant in 2010. His interests are in ultrasound, trauma and toxicology. He is a passionate and established teacher of critical appraisal courses. He has also written chapters for several textbooks.

Robert Galloway, MBBS, BSc, MRCP, FRCEM, PGCMedED (@drrobgalloway)
Emergency Medicine Consultant, Brighton and Sussex University Hospitals NHS Trust, and Honorary Senior Lecturer and Sub-Dean, Brighton and Sussex Medical School
Robert Galloway trained at St George's Hospital Medical School, London. He dual trained in Emergency Medicine and Intensive Care and was appointed an emergency medicine consultant in 2011. He has a passion for education and has completed a PGC in medical education. He is a Honorary Senior Lecturer, year five sub-dean and leads for Emergency Medicine undergraduate education at Brighton and Sussex Medical School. He has interests in a wide range of academic topics including resuscitation and human factors. He teaches critical appraisal courses with the aim of making them relevant to practicing clinicians.

Stephanie Goubet, MSc
Medical Statistician, Brighton and Sussex University Hospitals NHS Trust and Brighton and Sussex Medical School
Stephanie Goubet came to the Brighton and Sussex University Hospital National Institute for Health Research Clinical Research Facility from the Department of Primary Care and Population Health, University College London. While there she was a member of their Evidence-Based Medicine workshops. She shares her time between Brighton and Sussex University Hospitals NHS Trust and Brighton and Sussex Medical School.

Emily McWhirter, PhD, BSc, RN (@McWhirterEmily)
Director of Performance and Governance, Kent Surrey and Sussex Air Ambulance, and Research Fellow, University of Brighton

Emily McWhirter trained as a nurse at St Thomas' Hospital, London, and later as a midwife at the Nightingale Institute at King's College University. She completed a PhD in 2011 at the University of Brighton to understand the changing role and identity of the contemporary ward manager. Her research interests centre around understanding the connections between the patient experience and clinical outcome.

Introduction

To get from A to B, you only need to know how to drive your car – not how the engine, automatic braking system and gearbox work. If you felt like you had to have this type of understanding, you would not bother driving in the first place.

It is the same when it comes to reading papers; you do not need to understand how they obtained the complex statistical results, but merely what the results and statistics mean, whether they can be trusted and if the study should influence your practice.

Being able to read a journal paper, understand critical appraisal and practice 'best available' evidence-based medicine is a vital skill for all those who work in health care. Integrating this with the knowledge of available skills and resources, along with the patient's preferences, allows us to provide the quality care that our patients deserve.

Evidenced-based medicine (EBM) does not mean that we should only use medical treatments based on high-quality randomized controlled trials (RCTs). That would be madness; there has never been an RCT on whether removing an appendix for appendicitis actually saves lives – we know it does from the best available evidence, in this case, case series. But in order to provide the best possible care for our patients, we need to practice best available EBM, whatever the 'level' that evidence is. In order to do this, we need to understand critical appraisal.

Despite its importance in all fields of medicine, critical appraisal is misunderstood by many, if not most, health care professionals and this is to the detriment of our patients.

Traditionally, math experts who know a lot about how to derive the normal distribution, the math behind chi-squared calculations and how to calculate the exact Yates' correction coefficient have taught critical appraisal and medical statistics. And sometimes confusion is created, not because they do not understand it, but because they understand it too well and do not articulate the clinical relevance and what clinicians need to know.

We think a new approach is needed, hence why we have written this book and started to put on courses in this area. We are not experts; we are clinicians who need a working knowledge of critical appraisal in order to appropriately influence our practice. We know what is needed to read a journal article, understand the results, whether to believe the results and whether we should implement the results with our patients.

When it comes to p-values, confidence intervals and normal distributions, we do not know the formulas to derive them, but we know their importance and how to interpret the results. And that is a good thing because it is all you need to know when reading a paper as a clinician.

We want this book to take you on a journey: from frightened by the thought of a confidence interval, and so never looking at a journal article, to being able to enjoy reading and understanding articles (and not just the abstracts) speedily, while deciding whether they will change your practice.

We will not be teaching you how the engine and gearbox of critical appraisal work, merely how to drive. That is all you need, and that is all you want. After reading this book, you will be able to trust your own assessment of papers and not be influenced by the sandwiches and cheap pens of reps, however alluring their smile.

The book takes a stepwise approach to critical appraisal, starting with why we need to bother with critical appraisal, moving to the types of papers and research you may encounter, and ending with how to dissect the papers.

Throughout the book, the main text contains the essential information, with 'Nerd's Corner' boxes for those who are interested and want to get a deeper understanding. There is also a Glossary of Terms for easy reference and revision.

