

Research issues in Physical & Rehabilitation Medicine

Franco Franchignoni
Editor

Advances in Rehabilitation

AGGIORNAMENTI IN RIABILITAZIONE

Advances in Rehabilitation

AGGIORNAMENTI IN RIABILITAZIONE

Research issues in Physical & Rehabilitation Medicine

Franco Franchignoni

Editor

Maugeri Foundation Books
I libri della Fondazione Maugeri

Vol. 19, 2010
Pavia, Italy

ADVANCES IN REHABILITATION • AGGIORNAMENTI IN RIABILITAZIONE
Vol. 19, 2010

ISBN 978-88-7963-256-0

Series Editor: Marcello Imbriani, M.D.

PRINTED IN ITALY

© 2010

Tipografia PI-ME Editrice S.r.l.
27100 Pavia, Italy - Via Vigentina 136^A
Tel. 0382.572169 - Fax 0382.572102
E-mail: tipografia@pime-editrice.it
www.pime-editrice.it

All rights reserved.
No part of this book
may be reproduced
without the prior written
permission of the Publisher.

ADVANCES IN REHABILITATION
AGGIORNAMENTI IN RIABILITAZIONE
Vol. 19, 2010

**RESEARCH ISSUES IN PHYSICAL
& REHABILITATION MEDICINE**

Franco Franchignoni
Editor

CONTENTS

Introduction	v
<i>F Franchignoni</i>	
Medical ethics and experimentation in Rehabilitation Medicine	1
<i>R. Rubens, G.G. Vanderstraeten</i>	
Publishing scientific papers in Physical and Rehabilitation Medicine .	15
<i>G. Grimby</i>	
The peer review process	31
<i>F. Franchignoni, S. Negrini</i>	
Towards an epidemiology of functioning	53
<i>B. Fellinghauer, J. Reinhardt, G. Stucki</i>	
Functional assessment and outcome measurement: models and perspectives across the rehabilitation continuum	69
<i>C.V. Granger, M. Carlin, H. Feliciano, S. Markello, L. Tesio</i>	
The Rasch measurement model	89
<i>A.H. Elhan, A.A. Kucukdeveci, A. Tennant</i>	
Experimental design in Physical and Rehabilitation Medicine	103
<i>M. Zampolini, P. Sale, A. Giustini, M. Franceschini</i>	
Non-pharmacological placebo controlled randomized trials and the role of pragmatic trials in PRM	127
<i>S.J. Han, M. Imamura, F. Fregni</i>	
How to do and report systematic reviews and meta-analyses	137
<i>A. Liberati, M. Taricco</i>	

Qualitative research 165
T. Meyer, M. Schwarze, K. Bieleit, C. Gutenbrunner

**Internet resources for Physical and Rehabilitation Medicine:
databases, search engines, websites, e-books, images
and the tools to organize them 185**
E. Giglia, N. Barotsis, X. Michail

Clinical case reports in Physical and Rehabilitation Medicine 217
H. Collado, L. Bensoussan, G. Lotito, J.M. Viton, A. Delarque

How to prepare and present a lecture 223
N. Christodoulou, R. Frischknecht, K. Christodoulou

How to prepare a scientific presentation as a poster 241
A. Giordano, G. Ferriero, A. Nardone

INTRODUCTION

This book represents a dream come true, thanks to two major groups of support, for which I am deeply grateful. On one side, there is my Foundation (Salvatore Maugeri Foundation, IRCCS, Italy), with its President (Umberto Maugeri) and Scientific Director (Marcello Imbriani) who trusted me and believed in this editorial project right from the very first moment. On the other side, there is a group of highly-motivated friends and leading experts from many countries who, from the outset, accepted to be part of this educational project and help me with their wide expertise, dedicating considerable time to thoroughly prepare their contribution to this book.

There is a huge demand for high-quality free medical textbooks, and the specialty of Physical and Rehabilitation Medicine (PRM) is no exception. For similar reasons, a few years ago the WAPIE project (World Action Plan for Initial Education in PRM) was launched in Europe and then worldwide, one of its goals being to continuously inform our colleagues on advances, guidelines and evidence-based medicine related to PRM.

At the same time, a major source of satisfaction for the passionate authors of a book chapter is not so much the financial benefit as that of having the maximum conceivable audience of qualified colleagues. This is not easy when a book sells only a few hundred hard copies, but it is more likely to achieve if, associated with the print publication, free electronic copies of the book are distributed on selected websites. That is why we have decided to produce this book in both formats: the hard copy will be available for sale in bookshops, while the electronic one will be downloadable free-of-charge from the websites of the main European PRM bodies. This non-profit project was in fact carried out under the auspices of the following European PRM organizations, and I take this opportunity to thank them very much for their involvement:

- UEMS PRM Section (President Alain Delarque and incoming President Nicolas Christodoulou),
- UEMS PRM Board (President Franco Franchignoni and incoming President Jean Michel Viton),
- European Society of PRM (President Alessandro Giustini and incoming President Xanthi Michail),
- European Academy of Rehabilitation Medicine (President Gustaaf Lankhorst and incoming President Guy Vanderstraeten).

Next year, it is also planned to make available a free e-version of the book in the educational pages of the website of the International Society of PRM (ISPRM).

Why 'Research issues in Physical and Rehabilitation Medicine'? The advancement of PRM (and more generally of medical science) depends on the availability and application of new information gained from research. A successful research project depends not only on a well-designed agenda (that responds to clinical and societal needs) but also on having the research capacity essential to perform the study (in terms of a critical body of good investigators working as teams in supportive environments). This book is our small contribution to support basic research capacity in PRM, in order to build a strong future for our scientific field.

Overall, this project represents a generation change. It is our belief that the two formats - hard copy and electronic copy - of this book are complementary and will benefit from each other. Based on our previous experience with '*Assessment in Physical Medicine and Rehabilitation*' (Maugeri Foundation books), thousands of copies of which were freely distributed in e-version across five continents, we expect that the present book will be read all over the world by a very large number of PRM specialists and trainees.

The ideal book should be up-to-date, of high quality and low cost. I am sure that this book has these characteristics. The spirit of science is to share information, thoughts, wishes and visions, and we have tried our best to do this here in this project.

I thank again all who took part in the realization of this book. I wish them '*ad maiora!*', and all you who read it an enjoyable and fruitful reading.

Franco Franchignoni

Veruno, September 2010

ROBERT RUBENS¹

GUY G. VANDERSTRAETEN²

MEDICAL ETHICS AND EXPERIMENTATION IN REHABILITATION MEDICINE

¹ Ethics Committee, University Hospital
Gent, Belgium

² Department of Physical and
Rehabilitation Medicine,
University Hospital Gent, Belgium

I. MEDICAL ETHICS: HISTORICAL BACKGROUND

Throughout the history of medicine, both the physicians and the other actors in caring (nurses, allied health professionals, hospital administrators) have reflected on the ethical aspect of their actions. As mentioned first by Hippocrates and later confirmed by Galenus, acting without an ethical dimension is not feasible in the health care sector, where ethics is bound not only by the fundamental ethical principles of the society in which the care provider is engaged, but also by his professional actions. The term used to refer to the discipline concerned with these reflections differs depending on the line of approach.

Philosophers preferentially use the term 'bioethics', while physicians prefer 'medical ethics' and nurses 'nursing ethics'. These three labels encompass a comparable but not completely congruent field.

Usually, professional acting in the health care sector is ethically based on the four *prima facie* principles - autonomy, nonmaleficence, beneficence, and justice (1-4). These four principles can be applied by anyone whatever his or her religious, political, moral or philosophical beliefs (1).

The *prima facie* concept (Latin, meaning ‘at first sight’) implies that one should always test one’s actions against each of the principles concerned and approach each principle with respect. The *prima facie* approach refuses to make a choice between divergent principles and rejects the establishment of an absolute rank order; it refuses an absolute reference point. Adherence to this non-absolute principle makes it a tolerant approach (allows different decisions), but at the same time more difficult to apply because of its lack of total coherence. It is, however, a generally accepted ethical theory to analyse and evaluate issues arising in the medical field (5). A variety of other ethical theories are also used to this end, but according to most authors, they usually arrive at comparable principles through rational reasoning.

A. Autonomy

Origin and definition

The word ‘autonomy’ is derived from two Greek words: *αυτοζ* (*autos*) = self, and *νομος* (*nomos*) = law. The term comes from the Greek city-states and means the capacity of self-governance. Its meaning has been expanded and in medical ethics, it is implemented in the concept of the right to self-determination. More specifically, it is defined as an individual’s capacity to make autonomous decisions in so far as these are compatible with medical professionalism.

Historically, it can be traced to the works of Kant (6), but this has been challenged by Beauchamp (7). The traditional first implementation of the concept in politics was the Habeas Corpus Act (1679) and the Bill of Rights (1689) in England.

How is autonomy achieved?

Autonomy can only exist if an individual is free and capable of rational assessment. It is thus important that the following conditions be met: (1) freedom (=absence of coercion) and (2) capacity to implement the decision.

An autonomous decision can only be made if it is deliberate, informed and uncoerced.

The average patient usually does not fulfill all these requirements, but the care provider must at least aim at the maximum fulfillment and the highest possible autonomy for the individual in need of care.

In the ‘old’, traditional model of medicine, a paternalistic patient approach is traditionally used, in which the care provider is committed to serving the patient’s best interests, but the patient himself is denied any input into the decision-making process.

This model will be used for patients who have lost their autonomy and have no “proxy”. Nowadays, this is not uncommon in our urbanized society, in which family ties are often severed.

Consequences and derived principles

Out of respect for the patient's autonomy, the following complementary principles must be met: (a) always tell the truth; (b) respect the patient's private sphere; (c) safeguard medical confidentiality with all legal means possible (=respect the patient-physician privilege); (d) obtain the patient's consent prior to any intervention; and (e) try to provide the patient with as much information as possible to rationally support any decisions he may take.

Issues with truth telling arise when the prognosis is gloomy. However, even in such cases, the patient must be informed if he expresses the desire to be told the truth, but this information must not be forced upon him and his psychological structure must be respected. Under some circumstances, a person may not be able to cope with a poor prognosis, which can either be a limited life expectancy, a poor quality of life, or living with a permanent disability as is often the case in the field of physical and rehabilitation medicine. Although human mortality is generally accepted, some persons will always be inclined to suppress this fact. For truth telling, good communicative and empathic skills are of critical importance. Conveying the message to a previously healthy young man that he will have to spend the rest of his life as a paraplegic, requires compassion, understanding and an insight into communication principles.

Although it is the ground rule, absolute respect for a patient's privacy may give rise to problems, even in our pluralistic tolerant society. Good factual knowledge of his personal sphere is thus essential.

Medical confidentiality dates back to at least the traditional Hippocratic Oath. It implies that therapists must keep confidential any knowledge acquired during the course of providing services, in order to protect their patients, with some strictly defined exceptions to this rule. Confidentiality solely aims at protecting the patient, not the care provider!

Keeping information confidential not only applies to doctors, but to every caregiver engaged in health care. Obtaining information through a confidentiality breach will even cause lawsuits to be completely dismissed. Because of the obviousness of the impairment, rehabilitation professionals sometimes tend to attach much less importance to confidentiality, but even paraplegics have a right to privacy, e.g. with regard to bladder and bowel problems.