For those who have critical appraisal exams coming up, there is guidance on how to approach and pass these, including practice papers.

We hope you enjoy this textbook as well as find it a valuable resource for passing exams, writing essays and, most importantly, treating patients in the best possible way.

For further information on our associated courses, please go to the website www.medicalcriticalappraisal.com.

Thanks for reading.

Section 1

Introduction to Critical Appraisal

1 Types of Papers and Grades of Evidence

It is nearing the end of a Friday evening in the medical assessment unit, and lying on the trolley in front of me is a 47-year-old man who has attended complaining of breathlessness and a sharp left-sided chest pain. I am tired and want to go home, but even as a consultant, I am made anxious by the knowledge that chest pain is one of the 'banana skins' of our profession.

As much as any other, this situation reflects the challenges we face many times a day: What clinical information do I need to glean from my history and examination? What clinical decision tools should I use? What diagnostic tests should I subject my patient to and how confident can I be in them? When I finally make the diagnosis, what is the best treatment?

The ability to interpret the evidence and the guidelines that already exist and the new ones that are published every month, and to do this in a way that leaves enough time to actually see some patients, is challenging. We need to remember that our desire to know how to treat our patients better should drive our aim to understand evidence-based medicine (EBM), and then an understanding of EBM should drive improvements in patient care.

Before we begin trying to work out how to dissect various papers, we need to understand how research is divided into different types, and which types are best for answering which questions.

TYPES OF PAPERS

Reading through the medical literature, it will not take you long to realize that there are a number of different types of studies out there – the EBM world is not just randomized controlled trials (RCTs), diagnostic studies and meta-analyses.

The way different study types are grouped together depends in some part on whom you speak to, but we find the following to be useful in ordering our thoughts when it comes to the different types. Think about three questions:

1. What sort of environment (a clinical laboratory, a highly controlled but still clinical setting, a normal ward or general practitioner's surgery, etc.) is the trial being performed in?
2. What sort of results are the researchers interested in? Is it a result like mortality rate or blood pressure which will provide a numerical value, or is it a subjective feeling or description of outcomes and emotions which cannot be reduced to a numerical value (qualitative research)?

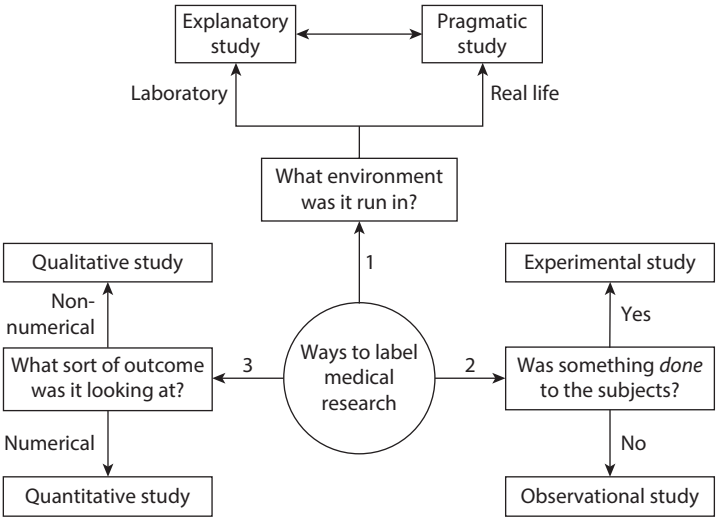

FIGURE 1.1 Different ways to label a trial – three questions to ask.

3. Are the researchers planning to do something to the patient, for example, give them a new drug or do a new diagnostic test, or are they just observing what happens to them?

The answers to these three questions should then allow us to describe the trial as explanatory or pragmatic, qualitative or quantitative and experimental or observational (Figure 1.1). Remember that these labels are not exclusive, and so it is possible to have, for example, a pragmatic, qualitative, experimental trial. The labels simply allow us to begin to order our thoughts about what has gone on within the trial.

Question 1: What Sort of Environment Was the Trial Run In?

As a rough rule clinical trials can be run in two sorts of places: highly controlled settings such as research laboratories (often with carefully selected, specific patients) or actual clinical settings that are more prone to the normal variations in atmosphere, patients, time of the day and staffing levels. Explanatory research is conducted in the more highly controlled environment and gives the best chance of seeing if the intervention actually works if as many as possible outside influencing factors are taken out of the equation. Pragmatic research, however, is performed in a setting more like a clinical environment (or actually *in* a clinical environment) and gives a better idea of whether the intervention will work in the real world. Clearly there are a variety of settings in which, and patient types on whom, studies can be performed. It is wise to consider the type of study as being on a spectrum between the explanatory and pragmatic; the more tightly controlled the conditions of the study, the closer to being explanatory and vice versa.