The informed-consent principle, i.e. the patient's consent prior to any intervention, forces us to discuss the therapeutic approach with the patient. In this respect, the care provider not only is the technician, but also the patient's counsellor (8).

Informed consent is absolute for every treatment or investigation in which experiments are carried out on patients. The term 'Nuremberg code' (later integrated in the Declaration of Helsinki) was based upon events that occurred during World War II, although the history of the USA has not been spared from serious violations of the right of human subjects either (the Tuskegee and plutonium experiments) (9, 10).

Obviously, a patient can only make a sensible and rational decision after having been properly informed (11).

B. Nonmaleficence (= do no harm)

Nonmaleficence (1-3) is the oldest concept in medical ethics and is embodied by the phrase ‘*primum non nocere*’. The origin of the phrase can be traced to centuries ago; some believe to Hippocrates, others to a much older tradition originating in ancient Egypt and Mesopotamia.

Hippocrates defined it as follows: “I will apply medical treatment for the benefit of the sick according to my ability and judgment; I will keep them from harm and injustice”. Nonmaleficence is usually placed on the same level as beneficence since both concepts are complementary.

Subconcepts

Although the concept of nonmaleficence is generally accepted, its interpretation by the different parties is sometimes widely divergent.

The concept incorporates three obligations: (a) do no harm; (b) avoid harm as much as possible; and (c) remove harm.

In addition, the rule of beneficence is applied: (d) do good or promote the wellbeing of patients.

In traditional ethics, the rule of nonmaleficence is considered to be more important and stronger than beneficence. This is not always true in the medical field (e.g., needle prick and administration of an antibiotic, incision and performing an operation).

Consequences and derived principles

From nonmaleficence the concept of harm has been derived. In a medical context, doing no harm encompasses three requirements: (a) do not cause death; (b) do not cause patients physical and emotional pain; and (c) refrain from doing anything that causes a patient to lose control against his will, do not deny him the capacity of deliberated self-rule (do not induce any form of addiction, do not administer psychoactive medications or use psychology to submit a patient to you (e.g., psychiatry in dictatorial regimes!). At the completion of rehabilitation, patient empowerment should be increased.

Unintentional harm can also be caused by professional negligence or “malpractice”, a term used in Anglo-Saxon litigation. This is usually due to lack of knowledge or training, and also includes acts for which the care provider is insufficiently qualified. Preventing negligence is an obvious duty of any care provider.

C. Beneficence (= do good)

The English term ‘beneficence’ (=do good) covers acts of compassion, kindness and charity (1-3). It can be broadened to include all interventions that promote the wellbeing of others. Beneficence is the core principle of medical ethics and is based on the rule that physicians, nurses and allied health professionals must aim at promoting the patient’s health. However,

beneficence should be defined very broadly because it not only implies curing the patient, but also caring for the patient and even inhibiting the disease course or supporting the incurably ill patient (palliative care).

Subconcepts

Beneficence can be subdivided into different components: (a) protect and defend the rights of others; (b) avoid causing others any harm; (c) remove the cause of harm; (d) help diseased or disabled persons; and (e) rescue persons in danger.

Consequences and derived principles

In the health care sector, the do-good principle is understood to cover the relationship to the patient, environment and society.

Although the foundations of the maxim have also been ascribed to Hippocrates, it will also find a ground for justice in compensating society for the costs of educating and training the health care provider.

However, the implementation of the do-good principle may sometimes conflict with the patient's desire for autonomy. This is frequently due to an inadequate understanding and insufficient communication skills of the care provider. Under some circumstances, the do-good principle is likely to clash with the autonomy principle, for which no cut-and-dried solution exists. One should always attempt to find the least disturbing solution. In this situation, not taking a decision is unacceptable because it will interfere with nonmaleficence (e.g., non-transfusion policy for Jehovah's Witnesses, distinction between an adult competent patient and the child of a Jehovah's Witness patient).

D. Justice

Although justice is said to be a fairly recent concept in medical ethics, it had already been mentioned in the WHO oath, in which the physician pledges to treat friend and enemy equally. Justice is therefore frequently also called fairness.

Subconcepts

Justice can be divided into three categories: (a) fair distribution of limited health resources, rational use of resources; (b) respect for the patient's specific rights; and (c) respect for the laws or rules of society. Technically, it is not easy to decide who will get what treatment. Not everyone can get every treatment under every circumstances. Besides the financial restraints imposed by society to the health care sector in Europe, the possibilities are also limited by the shortage of donors.

Consequences and derived principles

Following Raanan Gillon (3), we feel that a distinction should be made between the levels on which decisions are taken, while differenti-

ating between decisions taken by an individual and those taken by an organization, professional group or society. A care provider has to observe four roles. Firstly the decision must be morally justified and must not be based on unacceptable premises from general ethics. It is not the care provider's task to punish a patient for a maladjusted life or lifestyle! Secondly, the limited resources available must not be wasted. If a cheaper treatment is likely to produce the same effect as a more expensive one, the cheaper one will automatically be chosen, provided that the burden imposed on the patient is the same. Thirdly, the patient's rights must be respected, and fourthly, the rules of society must be followed, unless they conflict with general ethics (e.g., Nazi program for the mentally ill). Concerning organizational, professional and societal decision-making, any improvements and modifications can only result from interactions between these three structures. These rules should then be followed by all professionals.

II. RESEARCH AND "THERAPEUTIC MISCONCEPTION"

A. Research and experimentation in medicine

Since the development of scientific medicine in the 19th century, the consensus has been that experimentation is essential to expand medical knowledge. Initially, animal models were preferentially used, but it was soon realized that these failed to produce the necessary clinical information and human experimentation has to be introduced. Through advances in epidemiological thinking in the 20th century, Gaverett's 'law of large numbers' (12), from which evidence-based medicine (13) has developed, has become extremely important in medicine also. Experimentation is the basis of scientific medical thinking. However, experimental clinical trials use human subjects as a means to an end rather than to focus on their individual therapeutic needs. It is thus not surprising that experimentation, important as it is, has sometimes led to serious ethical issues.

Historically, conducting unethical experiments on human subjects has always been ascribed to Nazi Germany (14), although such experiments have also been done in the USA (15) and England (16) without observing the rules of informed consent (e.g. the Tuskegee experiments) (17).

Nevertheless, experimentation is a necessity if progress is to be maintained. This also applies to rehabilitation medicine. In addition to general ethical issues, these patients also pose specific challenges. In the classical medical model, the physician compares treatment A with treatment B for curing a patient with a specific disease. If a randomized double-blind study shows treatment B to cure 40% more patients without causing adverse reactions, then treatment B is accepted to be markedly superior and becomes the new standard treatment.

Although complete cures (=full 'restitutio ad integrum') may also be achieved in rehabilitation medicine, more realistic goals are stabilization, slowing down of progression, or improvement of the quality of life despite disease progression. Therapies to reach these goals must also be tested, but the end point is much less clear. Chronically ill or disabled patients obviously expect to benefit from any intervention that is proposed to them. For the sometimes very vulnerable but still hopeful population, every gleam of hope holds the promise of a new future. The experimenter should proceed with great caution and reserve because these patients readily feel positive about each new approach, precisely because of their vulnerability. The same caution was also expressed by Blackmer (18) several years ago.

Because of the need for experimentation and the associated issues, strict international regulations were developed and subsequently translated into local laws. The ethical principles were concretized in the World Medical Association Declaration of Helsinki of 2008 (19), and were at European level implemented in the Directive 2001/20/EC (20). In addition to general rules, each of these policies contain the requirement of having each trial proposal reviewed and approved by an ethics committee.

B. Ethics committees and experimental trials

Ethics committees must assess the trial proposal in the country where the trial is to be conducted. In the USA, regulations and procedures of the "Institutional review boards" (IRB), as are ethics committees called there, are centrally regulated by the Food and Drug Administration (FDA). Although steps are being taken to take similar actions in Europe, it is still the prerogative of each individual EU Member State to establish the rules of conduct of experimental trials. However, ethics committees must take into account the universal principles dictated by the Declaration of Helsinki, as well as local factors resulting from conditions present in the country where the trial is to be conducted. Ethics committees must have the necessary expertise to review the trial proposal and must be completely independent of the investigator and his sponsor. The composition and basic principles of ethics committees are different in each Member State of the European Union.

C. Types of experimental trials and rules for assessment

Experimental trials can have a therapeutic or non-therapeutic objective. A therapeutic experimental trial encompasses four components: (1) improvement of care (e.g. new nursing technique); (2) investigational drug; (3) novel surgical intervention; and (4) new material or device. Every change in treatment is best preceded by an assessment under experimental and controlled conditions.

A non-therapeutic experimental trial may be an additional measurement, registration, imaging technique, or interview.

Depending on the type of trial, the ethical assessment has to follow different detailed criteria, again guided by the principles of medical ethics.

To respect the subject's autonomy and value, an experimental trial can never be conducted without his explicit written informed consent, given without any pressure from the investigator. The therapeutic experimental trial must always respect the 'equipoise' principle.

D. Equipoise

The most suitable scientific method to gather scientific data is a randomized controlled trial (RCT). For scientific-philosophical reasons, double blinding is preferred (i.e. neither the physician nor the patient is aware of the group into which the patient has been randomized), the precept being that each treatment arm or alternative must be equal or in equipoise.

In the last 30 years, the angle of equipoise has shifted. Originally, equipoise was a concept in an individual's treatment. According to Fried (21), it refers to a situation in which the treating physician is in this particular patient unable to choose between two treatment alternatives. Both treatments are thus legitimate. In an RCT, the choice that otherwise would be made by the physician and/or patient on the basis of subjective factors, is entirely random. This type of equipoise is termed individual, theoretical, or Fried's equipoise.

In a scientific context, the term 'clinical equipoise' is more frequently used. It refers to scientific uncertainty as to which of two treatment alternatives will be more beneficial for a particular group of patients. This is the concept as first introduced by Freedman (22) in 1987, who defined clinical equipoise as a professional consensus on the choice between two treatments. Placebo can only be used in one treatment arm provided that no proven effective treatment option exists for this particular group of patients. An exact definition of the group and a fortiori of the disease itself is of paramount importance.

E. The dilemma of the physician-investigator

Every 'good' physician is obligated to offer the best treatment, tailored to the sick or disabled person's individual needs and based on his scientific knowledge and professional competence. This duty constitutes the basis of the physician-patient fiduciary relationship and can be interpreted as an implicit 'provision of care contract'. This Hippocratic tradition has characterised the medical profession since more than 2000 years. Under some circumstances, limits can be imposed by society but the patient should know that the physician will do everything within his power to honour his contract with the patient.

The practice of medicine is designed to improve the patient's condition, alleviate his suffering, or rehabilitate the patient. It is precisely because of this obligation that society has made provisions to grant physicians "medical exception" (23). The purpose of medical treatment is to give the individual patient the best possible care and therapies. In addition, the physician can conduct and record observations, but their value as evidence is scientifically weak.

Truly scientific work will be done if the goal is to acquire knowledge and human subjects are used to this end. The primary goal of the physician will no longer be to maximally benefit the individual patient, but to gather data for the purpose of improving the outcomes for future patient populations with the same condition.

In this model, the practice of medicine exceeds the provision of care, because the treatment itself is only an adjunct or a means to contribute to scientific knowledge. Medical care is thus designed to generate new knowledge for the benefit of future patients through using individual patients.

Conducting clinical research is markedly different from providing standard medical care. Patients enrolled in a clinical trial are examined at standardized intervals; their treatment and follow-up are not tailored to their individual needs but are predetermined by the designer of the trial. It is also not uncommon for the patient to be subjected to more investigations than strictly necessary for his individual medical care. The primary focus is not on the patient's best interest but rather on the possible intermediate endpoints of the trial.

A placebo is often used for comparison purposes. However, there are only few conditions for which no effective treatment exists. The results of the current standard treatment may sometimes not be good, but placebo means no treatment at all.

The information provided to the patient is highly standardized and adjunctive therapy or fertility choices are often limited.

It is not uncommon for the physician-investigator to exert slight moral pressure on the patient for obtaining his consent to participate in a trial (the physician has a hidden agenda).

The issues about ethical constraints have become so strong that Brody and Miller (24) devised a dual model to illustrate the dilemma (fig. 1).

Both positions are justifiable. In the 'similarity position', the ethics of research is assumed to be part of the therapeutic practice, e.g. the standard oncological protocols. A similar structure can also be accepted in the rehabilitation setting, e.g. an additional intervention in a standard rehabilitation protocol. The standard protocols for investigating new drugs using a 'pure' protocol (often with a placebo arm) will rather adhere to the 'difference position'. In these circumstances, the standard medical care and the experimental protocol are markedly different. The care delivered will often be programmed and standardized, which is fairly different from the normal personalized care. Conversely, it is also

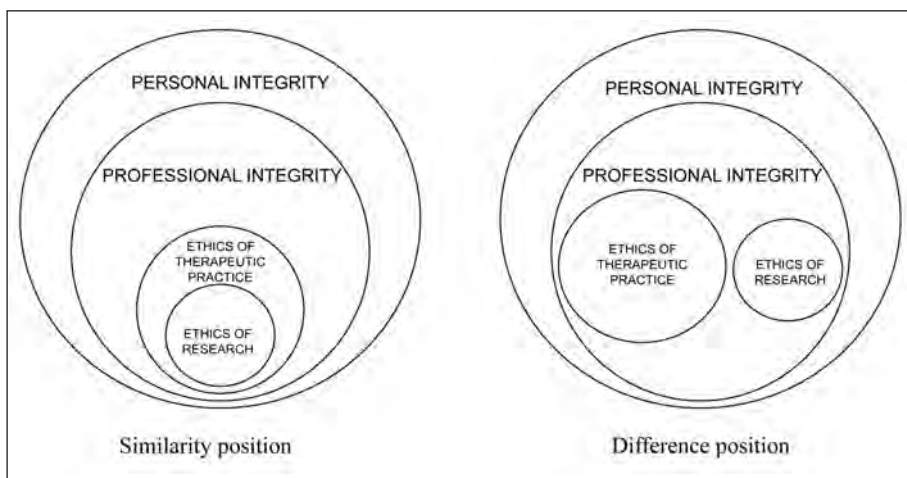


FIGURE 1. Issues about ethical constraints: the dual model [modified from (24)].

true that the incorporated measures with strict and controlled policies frequently result in more effective and qualitatively better care.

Furthermore, a patient participating in an experimental protocol often also benefits from his participation: closer follow-up, much more technical investigations, more accurate and strictly defined diagnosis at entry, alertness of the treating physician to any adverse effects of the treatment. With each new trial, the 'medically' acceptable supplementary burden must be considered and the patient must be accurately informed of the complementary investigations, risks and pain he will experience from his participation.

An experimental clinical trial of a new treatment for vertebral fractures will always include repeated blood sampling and more frequent MRIs (not only the procedure itself but also transportation which entails pain and risks for these patients), and sometimes also complementary investigations such as lumbar puncture, EMG, etc...

It is the physician-investigator's duty to properly explain and confirm in writing all these factors associated with trial participation. Particularly in academic medical centres, the treating physician is frequently also the physician-investigator. The fact that an academic career is much more based on research than on excellent patient care may sometimes interfere with the physician's Hippocratic responsibility. To prevent this conflict of interests, the American Medical Association stipulates that someone other than the treating physician should obtain the prospective participant's informed consent to participate in a trial (25). Although in Europe the rules are less stringent, this dilemma is also a matter of concern (26).

Conversely, patients often feel privileged to participate in a trial, because it allows them to receive cutting-edge treatment for their condition. Sometimes, patients may even ask to be enrolled in a trial (27), e.g. when a definite cure for their disease does not yet exist (AIDS, cancer), or when

therapeutic knowledge is so limited and their pain so severe that any intervention can be considered helpful (progressive degenerative nerve diseases). In this context, the physician is faced with an additional dilemma, i.e. that of “therapeutic misconception”.

F. Therapeutic misconception

The term ‘therapeutic misconception’ was introduced by Appelbaum et al. (28, 29) some 30 years ago. Although their psychiatric patients appeared to have a good understanding of what informed consent involved, the authors found that in a randomized study the majority of these patients explicitly believed that they had been assigned to the medication group despite the detailed explanation by the physician-investigator about randomization to either placebo or active medication. In subsequent years, the concept of therapeutic misconception was broadened in that participants in a randomized trial are always convinced that the treatment they receive surpasses standard medical care, although some of them actually receive placebo. They thus fail to recognize that the goal has shifted from serving the individual patient’s best interests (Hippocrates) to producing generalizable knowledge about how people with the same condition can best be treated (scientific). This group focus is frequently present, also in the rehabilitation medicine setting, which requires a slow and sustained effort from the patient. A sick or disabled person also firmly believes in every novel approach designed to improve his condition.

This has led to a strong belief in the value of experimental therapies and not uncommonly also to selective blindness to the risks associated with trial participation. Many participants fail to consider the placebo arm of a trial. Therefore, limited rational informed consent is actually given instead of truly informed consent.

Therapeutic misconception thus defines the tension between medical care and scientific research. Objective scientific methodology is important in research but conflicts with patient-centred orientation of medical care which has the patient’s best interests in mind.

G. Competing interests: yes, but should they be divided?

The problem described above is real but inherent in the academic process. Unless health care students are properly educated, care will no longer be possible in the future. The academic educator is required to engage in research. The earlier period (scholasticism), in which the search for new lines of empiric facts was avoided, led to very inferior standards of medical care. If during his education a student would not be confronted with research, he would inevitably develop a rigid non-intellectual vision of scientific thinking. Tomorrow’s knowledge actually stems from experimental findings. Research will eventually result in better care. If

Fleming had not tested the antibacterial effect of penicillin, the mortality of infectious diseases would still be extremely high. Should a distinction be made between the ethics of care and the ethics of research? Two schools have been described in the literature. The NIH group (USA model) under the guidance of Miller (30) pleads for a strict distinction between the two areas. Patients do not come to the NIH for standard medical care, but to participate in well-designed trials in which novel treatments (medication, medical device, or surgery) or new investigational techniques are tested, independent of their treating physician. Miller claims that a strict distinction between treating physician and research centre places the patient in a better position to give written informed consent because he knows exactly what to expect.

At the other side of the divide are the traditional clinicians (31) who regard research as an integral part of clinical care. This is where the concept of equipoise enters into play. Refusal of a clinician to engage in research would be a major expression of arrogance and contravene every Greek philosophical principle, as it would assume current medical knowledge to be perfect. In their eyes, every good physician will sooner or later be an investigator.

However, there are some boundaries that cannot be crossed. In addition to complying with clinical equipoise, the investigator should also remain completely independent of the sponsors. The investigator should obviously be compensated for his work, but compensation or moral pressure to publish must never compromise objectivity. In this respect, ethics committees or institutional review boards have an important role to play.

H. Is the problem real or merely a dispute among lawyers, philosophers and medical ethicists?

Incidents are fortunately rare, but they continue to occur even in our strictly regulated society, despite the strict control and safety rules. The two most striking cases are the TGN1412 experiment in London and the Propatria trial in Utrecht (33). Serious and even fatal adverse reactions can also occur in standard conditions of care. The case of Jolee Mohr, which received ample consideration in the lay press, must never be forgotten. This 36-year-old lady presented to her rheumatologist in February 2007 and was diagnosed with rheumatoid arthritis. On February 12, she signed an informed consent form for enrollment in a phase I/II trial of an adenovirus-associated gene therapy stimulating the production of anti-TNF. She received a first and second injection on February 26 and July 3, 2007, respectively. On the evening of the second injection, she developed fever and vomiting. She was admitted on July 12, developed multiple organ failure on July 18, and died on July 24, 2007 (34). A subsequent detailed assay disclosed numerous concomitant factors and definitely ruled out viral vector injection as the unique cause of death. This was very important to the resumption of the trial (35), but did not exclude that in this

particular patient the addition of the vector to other therapeutic agents had a fatal outcome (36).

III. CONCLUSION

Particularly in rehabilitation medicine, there is a dire need of experimental protocols. Too much is still based on old traditions and a sound empirical basis is lacking. The rehabilitation clinician is thus morally obligated to increase this knowledge. Rehabilitation patients on the other hand are strongly dependent on their care providers. Cognitive function may be impaired due to various disorders and injuries. It is essential that these patients be given a detailed explanation of what informed consent involves, and be properly counselled. The pitfall of therapeutic misconception must always be kept in mind, particularly in this patient population. Finally, the process of peer review by the ethics committee or institutional review boards is of paramount importance to assist the investigator. The trial design can often be much improved through thorough discussion and consultation with the investigator.

REFERENCES

1. Dawson A, Garrard E. In defence of moral imperialism: four equal and universal prima facie principles. *J Med Ethics* 2006; 32: 200-204.
2. Raanan G. Medical ethics: four principles plus attention to scope. *Br Med J* 1994; 309: 184-188.
3. Beauchamp T, Childress J. *Principles of Biomedical Ethics*. Oxford Univ. Press 1994, 562 pp.
4. Seedhouse D, Lovett D. *Practical Medical Ethics*. John Wiley & Sons 1992.
5. Gillon R.A, Lloyd A. *Principles of Health Care Ethics*. John Wiley & Sons 1994, 1118 pp.
6. Baker R, Mac Cullough L. Medical ethics' appropriation of moral philosophy: the case of the sympathetic and the unsympathetic physician. *Kennedy Institute Ethics J* 2007; 17: 3-22.
7. Beauchamp T. History and Theory in "Applied Ethics". *Kennedy Institute Ethics J* 2007; 17: 55-64.
8. Rubens R. Het "informed consent" in de kliniek. Oorsprong, analyse, verklaring, problematiek en toekomst van een begrip in de medische ethiek. *Tijdschr. Voor Geneesk* 1998; 54: 1068-1078.
9. Kaufman S. The World War II plutonium experiments: contested stories and their lessons for medical research and informed consent. *Culture, Medicine, Psychiatry* 1997; 21: 161-197.
10. Dolgin J. The legal development of the informed consent doctrine: Past and present. *Cambridge Quarterly of Health Care Ethics*, 2010; 19: 97-109.
11. Callahan CD, Hagglund KJ. Comparing neuropsychological and psychiatric evaluation of competency in rehabilitation: a case example. *Arch Phys Med Rehabil* 1995; 76: 909-912.
12. Gavaret J. *Principes Généraux de Statistique Médicale*. Paris, 1840.
13. Sackett DL, Richardson WS, Rosenberg W, Haynes RB. *Evidence based medicine: How to practice and teach EBM*. New York, Churchill Livingstone, 1997.
14. Schmidt U. The Nuremberg Doctor's Trial and the Nuremberg Code. In: *History and Theory of human experimentation*. Ulf Schmidt, Andreas Frewer (Eds). Franz Steiner Verlag, Stuttgart 2007, pp. 71-116.
15. Beecher HK. Ethics and Clinical Research. *New Engl J Med* 1966; 274: 1354-1460.
16. Papworth MH. *Human Guinea Pigs*. London, Routledge and Keegan Paul, 1967.