Question 2: What Sort of Results Are the Researchers Interested In?

Whether the result that is generated is a hard clinical number such as blood pressure, height or mortality rate or is instead something less quantifiable like sense of well-being or happiness is our second way of dividing up research. Quantitative research will generate numbers such as difference in blood pressure or mortality, which can then be analyzed to produce the statistical results that we talk about later in the book and which form much of the published results from medical literature. Studies that produce a result given as a description which cannot be boiled down to numerical data (such as opinions, explanations of styles of working and why people do not follow best practice, emotional well-being or thoughts and feelings) are termed qualitative research and can be harder to succinctly analyze and summarize. However, there are established ways of looking at this type of paper to ensure that it is as rigorously conducted as quantitative research. See chapter 7 for information on this increasingly important subject. However, in many studies looking at what you might think of as qualitative outcomes (quality of life being a good example), scoring systems have been developed so that a quantitative result can be recorded (and hence analyzed in a statistically clearer way).

Question 3: Are the Researchers Planning to Do Something to the Patient?

The final way for us to define the type of research is by whether the researchers intervene in the patients' care in any way. In RCTs and in diagnostic studies (both of which we talk about in detail later), but also in other types of study, we take a group of subjects or patients and apply a new treatment, intervention or diagnostic test to their care. These sorts of studies are termed experimental studies. Studies where the researchers look at what happens to a group of subjects or patients but the care that they receive is influenced only by normal practice (and normal variations in practice) and not by the researchers are called observational studies. Surveys, case reports, cohort studies, case control studies and cross-sectional studies are all observational studies, whereas RCTs and diagnostic studies are both experimental studies.

There are a number of different kinds of each type of study, whether experimental/observational, pragmatic/explanatory or qualitative/quantitative. A description of the main features of the important types of studies can be found in the Glossary, and an overview is given in Tables 1.1 and 1.2. The type of study that the researchers perform should be decided before they start and should be based on the strengths and weaknesses of each type of possible study.

Finally, we need to mention secondary research. Although single papers can be highly valuable, often greater value can be derived from combining the results or knowledge from a number of papers. Results of different studies can be brought together in review articles, systematic reviews or meta-analyses, all of which we discuss later in the book.

TABLE 1.1
Different Types of Experimental Studies

Randomized controlled study (RCT)	The study group is divided into two or more groups by a randomization process. One group is given one intervention, and the other group(s) another. The outcomes in the different groups are then compared. The randomization works to balance and thereby remove the influence of confounding factors. The RCT is considered to be the best sort of experimental study and commonly used in many examinations. They are, however, normally expensive to run.
Diagnostic study	The diagnostic test is performed on the whole study group alongside the gold standard test. The performance of the diagnostic test under investigation is then compared with the gold standard test. Diagnostic studies appear less commonly than RCTs in the literature but appear commonly in many examinations.
Crossover study	The study is designed to compare two interventions (usually drugs). Instead of one group receiving one drug and the other group the second drug, both groups receive one drug for a period of time, followed by the alternate drug for a period of time. The difference between the two groups is the order in which they receive the two drugs. Essentially, each patient is acting as their own control. Importantly, the drugs being studied must not interfere with the process of the underlying disease and the disease must be unlikely to alter in nature during the time course of the trial.
Cluster study	This is another form of trial that compares two or more interventions. Instead of individual patients being randomized to the different arms of the study, groups of patients are randomized. An example is all emergency department (ED) patients from one hospital being randomized to having tea and coffee provided free of charge in the waiting area, and all ED patients in another hospital not receiving that intervention.

GRADES OF EVIDENCE

The main aim of the EBM is to deliver study results that further our understanding about how to diagnose and treat patients. As thousands of studies are published every week, we need a way to judge both the importance of the results and the strength we should ascribe to the recommendations given in the paper's conclusion (or indeed to any guideline produced by an august professional body). Although there are variations in how different people rate different types of study and how strongly they make their recommendations, most follow a similar pattern to that shown in Tables 1.3 and 1.4.