17. Jones JH. *Bad Blood: The Tuskegee Syphilis Experiment*. New York: Free Press, 1981.
18. Blackmer J. The unique ethical challenges of conducting research in the rehabilitation medicine population. *BMC Med Ethics* 2003; 4: 2-6.
19. Anonymous. The Declaration of Helsinki of the World Medical Association, 2008. *World Med J* 2008; 54: 122-125.
20. Fontaine N, Rosengren B. Directive 2001/20EC of the European Parliament and of the Council of 4 April 2001. *Official Journal of the European Communities*, EN, 2001, L121/34-L121/44.
21. Fried C. *Medical experimentation, personal integrity and social policy*. New York, NY: American Elsevier Publishing Co Inc, 1974.
22. Freedman B. Equipoise and the ethics of clinical research. *New Engl J Med* 1987; 317: 141-145.
23. Moratti S. The development of "medical futility": towards a procedural approach based on the role of the medical profession. *J Med Ethics* 2009; 35: 369-372.
24. Brody H, Miller FG. The clinical-investigator: unavoidable but manageable tension. *Kennedy Institute Ethics J* 2003; 13: 329-346.
25. Morin K, Rakatansky H, Riddick FA, Morse LJ, O'Bannon JM, Goldrich MS, Ray P, Weiss M, Sade RM, Spillmann MA. Managing conflicts of interest in clinical trials. *JAMA* 2002; 287: 78-84.
26. Millner F. Recruiting research participants. In: *The Oxford textbook of clinical research ethics*. Oxford University Press, 2008, pp. 396-403.
27. Annas GJ. Cancer and the constitution - Choice at Life's End. *New Engl J Med* 2007; 357: 408-413.
28. Appelbaum PS, Roth LH, Lidz CW. The therapeutic misconception: informed consent in psychiatric research. *Int J Law Psych* 1982; 5: 319-329.
29. Appelbaum PS, Roth LH, Lidz CW, Benson P, Winslade W. False hopes and best data: consent to research and the therapeutic misconception. *Hastings Center Report* 1987; 17: 20-24.
30. Miller FG, Rosenstein DL. The therapeutic orientation to clinical trials. *New Engl J Med* 2003; 348: 1383-1386.
31. Grunberg SM, Cefalu WT. The integral role of clinical care. *New Engl J Med* 2003, 348: 1386-1388.
32. Kenter MJH, Cohen AF. Establishing risk of human experimentation with drugs: lessons from TG1412. *Lancet* 2006; 386: 1387-1391.
33. For the Dutch Acute Pancreatitis Study Group; Besselink MG, van Santvoort HC, Buskens E, Boermeester MA, van Goor H, Timmerman HM, Nieuwenhuijs VB, Bollen TL, van Ramshorst B, Witterman BJ, Rosman C, Ploeg RJ, Brink MA, Schaapherder AF, Dejong CH, Wahab PJ, van Laarhoven CJ, van der Harst E, van Eijck CH, Cuesta MA, Akkermans LM, Gooszen HG. Probiotics prophylaxis in predicted severe acute pancreatitis: a randomized double-blind, placebo controlled trial. *Lancet* 2008; 371: 651-659.
34. Weiss R. Death points to risks in research. *Washington Post*, 2007, 6 august.
35. Hampton T. Arthritis gene therapy trial resumes. *JAMA* 2008; 299: 28.
36. Wadman M. Gene therapy might not have caused patient's death. *Nature* 2007; 449: 270.

Further reading:

Medical Ethics Series Supplement. *Am J Phys Med Rehabil* 1995; 74(1 Suppl): S3-S58.

PUBLISHING SCIENTIFIC PAPERS IN PHYSICAL AND REHABILITATION MEDICINE

Department of Clinical Neuroscience
and Rehabilitation, Sahlgrenska
Academy at University of
Gothenburg, Gothenburg, Sweden

This chapter reviews the process of writing and submitting an article, explains how papers are dealt with by editors and editorial offices, summarizes the peer review process, and provides advice on revising and re-submitting a manuscript. This information is intended to be of value for research and clinical trainees as well as for established scientists. There are several handbooks available on this topic, published by the British Medical Journal (BMJ) Publishing Group (1, 2), and detailed advice can be found in the publication *Uniform requirements for manuscripts submitted to biomedical journals* (3), published by the International Committee of Medical Journal Editors (ICMJE), and which is widely accepted as a standard set of guidelines.

SCIENTIFIC JOURNAL PUBLISHING

Original scientific publication began in the 17th century and was, for a long time, handled by the universities on a non-profit basis. During the 20th century a large number of commercial publishers were established. However, a number of the journals published remained in the ownership of scientific or professional organizations, foundations or universities. Since then, a

number of journals published by such research organizations themselves on a non-profit basis have continued to be produced or have been developed further.

Before selecting a journal to submit your manuscript to it is advisable to find out about the type of publisher involved and their policy on open access (as exemplified below).

PLANNING A SCIENTIFIC ARTICLE

Before writing a scientific article there are certain steps that you are advised to take:

- Define your subject matter and make sure that you have a “story” to tell.
- Put yourself in the reader’s position and ask what new and interesting information you have to relate and what is its relevance for physical and rehabilitation medicine.
- Define the aim(s) (e.g. main aim and secondary aims) of the study so that you can express them clearly.
- Review published articles in the same area, to ensure that you have some new or relevant confirmatory information, and to gather information for the literature survey you may want to include in the introduction.

CHOOSING A JOURNAL

In choosing the initial journal to which to submit your article, read the information on the scope and the instructions to the authors for that journal and some other relevant journals. The following aspects ought to be considered:

- Does the journal have an appropriate profile for your study and does your paper fit into the main area for that journal?
- Is the journal known and read within your research area?
- Do you or your colleges have previous experience of the journal, both with respect to the type of papers published and the handling procedure and time it usually takes to reply about submitted manuscripts?
- Take a look at the Editorial Board; are there members representing your specific area of research?
- What are the impact factor and other relevant measures of quality of that journal?
- What is the journal’s policy on open access?

You can obtain some information about journals in the field of human functioning and rehabilitation research from a paper published in 2007 (4). An extensive list of journals relevant to the field of physical and rehabilitation medicine is included in that article, although some relevant journals may be missing as new journals have been started since then and

only those published in English are listed. Publishers' names, addresses, specific areas of research and impact factors at the time of publishing are presented in that report.

In another paper published at the same time (5), five distinct fields for research and publishing in rehabilitation were proposed:

- Biosciences in rehabilitation as a basic scientific field;
- Human functioning sciences, as a basic and an applied field;
- Biomedical sciences and engineering as an applied field;
- Integrative rehabilitation sciences as an applied field;
- Professional rehabilitation sciences as an applied field.

Identifying the subject of your study within one of these areas may help you to choose where to publish it. This classification may also be used to characterize articles and abstracts in scientific journals and at congresses.

Impact factor

Impact factors are used for distributing resources to universities and departments, for helping make decisions about research grants, for appointment of positions, as professors and other teachers or researchers, but also for authors to decide to which journal to submit a paper.

The impact factor is a measure of the level of citations from all indexed journals to the articles in a specific journal. It is calculated by taking the number of citations made during a specific year (e.g. 2009) in all indexed journals to articles published by the specific journal during the previous two years (e.g. 2007 and 2008), and dividing this figure by the total number of articles published by the specific journal during that two-year period. For example, if a total of 100 articles were published in *Journal A* during 2007 and 2008 and there were 190 citations of these articles in all indexed journals during 2009, the impact factor of *Journal A* for 2009 would be 1.9.

One of the limitations of this traditional impact factor is that the window for assessing the number of citations is rather small and occurs during an early stage after publication. It has been shown that the peak of citations of articles in clinical research occurs approximately three years after publication, with a relatively high level of citations continuing for up to six years after publication, but this is not taken into account in calculating the traditional impact factor. There may be several reasons for the delay in citation, such as the time it takes to perform and publish clinical and epidemiological studies, but also that clinical research information may be considered new for a longer time than for more experimental research. Key references may thus be older in clinical articles. In recent years a five-year impact factor has also been used as a complement to the two-year impact factor, and this may be more relevant for clinical publications.

It should, however, be noted that there are other bibliographic means of evaluating the quality and reputation of a journal; for example, cited half-life and different means of recording downloads of an article. It is

also important to remember that although the impact factor may provide relevant information on the citation rate and quality of a journal it cannot be used in evaluating a specific article; examining the citation frequency of that particular article should do this.

Open access

A journal's policy on open access is of interest when choosing where to submit an article. In recent years this topic has been debated at professional meetings and in the scientific literature, but also in the public media. The idea is to enable full and open accessibility of scientific information, and to uphold the principle that this information belongs to the scientific community without commercial profit. In the fully developed form of open access anyone should be able to read the full text of a paper online as soon as it is published. There are now a number of open access journals published only online, with or without fees for submitting articles. It is worthwhile considering such journals if they are appropriate for your manuscript. In addition, such journals are peer reviewed, and the publication time may be reduced. In a hybrid form ("partial open access") the authors will have the right to put their manuscript on their homepage or in the repository of the department or university at the time of submission, followed by the article in its published form. At present, the most prestigious research councils, such as the National Institute of Health in the USA and the Wellcome Trust in the UK, and grants from the European Union (EU) require that there is at most a six-month embargo prior to an article being made fully available for open access. The same policy is also adopted by some national research councils. It is therefore recommended that authors give priority to journals with such a policy on open access. This type of policy may also benefit the citation rate.

Further movement towards open access will require a change in the way journals are funded, from subscription fees forming the basis for a journal's economy to page charges being paid by the authors, unless journals are in receipt of considerable income from advertising or professional organizations. It would be assumed that page charges will be paid to the authors through research grants and universities (e.g. instead of subscriptions for the paper versions of journal). The move towards full open access will take some time, but authors of scientific papers are recommended to be aware of this ongoing process.

PREPARING A MANUSCRIPT

Before starting to prepare a manuscript you should read the instructions to the authors, which will be found in the journal and/or on its homepage. Each journal has specific instructions, although there may be some common information. General guidelines can be found in the *Uniform requirements for manuscripts submitted to biomedical journals* (3). The maximum length of manuscript allowed is usually stated, and this

should be adhered to. Some journals may allow some extra length for an additional cost.

Title

The article title should be concise and informative, and should be written to attract the reader's interest. It may be possible to indicate the main finding or the main hypothesis in the title, but do not try to be sensationalist. You should also provide a short title, to appear as a running head at the top of the printed pages of the article, which in some journals may also be listed on the front page.

Abstract and key words

Follow the recommended structure for the abstract. It may be unstructured, that is without any headings, or structured with recommended headings. Often no more than 200 words are accepted. The required number of key words should be included, all or most of which should be from the Medical Subjects Headings (the National Library of Medicine's vocabulary for indexing; 6). Avoid words included in the title, as they will also be indexed.

Introduction

The Introduction should set out the background to the study, including its purpose, justifying that the study ought to be performed and published so as to present new results or to confirm results from previous studies, as appropriate. Including a review of the relevant literature is appropriate, but the length of such a review may vary, and the information must not be repeated unduly in the Discussion. It is important to decide what belongs in the Introduction and what in the Discussion. Give only relevant references for your study. Do not review the subject extensively. At the end of the Introduction present the aim of the study, preferably in point form.

Materials and Methods

In the Materials and Methods section, which in some journals is combined into one heading, provide detailed information on the subjects/patients/persons/individuals (different expressions may be used), who participated in the study, including the number of non-participants and/or drop-outs. Detailed information may be given as a table. Present characteristics of the subjects/patients; if appropriate use the relevant terms from the International Classification of Functioning, Disability and Health (ICF) (7). Do not use the term demographics, but rather characteristics of the subjects/patients.

The methods should be described so that the study could be reproduced, using either detailed descriptions in the text and/or relevant references. In any case, give the readers a general idea of the methodology, and how well the methods used have been studied with respect to reliability, validity and other relevant characteristics.

Manufacturers and/or suppliers of equipment and materials used in the study should be identified by name, town, country, etc. Where and how this is presented may vary between journals.

Describe blinding of the observers and, if pertinent, of the subjects. In reporting the structure of a randomized controlled study (RCT) follow the recommendations of the Consolidated Standards of Reporting Trials (CONSORT) Group (8), which presents a standard way to report trials. The checklist includes items, based on evidence, that need to be addressed in the report; the flow diagram provides readers with a clear picture of the progress of all participants in the trial, from the time they are randomized until the end of their involvement. There are specific recommendations for assessing non-pharmacological treatments, which are relevant for rehabilitation research (9). Describe clearly how randomization was performed, as this is a critical issue.

You are recommended to register clinical trials with a trial register (e.g. www.clinicaltrials.gov); indeed for some journals this is a requirement for publication of the results. When reporting studies in human and animal subjects, indicate approval by the relevant ethics committee. Do not use patients' names, initials or other indicators that may reveal subjects' identities. Identifying information should not be published in patient history, drawings, or photographs, if not essential for scientific purposes, and unless the patient (or parents or guardians) have given written informed consent for publication and this is stated in the manuscript before publication.

Statistical analysis

Check the instructions in the journal for the use and description of statistics. Ensure that you have used correct and optimal statistical methods. Often it is recommended to consult a professional statistician before starting the study and on an ongoing basis. Such persons may, with their agreement, be acknowledged in the article if full authorship is not justified. Whenever possible, quantify findings and present them with appropriate indicators of measurement errors or uncertainty, such as confidence intervals. Calculations of power should be performed whenever appropriate. Be aware of the risk of statistically significant results due to chance at multiple comparisons, such as multiple *t*-tests, and correct for that. Use the appropriate number of decimals. Indicate which computer programs were used. There may be specific advice on, for example, the use and treatment of data from ordinal scales (e.g. 10). Median level and quartiles are appropriate measures for describing the distribution of ordinal data. In principle, data from ordinal scales should not be treated and reported as interval data using traditional statistics unless specific reasons are expressed, but non-parametric statistics should be used. In some cases, treatment of data with techniques that aim to transform ordinal data to interval data using item response theory (IRT) (as in Rasch analysis) or other specific techniques can be used.

For qualitative studies, present data collection strategies, such as selection of subjects, methods of collection, type of data, and relationship of the

researcher to subjects/settings. Also describe the data analytical process, defining clearly the concepts and categories/themes, how these were developed, and how they relate to the data. Report the quality and validity of the data; for example, triangulation, respondent validation, application of critical thinking to analysis, influence of the researchers on data collection and analysis, and the critical approach to the status of the data collected.

Results

The results should be presented in a logical sequence. View the way to present data from the perspective of readers. Avoid repeating information in the text as well as in tables and figures. Do not write in the text what variables are seen in a table or figure, but which are the main findings in tables or figures. Thus, write “In Table xx is seen that” or refer to the specific table or figure in parentheses after the main information seen in the table or figure is presented. For individual data, figures may be preferred unless there are few persons in the study. However, if you want to allow for further comparison, e.g. meta-analysis, relevant numbers must be given in the text or tables.

Discussion and Conclusion

The Discussion should start by summarizing the main findings briefly, not repeating in detail the information from the Results section. Put the findings in the context of the current literature, with an historical retrospective. When indicated, discuss the clinical and practical relevance and applicability of the results, and the limitations of the present study and results. You may also indicate areas for future studies. Be careful not to repeat information already given in the Introduction, unless it is presented in another context. Some journals will have a specific heading for a Conclusion or Summary after the Discussion, other journals end the Discussion with a paragraph starting without a separate heading and with “In summary...” or “In conclusion...”.

Be aware that the most-, and probably first-read, sections of a scientific paper are the Abstract, followed by the beginning and end of the Introduction section, the Discussion and perhaps the figures.

Acknowledgements and Conflicts of interest

Acknowledgements may be given to persons who have contributed to the study or article in different ways, but who are not eligible to be included as authors. For multicentre studies, there may be a large number of contributors or collaborators to mention under Acknowledgements or in a specific section on the title page. Some journals require that the persons mentioned in the Acknowledgements should give their permission, as otherwise persons who do not accept that they have supported the manuscript may be presented as having done so.

Acknowledgements should also be given regarding any financial support received from scientific research councils, grant committees or commercial agencies. Any conflicts of interest in this context should be re-

ported when submitting the manuscript. If financial support has been received from a commercial agency, it is important to state that this has had no influence on the presentation and discussion of the results, and that the researchers have had full scientific freedom even if some practical support was given in conducting the study.

Conflicts of interest may be factors such as employment by a commercial firm with any connection to the study, having acted as a consultant or a lecturer paid by a commercial firm in any way involved in the study, receipt of financial or practical support from a commercial firm, or of drugs or equipment for the study, etc. Whether this information is presented in detail in the article or recorded at the Editorial office depend on the editorial principles of the journal, but full openness is recommended as readers will then be able to evaluate whether any conflict of interest may have had any influence on the study and on the published paper.

References

The reference list should include relevant references, both for understanding and justifying the study, for documenting the methods used, and for analysing and integrating the results in the context of current knowledge in the field. In journals published in English it may be permissible to list relevant references published in other languages with an accompanying translation of the titles into English. Follow the Instructions to authors as to how the references should be listed, as there are different systems. One of the most common is the Vancouver system, in which references are numbered sequentially throughout the text. There are usually a maximum number of references that can be accepted.

Tables and figures

Tables should be presented on separate pages and with headings. Abbreviations used in the tables should be explained in the heading or in a footnote to the table. If the abbreviations appear in many tables, reference may, however, be given to a previous table. Each table and figure should deal with a specific issue, and should be possible to read and understand on its own. In order that readers can easily see changes in the data, present results in a table in which changes are displayed in adjacent columns. In both tables and figures mark those changes that are statistically significant. Figure legends should be submitted on a separate page. The comment on abbreviations refers also to figures.

Be aware of the quality of the figures, so that they appear well in the printed article. Most journals will not publish figures in colour, except where this is essential (e.g. for histochemical preparations, or images from specific recordings, such as positron emission tomography (PET) or magnetic resonance imaging (MRI)).

A letter from the copyright holder must accompany material (e.g. tables and figures) that has been published previously, setting out permission to reproduce the material.

Language editing

For a paper written in a language that is not native to the authors, it is advisable to employ a professional language expert to correct the manuscript before it is submitted. Linguistic errors can make the manuscript difficult to read and this may affect peer review evaluation of the manuscript. This is also true for manuscripts written by native English-speaking authors, who may also need support from a language expert. You should be prepared to make further language corrections during any revision of the manuscript. Most journals will, in addition, use copyeditors for language correction during preparation of the paper for publication.

Authorship

Further areas for consideration are the requirements for being listed as an author, and the order in which authors should appear. Criteria for authorship are set out in the *Uniform requirements for manuscripts submitted to biomedical journals* (3). The conditions that should all be met in order to qualify for authorship of a scientific paper, are:

- having made substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data;
- having drafted or revised the article;
- approval of the final version.

When submitting a manuscript to the journal the covering letter should state that all of these requirements have been met by all authors. Unfortunately, in many cases only some of the authors fulfil all of the requirements. For example, to arrange only for funding or the collection of samples or data does not justify authorship, nor does general supervision of a research group. Such contributions to a study would be best met by a mention in the Acknowledgements.

There are certain informal “rules” concerning the order of authors. The main author, who has contributed most to the study, should be placed first. An author with a more essential contribution than the others can be placed second. Sometimes it may be difficult to separate the contribution of these two authors, and in such cases it may be possible to mention in a footnote that they have contributed equally. The project leader is usually placed at the end, unless that person has also acted as the main author, in which case he/she can be placed first. The other authors can either be placed alphabetically or according to ranking. Some journals require the contributions of each author to be stated in a letter at submission or in the article itself, especially if there are more than a certain number of authors. With many persons participating in a study, as in multicentre studies, some journals prefer a limited number of persons to be listed as authors and the rest as contributors.

Case reports

If you have an unusual or very illustrative case (or a number of cases), you may want to write a Case report, which most journals will publish.

Case reports will also go through the peer review process. Try to decide what your message is, and bear in mind that it should have general interest and usually be written from an educational or clinical point of view. Do not write a long Introduction but just one or two sentences explaining why the Case report should be published and indicating its uniqueness. Confidentiality is essential. Keep the Discussion short; usually it is the clinical implications that are essential. A full scientific review or discussion cannot be given in a Case report.

Letters to the Editor

Most journals also accept Letters to the Editor, and these usually will be published without the regular peer review process. The purpose of a Letter may be to discuss a published paper, giving some complementary information or alternative interpretation of the results, or to briefly present some general scientific or political comment or information, where the Letter maybe the most appropriate form. If the Letter concerns a previously published article, the authors of that article should be given the opportunity to publish a reply; if you are in this position, do take the opportunity to reply. Overall, Letters should be kept short and precise.

Review articles

Review articles may be invited or non-invited. If you intend to submit a non-invited review consult the journal first and follow any specific instructions on such a paper. A review can be systematic, based on a systematic literature survey with the methodology for that described, or narrative. There is usually a preference for systematic reviews unless a narrative review is written for a specific purpose as educational.