Often people find the ranking of evidence and recommendations a little confusing. Although it is fairly standard that an RCT is considered of more value than a cohort study, and the grading system reflects that, remember it is only a tool to help us when summarizing the evidence that we are reading. To give one famous example, the British Doctors Study was a cohort study performed in the 1950s that

TABLE 1.2
Different Types of Observational Studies

Case control studies	A group of patients with an outcome variable (like the presence of a particular disease) are compared with a group without that outcome. The medical and social histories of both groups can be compared to look for preceding risk factors that may have made the patients with the outcome more susceptible. Case control studies are by definition retrospective and therefore suffer from recall bias. They are good for looking at patients with rare diseases. They are not very good at looking at exposures to rare risk factors.
Cohort study	A group of patients with an exposure to an intervention or risk factor are followed in time and compared with those without the exposure to that intervention or risk factor. The researchers do not alter what exposure any patient gets, and the variation between the two groups is just due to normal differences in care or exposure. The outcomes of both groups are then compared. Although these are normally prospective, if a large database of patients has been created for something else, the database can be used retrospectively to perform cohort studies. Cohort studies are valuable as they allow us to look at a number of different outcomes in relation to one exposure and the time sequence of events can be assessed. Problems include being potentially costly to run and that you need a large sample size if you want to look at rare outcomes.
Cross-sectional studies	Disease and exposure status of a population are studied at the same time. They are good for establishing prevalence or association. Unfortunately, a large number of subjects is normally needed, and they cannot be used to demonstrate causation.

TABLE 1.3
Ranking the Types of Study

Grade	Type of Paper
1	Systematic review (including at least one RCT) or a single good-quality RCT
2	Studies without randomization
3	Well-designed, controlled observational studies such as cohort studies or case control studies
4	Observational studies such as case series or case reports
5	Expert opinion

TABLE 1.4
Ranking the Strength of the Recommendation Made from the EBM Available

Grade	Usual Basis for Strength of Recommendation
A	Based on meta-analyses, systematic reviews or randomized controlled trial
B	Based on level 2 or 3 evidence (see above)
C	Based on level 4 evidence
D	Based on expert advice

proved to be a major factor in demonstrating the link between smoking and lung cancer. If you applied Tables 1.3 and 1.4 literally, it would only be considered level 3 evidence and a grade B recommendation not to smoke. From this example you can see that there are situations where an RCT is neither the best nor only way to provide a world-changing result.

Take-Home Message

1. Studies can be described as explanatory or pragmatic depending on whether the trial was run in highly controlled conditions or real-life conditions, experimental or observational depending on whether the researchers intervene in the care of the patients, and qualitative or quantitative depending on whether the results are the patients' descriptions of feelings or numerical data.
2. There are a number of different study designs for both observational and experimental studies, and which ones the researchers choose will depend on what they are studying and the type of result they are looking for.
3. The different types of papers are graded by how valuable within the EBM world they are, with meta-analyses and RCTs at the top.
4. The strength of recommendation that is made by groups such as the National Institute for Health and Care Excellence (NICE) or Royal Colleges based on the available EBM is usually also graded, with recommendations made on the basis of results from meta-analyses and RCTs considered the strongest.

2 An Approach to Appraising Papers

A very simple approach when reading papers is to think about the following questions:

1. Is the question they are trying to answer relevant to your practice and patients?
2. Are the results significant (clinically and statistically) and if so, should this affect how I treat my patients?
3. Is it of high enough quality that I can trust this paper (internal validity)?
4. Is it relevant to the patients whom I see (external validity)?
5. If the answer to all of these is yes (and the side effects of the treatment/diagnostic test are acceptable), then how do we implement the findings?

With this is mind, you can now think about dissecting individual papers. Papers are not single units; they consist of multiple parts. In this chapter, we look at what these parts are. This chapter refers predominantly to interventional trials, but other papers follow a similar approach. The specific details of how to assess each type of paper are given in the specific chapters.

The purpose of research is to answer a clinical question, which we all meet in our day-to-day clinical practice. For example, in a patient with a wrist fracture, we might ask ourselves a number of questions: Which is better to make the diagnosis, x-ray or ultrasound (diagnostic study)? Is a haematoma block or Bier's block a better anaesthetic (therapeutic study)? To answer this we would need to perform a literature review, and if we find there is insufficient evidence to choose one over the other, it may prompt us to do a study.

In a study, we enrol a sufficient quantity of patients, and once we have the raw data we use statistics to present the evidence and decide whether one is better than the other. We can then answer our question in light of our results plus any previous research. We are now in a position to write our paper and submit it for publication.

Prior to 1993, the way in which studies were written for publications varied tremendously. The Consolidated Standards of Reporting Trials (CONSORT) statement has meant that randomized controlled trials (RCTs) should now be reported in a specific way. A full description of the CONSORT statement is available online at http://www.consort-statement.org. This systematic way of presenting studies is aimed at therapeutic studies; however, it is equally valid for diagnostic papers and observational studies.

Putting it into the simplest terms, papers can be divided into the following sections:

- Title: Quick-to-read explanation of what is being studied
- Abstract: Short summary of the whole paper