SUBMITTING YOUR ARTICLE

Before submitting the manuscript, check that you have included full names (initials or first names and initials), academic degrees and affiliations for all authors, full contact details for the corresponding author (telephone, e-mail, full postal address), evidence of “in press” citations, and preferably copies of such articles.

The manuscript should be accompanied by a covering letter, stating that the manuscript is being submitted to that specific journal and why, giving a concise description of the key message of the results and their relevance for readers, that the research has not been published previously, and that the paper is not under consideration for publication by any other journal. If the study has been published previously as an abstract or in another language, but you want to publish it again, this should also be mentioned in the covering letter. It should be stated that all authors have approved the final version of the manuscript (some journals may require a signed statement from each author), and that all authors fulfil the criteria

for authorship. Conflicts of interest could also be mentioned here as well as in the text. Some journals will ask you to provide a list of potential reviewers from which the Editor can choose.

Authors must submit their paper to only to one journal at a time. This should be stated in the covering letter. A journal may have specific rules about studies presented previously as abstracts or papers previously published in another language. Consult the journal on these issues. Direct double publication is allowed only in exceptional circumstances that have been approved by the journals in question; otherwise it would be regarded as scientific misconduct.

If you want to withdraw a manuscript from publication after submission, you may do so by informing the journal. If a paper is rejected you may submit it to another journal after any necessary revision.

HANDLING OF THE MANUSCRIPT BY THE JOURNAL

Manuscripts are usually submitted to the journal electronically with a covering letter, as described above. At the Editorial office the manuscript will be registered and given a number. All manuscripts will then be evaluated by the Editor, who will determine their further handling. Some manuscripts may be rejected immediately without peer review, perhaps after discussion within the Editorial Board. Probably the most common reason for an immediate rejection is that a paper does not fall into the main scope of the journal. Another type of journal may sometimes be recommended, partly for the authors to better understand why their paper has been rejected. Another reason would be that the study and manuscript do not fulfil the basic criteria for a scientific publication.

Peer review

When deciding to start the peer review process the manuscript may be sent to an Associate Editor or Co-Editor, or handled further by the Editor. The handling of the manuscript will, in principle, be the same.

Reviewers will be chosen by the Editors. They should be experts in the field, maybe with some different expertise. Specific fields of expertise could be scientific and/or clinical experience of similar studies and of the methodology. The reviewers should not be biased, and if they consider so themselves after selection, it is important that this is reported to the Editor so that another reviewer can be chosen. Some journals ask the authors for suggestions for reviewers and one of these may be selected. Sometimes authors also indicate potential reviewers who they consider to be biased or inappropriate. Journals usually have an Editorial Board or corresponding group of scientists who are specifically selected according to their scientific experience in the area of the journal. These persons should also be prepared to act as reviewers. The majority of reviewers are, however, external and not selected from the Editorial Board,

and journals will then have a list of reviewers with their area of expertise. In addition, specific reviewers or consultants may be selected for evaluation of the statistical methodology and presentation or for other specific methodological aspects.

The original meaning of “peer” is a person of equal standing. The term “peer review” is now usually used to refer to persons in the same field of research who are of the same or higher rank. The review process can be double-blinded, which means that both the authors and the reviewers are unaware of their respective identity. However, that must mean that references to the authors ought also to be blinded and there may be problems in referring to previous studies or methodology published by the authors, rendering the manuscript not fully comprehensible to the reviewers. Experience has also shown that the reviewers may rather easily guess the origin of a manuscript. A single-blinded review process means that the reviewers are blinded for the authors but not the authors for the reviewers. This is a rather common practice, but there may be objections in principle to not treating the authors and reviewers equally. However, this does appear to be a practical and accepted procedure. Some journals practice full open reviews and may even make the full review process available online. There may be a risk in this, that some reviewers will decline to participate, but usually even this procedure is well accepted and in principle is the most correct. There are reports that, in this case, the main conclusion about the quality and recommendation about acceptance of a paper is no different from a blinded review, but the wording in the review may be more carefully written.

The main benefits of the peer review process are to support the Editor in choosing the manuscripts to be published and to improve the quality of the manuscripts. In the guidelines for reviewers the aspects usually mentioned are:

- How relevant is the study to the field (rehabilitation)?
- Is there new or confirmatory information?
- Is the title adequate?
- Does the abstract include essential parts of the article and does it follow the instructions for an abstract for that particular journal?
- Does the introduction state why the study was performed and the aim of the study?
- Are methods and subjects/patients adequately described using adequate references and so that the study can be replicated?
- Are the results clearly described and without duplication in text, tables and figures?
- Is the discussion adequate with respect to results and does it include comments on merits and weaknesses (limitations) of the study and its clinical or practical consequences?
- Are the references adequate and do they comply with the journal's rules?

There are usually specific forms for the reviewers to complete, perhaps with a checklist and ratings of different aspects of the manuscript to be completed. These are not presented to the authors but serve as a

standardized type of information for the Editor. Agreement between the different reviewers can thus be evaluated.

Acceptance and rejection

After receiving the comments from the reviewers, some confidential to the editor, some for the authors, the Editor will decide whether the paper can be accepted as it is (although this is very rare), whether it needs minor or major revision before being further considered for the decision of publication, or if it will be rejected at this point. Sometimes the reviewers do not agree on the recommendation for publication and the Editor has decide based on his or her own competence and judgement, or use an additional reviewer. Thus it may take some extra time for authors to be given a final decision. This delay should be communicated to the authors. If a delay in receiving a decision is unexpectedly long (e.g. more than two to three months), authors should enquire of the Editorial office about the progress of their manuscript.

The reviewers' and Editor's comments will provide the basis for revision of the manuscript; authors should read them carefully and answer them point-by-point. It is important to take all comments into consideration and if you do not agree state your reasons. This may be the basis for further discussion between the authors and the Editor and reviewers, respectively. It may be suggested that the manuscript should be shortened and tables and figures excluded; as they may be too many for the specific journal or unnecessary. Occasionally it may be suggested that a manuscript is changed into a Short communication or similar type of publication.

The revised manuscript must to be returned to the Editorial office within the advised time. The paper may be rejected if the revision does not satisfy the Editor and reviewers (the final decision being made by the Editor) and further revision is not considered appropriate, it may be sent back to the authors for further revision, or accepted in the revised form. Even a third version of the manuscript may be sent out to reviewers for further comments. Unfortunately, even after revision some manuscripts have to be rejected.

After acceptance, the manuscript will be handled by the Editorial office and, if relevant, by a separate publisher with further language checking, copyediting to format the manuscript in detail according to the rules of the journal, proof reading at the office and by the authors, when specific questions about details in the paper may be put to the authors before the manuscript is finalized. After acceptance there should not be too long a delay before the manuscript is published, but a period of some months is usual. Many journals nowadays make accepted manuscripts available for electronic preview as quickly as possible and publish an electronic version for the subscribers before the paper version is distributed.

The rejection rate varies considerably between journals. Long-established medical journals may have a rejection rate of 80-90% or even more. Clinical journals, for example in physical and rehabilitation medicine, may have rejection rates in the order of 40-65%. Electronic journals, which in principle have unlimited space, may have lower rejection rates, mainly based on the quality of the manuscripts.

DEALING WITH REJECTION

What should you do if you consider the decision to reject your paper unfair? It is always possible to make a complaint to the journal. Very occasionally you may then be invited to submit a new manuscript, which will probably be treated as a new submission, but usually you will have to accept the decision, perhaps with some further explanation. There is no official means of complaining about an article not being accepted for publication.

If a manuscript is rejected, take advantage of the comments from the peer review process. Try to understand why it was rejected; this may have been explained by the Editor. Possible reasons are that the manuscript does not fit with the main scope of the journal; the journals must, due to lack of space and to keep the publication time down, reject a certain proportion of manuscripts; or there were serious limitations in the study or manuscript, even after revision. Ask yourself if the publication was premature with too small a material and whether you should extend the study before publication; if there is an inadequate design or principle errors in the treatment of data; or if there is too little new and relevant information to justify publication.

Re-submission

If you submit the paper to another journal, which you are now free to do (and you do not need to mention the previous submission to the “new” journal), take advantage of the comments and make adequate revision before re-submission.

AFTER PUBLICATION

After publication you may receive reactions to your article in several ways. If you receive informal letters directly from interested readers, you can take the opportunity to discuss the paper further; this may result in future collaboration. Sometimes a reader will write a formal Letter to the Editor of the journal in which the article was published (as described earlier). A Letter may give additional evidence regarding your findings, or express some contradictory experience or evidence on the topic. It may also just extend the discussion in the area. The Editorial office will send you a copy of the Letter and you will be invited to write a reply (see above), and this will then probably be published in connection with the original Letter to the Editor. Journals usually look positively on such an extended discussion of published papers.

CONCLUDING REMARKS

The first logical step before writing and submitting a scientific paper would be to ask yourself questions such as:

- Is the study ready to be published with respect to the message to be presented?
- Is it possible to draw conclusions based on the size of the group(s) studied?
- What is the statistical power of the study?
- Are the methods used sufficiently documented and controlled in the current setting (country, type of patients, etc.)?

It is then recommended that you take some preliminary decision as to which journal you will submit the manuscript, taking such aspects into consideration as the main scope of the journal, previous experience with that journal regarding the delay in answering and quality of review comments, the publication policy of the journal including open access and, to some extent, the impact factor. When that decision is made, read the journal's Instructions to authors carefully, and after writing the manuscript read them again to check that they have been followed. Justify the presence of all authors and make sure that the original version of the manuscript has been approved by all authors and that any conflicts of interest are reported. Be aware of the quality of the language and use a native-speaker to check the language if the authors themselves do not include such a person. It is useful to understand the procedure at the Editorial office and the peer review process. It is becoming increasingly common to be able to track the progress of your manuscript online. Comments from the Editor and reviewers should be read carefully and answered point-by-point, and if the authors have a different opinion this should be indicated, with reasons. If a manuscript is rejected, consider why this might be; maybe it was premature, or the journal did not have the topic of the article in its main scope. Take all of the comments into consideration and re-submit a revised manuscript to another journal that may be more appropriate for the topic or which does not have such a high rejection rate.

REFERENCES

1. Hall GM, editor. How to write a paper. London: BMJ Publishing Group; 1994.
2. Wager E, Godlee F, Jefferson T. How to survive peer review. London: BMJ Books; 2002.
3. International Committee of Medical Journal Editors (ICMJE). Uniform requirements for manuscripts submitted to biomedical journals. Philadelphia, PA: ICMJE; 2009. Available from: URL: <http://www.ICMJE.org>
4. Reinhardt JD, Hofer P, Arenz S, Stucki G. Organizing human functioning and rehabilitation research into distinct scientific fields. Part III: Scientific journals. *J Rehabil Med* 2007; 39: 308-322.
5. Stucki G, Grimby G. Organizing human functioning and rehabilitation research into distinct scientific fields. Part I: Developing a comprehensive structure from the cell to society. *J Rehabil Med* 2007; 39: 293-298.
6. Medical Subjects Headings. Bethesda, MD; 1966. Available from URL: <http://www.nlm.nih.gov/mesh/MBrowser.html>.
7. World Health Organization. International Classification of Functioning, Disability and Health: ICF. Geneva: WHO Publishing; 2001.

8. Altman DG, Schultz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. CONSORT Group (Consolidated Standards of Reporting Trials). The revised COINSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001; 134: 663-694. Available from: URL: <http://www.consort-statement.org/>
9. Boutron I, Moher D, Altman D, Schultz KF, Ravaud P for the CONSOPRT Group. Methods and processes of the CONSORT Group: example of an extension for trials assessing non-pharmacological treatments. *Ann Intern Med* 2008; 148: W-60-W-66.
10. Svensson E. Guidelines to statistical evaluation of data from rating scales and questionnaires. *J Rehabil Med* 2001; 33: 47-48.

FRANCO FRANCHIGNONI¹
STEFANO NEGRINI²

THE PEER REVIEW PROCESS

¹ Unit of Occupational Rehabilitation and Ergonomics, Salvatore Maugeri Foundation - Clinica del Lavoro e della Riabilitazione, IRCCS, Veruno (NO), Italy

² Italian Scientific Spine Institute (ISICO), Milan, Italy & Editor in Chief of European Journal of Physical and Rehabilitation Medicine

In every discipline judgement by peers is one of the main (perhaps the most important) tenets for quality acknowledgement. In the process of scientific publication, peer judgement has been gradually structured into the so-called 'peer review process'.



According to the 'Uniform Requirements for Manuscripts Submitted to Bio-medical Journals: Ethical Considerations in the Conduct and Reporting of Research', the peer review is an "*unbiased, independent, critical assessment that is an intrinsic part of all scholarly work, including the scientific process*" (1). Peer review

of manuscripts submitted to journals is usually done by experts who are not part of the editorial staff. Its aim is to help authors in revising and strengthening their manuscript (thanks to constructive and professional comments) and editors in deciding which manuscripts are suitable for their journals (putting each paper in the proper context and perspective to make a decision about its acceptance/revision) (2). Thus, first of all it should be an educational process, that allows authors to produce better works. As such it represents an extension of the basic principles of science and scholarship and an essential step prior to biomedical publication.

The peer review process is widely accepted, but its actual value has been insufficiently studied, and little is known about its effects on the quality and utility of published material, and much less about its social, psychological and financial effects (3, 4). Thus, opinions and feelings about peer review are quite different, although the main attitudes about it are as follows: *“Peer review is the worst system for the scientific literature... except for the alternatives”* (5), and *“Peer review is to science what democracy is to politics. It’s not the most efficient mechanism, but it’s the least corruptible”* (6).

In spite of flaws inherent in the process, most of us (whether working as author, reviewer or editor) agree that peer review is necessary because it improves the quality of research reported in medical journals (in terms of logic, clarity, accuracy, utility, etc.) (7, 8) and strongly contributes to determine the standing of a journal among other journals of the same category (9).

A peer-reviewed journal submits most of its published research articles for outside review. This process is essential, because editors of specialist journals can not be expert in all aspects of their scientific field. At the same time a key role is played by the Editor in Chief, often with the support of his/her close editorial team, usually Associate and Assistant Editors. In fact, the uniformity of editorial judgement (i.e. a fair and consistent decision on all different papers submitted) is a key issue, and this can be guaranteed only by the final overall supervision of one person, sometimes with help from close editorial team.

In this chapter we will review the characteristics of the classical peer review process (fig. 1). These have been organised according to the three main protagonists of this complex human behaviour involving many psychosocial interactions: editors, reviewers and authors. The paragraphs will discuss the main aspects of this process, including those connected with the recent evolution of Evidence Based Medicine. Finally, a checklist to perform a careful review process is given in the Appendix. Some other details on this topic can be found in the paragraph ‘Handling of the manuscript by the journal’ of Grimby’s chapter in this book. Of course, the electronic revolution and associated change in culture and readers’ habits are changing year by year many features of this classical process, and are also moving this system towards a greater global transparency.

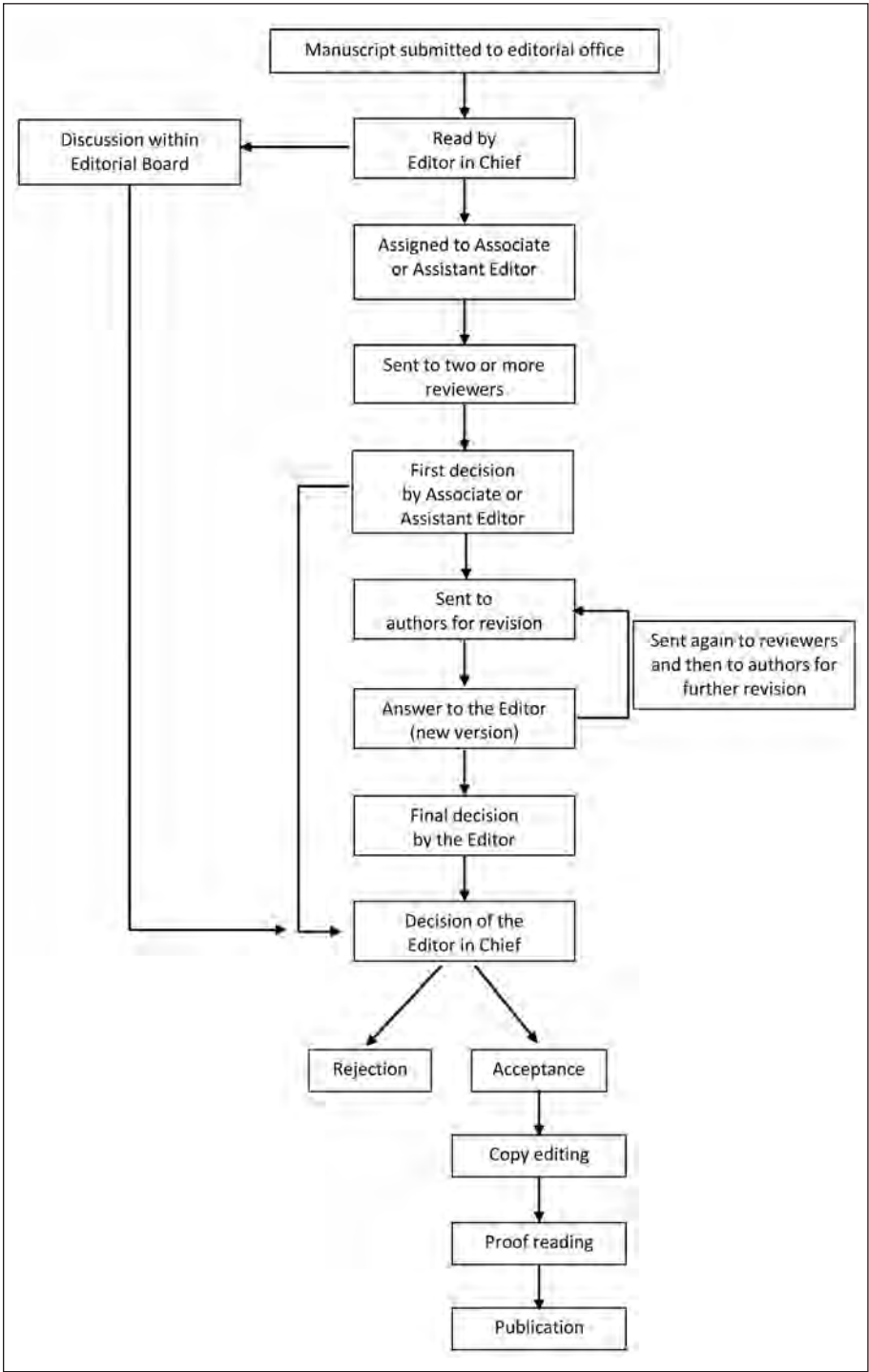


FIGURE 1. Flowchart of a classical peer-review process (with permission of the European Journal of Physical and Rehabilitation Medicine).

1. EDITORS

The editors' main perspective in overseeing the peer review process is: on one side, to increase the general quality of the published papers, in order to give readers appropriate material for their professional development; on the other side, to serve authors, offering them both a timely and fair peer-review process and advice to increase their writing and research skills (10). Beyond these two main aims, other factors can interfere (negatively or positively) in the process, namely:

- the general pressure to increase the bibliometric indexes used to judge the journal quality (including Impact Factor);
- the relationship of the journal with scientific societies and other professional bodies in its discipline;
- the evolution of the publishing world, driven by modern technologies, in particular Internet (open access, on-line submission, and so on);
- the shift of readership that more and more is characterized by people interested in single papers rather than in a full issue; and, last but not least,
- the economic sustainability of the work done.

Based on all these premises, the broad critical objectives for a PRM journal could be summarized as follows:

- to maintain the highest standards of editorial integrity and ethics;
- to publish original, well-documented, peer-reviewed articles;
- to provide physicians with continuing education in order to support informed professional decisions (often including non-clinical aspects of the discipline, e.g. political, philosophic, ethical, legal, environmental, economic, historical, and cultural aspects);
- to help elevate the quality of medical care and research in the specific field;
- to foster responsible and balanced debate on issues that affect the discipline;
- to produce a journal that is reliable and enjoyable to read.

The editor's task is therefore demanding and complex, because editors have to deal in making their final decision with, besides the above factors, also page limitations and priority considerations, whereas reviewer's reports generally aim at improvement of a manuscript and not at its rejection. Based on all these considerations, the editors set and maintain the standards for their particular journal.

1.1 First decisions within the Editorial Board

The first decision for the editor is whether the paper should be sent for external review. Crucial factors are timeliness and necessity of using reviewers with parsimony (due to their chronic shortage) (10). Top journals aim to reach the first 'in-house' decision (i.e. involving editors only) within three weeks of submission and reject a great part of all submission (up to 65%) without external peer review. This behaviour has the advan-

tage that it does not waste authors' time, allowing them to get on and submit the work elsewhere without undue delay. The other side of the coin is that usually these quick answers for rejection do not offer advice for improving the manuscript.

The usual reasons for rejection at this stage are: presence of fatal flaws, insufficient originality (i.e. importance to medicine), or the absence of a message that is central to the journal's mission and interesting to the readership (11). At this stage, for research articles editors focus mainly on the research question asked: even when the overall subject is relevant, topical and important the article may be rejected because the study didn't ask a research question that added enough to current knowledge.

Moreover, the authors' names and affiliation allow editors to look at other papers on that topic coming from the same group, in order to judge if the submission represents a logical and creative evolution of a scientific idea or a redundant (or duplicate) publication.

The careful reading of the text starts with an analysis of the abstract: this is one of the main reasons why abstracts should be as complete, accurate, and clear as possible. Of course, one cannot accurately assess the appropriateness of the abstract until the whole manuscript has been reviewed.

At this stage and later, the credibility of published articles depends also on how well conflict of interest is handled during writing, peer review, and editorial decision making (12). Conflict of interest exists when an author (or the author's institution), reviewer, or editor has financial or personal relationships that may influence their judgements on what is published: they are personal, commercial, political, academic or financial. It is a condition not a behaviour, that can vary from negligible to highly critical for influencing judgment. Clear financial relationships (such as employment, stock ownership, honoraria, payment for lectures or travel, company support for staff, and so on) are the most easily identifiable conflicts of interest and the most likely to undermine the credibility of the journal, of the authors, and of science itself. However, conflicts can occur for other reasons, such as personal relationships, academic competition, research funding, and intellectual passion. Thus, all participants in the peer-review and publication process must disclose all relationships that could be viewed as potential conflicts of interest. Editors may use this information as a basis for editorial decisions.

We will not speak here about problems connected with study design, ethical approval, informed consent, authors' responsibility (including authorship requirements) and so on, because these important issues are the topic of dedicated chapters in this book, and of an interesting publication available on the website of the World Association of Medical Editors (13).

If the manuscript is potentially suitable for publication in the journal, the editor will begin the process of external peer review. Strategies for this process can vary depending on the journal's history and resources. For example, journals may differ in the amount of time given

for doing the review, the number of reviewers involved, the reminders they use when reviews are late, the process for yielding unresponsive reviewers, and so on.

Among major factors that affect the quality of the peer-review process, we will discuss the preliminary editorial choices, the reviewer's review, and the editor's decision based on reviews.

1.2 Preliminary editorial choices

a. Reviewer selection - For example, the reviews might be improved by assuring that the manuscripts are well matched to the reviewers. The selection is usually based not only on their expertise in the manuscript topic, but also on their past performance (e.g., how many reviews the peer-reviewer has done, when and how). Thus, continuing to be asked to review depends also on the reviewer's track record.

In general, a careful reviewer screening is warranted, and technical expertise is increased by specific professional experiences. Sometimes an expert in statistics or epidemiology is also needed for advice. Furthermore, the most productive and skilled reviewers are always overloaded and thus journals need to avoid sending too many papers to their best reviewers (10).

Unfortunately, if a journal is successful in attracting more papers, it is reasonable that it will need more reviewers. As a result, editors know that the search for reviewers is an ongoing struggle and usually several attempts are required to find (competent?) referees who will agree to review a paper. Clearly, peer reviewing would benefit from both more recognition and reward (see later in this chapter).

In order to partially reduce the shortage of good reviewers, some journals ask authors to suggest names of experts to consider as possible peer reviewers for that paper, as well as to supply a short list of individuals who should not review their manuscript (due to potential biases). As an example, Archives of Physical Medicine and Rehabilitation writes "*Include in the cover letter accompanying a submission the names and addresses of 3 potential manuscript reviewers. The editors may seek reviews from others. Authors should not recommend as potential reviewers current members or associate members of Archives's Editorial Board, or people who are affiliated with the authors' institutions*". Sometimes, the suggested reviewers are not used for that specific paper but are inserted in the pool of potential reviewers for that topic. Compared to editor-selected reviewers, author-suggested reviewers would seem more likely to praise the research and recommend acceptance. But authors should be aware that the final decision on acceptance rests with the editor, not the reviewers.

b. Blinding procedures - The double-blind review seems to be the preferred form and the most effective (8). The most common form of blinding is to conceal the reviewer's identity from the author. Of course,

this should not allow the reviewers to be rude in their comments, because impolite feedbacks can hurt authors, especially young researchers. Anyway, it is a duty of the editor to safeguard also this point. In contrast, when the journal does not blind the reviewer's identity, it has been reported that more reviewers decline to review or are more reluctant to criticise a manuscript (particularly those of a colleague or personal acquaintance), and it is reasonable that acrimony in the scientific field would increase (14).

On the other side, blinding of reviewers to authors' identity (that sometimes may be ineffective, due to self-reference, the very specific topic, or the author's identifiable knowledge of the subject matter) is usually done to reduce bias against or for the authors. For example, possible biases when the author's identity is not masked are connected with partiality on the basis of personal characteristics (e.g. sex or minority biases), geographical bias (overall, reviewers assign higher priority ratings to manuscripts from their own country than from other countries (15), trend to praise well-known authors (senior vs. junior author bias), camaraderie and so on. But, often experts in a specific field - who are called upon to review a paper, as ideal candidates in providing a valuable review - are well aware of what research is being performed by fellow colleagues, and may potentially be the main 'competitors' for leadership in this field. Thus, in the end we feel that the moral integrity of the reviewers is more important than any blinding process. The alternative is often having reviewers who are not fully capable of understanding the intricacies of that research.

1.3 Quality of reviewers' review

Jefferson et al. recently examined a series of studies regarding the quality of the editorial peer-review process in biomedical journals (effects of concealing reviewer/author identity, use of checklist, etc.) (14). Another study (16) analyzed the papers measuring the effects of editorial peer review on the quality of published articles and of the reviewers' comments. Overall, the conclusion was that editorial peer review is largely untested, and that - until the main objectives of the peer review are properly defined - it is impossible to assess (and improve) the effectiveness of the process.

Suggestions for strengthening peer review have included improved training for reviewers (i.e. dedicated workshops), use of standardized assessment forms and structured feedback to the reviewers. Most of these interventions - even if they have failed in some studies to demonstrably improve the quality of the review (2, 17) - are in our opinion to be taken into serious account. In fact, at present the absence of evidence on efficacy and effectiveness cannot be interpreted as evidence of their absence (4).

For sure, detailed structured review forms could facilitate the delivery of useful feedback and reduce idiosyncratic answers. A set of descriptive review criteria for research manuscripts should inform and remind

reviewers which aspects of the manuscript should be carefully considered when generating a review. These criteria should be straightforward and regularly updated. The goal is not only standardization but also transparency. The most frequent questions are related to the following broad issues: How relevant is the study for the mission of the journal? What is the importance of the study with respect to new or confirmatory information? [Originality vs. redundancy]. Are title and key words adequate? Does the abstract reflect all essential parts of the manuscript? Is the aim of the study well defined? Are methods appropriate and adequately described and applied? Are results clearly presented? Are figures and tables - if present - adequate? [Is there any duplication and/or need for reduction in text, tables and/or figures?] Is discussion adequate with respect to results? Are the merits and weaknesses of the study considered? Are the references up-to-date and relevant? [Are any important references missing?]. Are there any aspects of the language that need to be addressed?

The forms vary in the type of issues that reviewers are asked to address. More detailed criteria regarding this field are listed in the Appendix, that is based on the suggestions of important publications (10) and our personal experience. Reviewers do not have to answer all questions, but use their judgement in deciding how to write their comments that are mainly focused on describing if the different parts of the manuscript meet expectations for a scientific paper and on making a recommendation regarding publishability.

This kind of checklist can be useful also to the authors when submitting their paper, in order to verify one by one before submission all main aspects of their scientific work.

Finally, the editorial office should assist the reviewers appropriately and provide them also with journal-specific reviewer training material. For these aspects see for example <http://resources.bmj.com/bmj/reviewers> and <http://resources.bmj.com/bmj/reviewers/peer-reviewers-guidance>. Similarly, it would be useful for reviewers to receive a copy of the decision letter(s) sent to the author, together with the other reviewers' comments.

Finally, the auditing process. Whenever a process is in operation, checking it is fundamental to improve quality (18). Auditing is always crucial to internally evaluate the consistency, objectivity and quality of the peer-review process.

Not all authors (or even reviewers) are aware that editors take care how submitted papers have been handled by reviewers. This is done for many purposes, not only for the selection of good reviewers (and rejection of bad ones), but also - in the longer term - selection of members of the Editorial Board. The analysis of the reviewer quality is usually done at a high level in the editorial team (Editor in Chief with the Associate & Assistant Editors).

The first way to assess a reviewer's work is through some statistics that can easily be drawn up by the publishing team: number of papers accepted/refused for review per month/year; percentage of accepted/rejected papers (and at which stage of the review process); average time re-

quired to perform a review. Moreover, specific scales to check quality of the reviewer's answers can be used: in these scales many factors are judged, like completeness of review, details given, focus on methodological issues, usefulness of the review to authors, and so on (16).

1.4 The editor decisions

Once the answers from the reviewer have been received, often their agreement is not high (19) in terms of both general judgement and specific comments. This is not surprising, because several different areas of expertise may be relevant for a given submission (i.e. methodological and clinical), and reviewers are likely to be more familiar with some areas than others (i.e. statistical analysis, outcome measures, study design, etc.). This broadening of expertise is also required by the increasing number of multi-authored manuscripts spanning several disciplines.

As a result, often reviewers (and editor) recognise different aspects of scientific quality. This does not appear as a limit, and explains why only the combined recommendations seem to adequately predict acceptance or rejection (18).

When the editors receive the comments from reviewers they should be aware of and take into account all intrinsic (and sometimes unconscious) biases in the review process, and make an effort to overcome prejudices and partiality. Moreover, while presentation quality is the author's responsibility, care needs to be taken to ensure that studies are reviewed primarily on their scientific merit, and not only on writing style. The letter to the authors should contain also the editor's personal comments, including the priority in considering the reviewers' comments (that may be conflicting) and an unambiguous instruction about what is required, encouraged but not required, and optional. For this reason, it seems incredible that a survey of North American medical journal editors found that almost half of them based their decision solely on reviewers' comments (10)!

If (major or minor) changes are recommended, the editor can ask for a resubmission for re-evaluation after revision according to the comments made by the reviewers, along with a point-to-point response to the reviewers' comments and a list of all changes made (with a deadline for the authors' reply). Where major revision is needed, a definite decision about the manuscript's acceptance usually can not be made.

Indeed, a certain heterogeneity in the quality of peer reviews exists. In general, 20-25% of reviews are great, 20-25% scarcely useful, and the rest are in between (20). Sometimes, the reviewer may have spent considerable time reviewing the manuscript but with little profit for the purposes of the journal. According to Provenzale et al. (21), the most frequent cases of scarcely helpful (and/or badly written) reviews are as follows. First, there are some reviews that are composed of a few generic sentences (sometimes only for the editor), with a recommendation of accepting or not (the so-called 'snapshot verdict'). This does not allow to

determine if the reviewer read carefully the manuscript, because the comments are non-specific and thus non-informative. A critical and detailed analysis including constructive comments is lacking and thus the editor often needs to send the manuscript to an additional reviewer for an in-depth evaluation (with considerable waste of time). Second, there are other reviews characterized by discordance between the specific comments and the final overall recommendation (the so called 'mixed-signals' or 'hidden-agenda' reviews). Most commonly, the reviewer is highly critical in the narrative summary but then provides a recommendation of 'minor revision required', or the reviewer provides opinions in the confidential section to the Editor that differ from those expressed in the review available to the authors.

Such cases raise the problem if the editor and editorial staff should transmit all comments as received or (as we suggest) edit them, particularly where some comments are rude, unclear, unhelpful or even inappropriate.

In the end, editors are the 'stewards' of journals, have a specific responsibility to the medical community, and - as described in paragraph 1 - have the charge of balancing the points of view of many parts while promoting the science and art of medicine and the general improvement of public health. The decision-making process is highly complex, multifactorial, and unique for each paper. It is subjective, but must be neither impulsive nor uninformed.

On the other hand, we have also to acknowledge that the peer review process is fallible. This imperfection should ensure the humility of the editors and reassure authors that some rejections can be due to questionable decisions. As readers, authors, reviewers and editors we have experienced many of these limitations firsthand.

We do not discuss here issues related to authorship requirements and publication ethics, because they are examined in other chapters of this book. On these topics, we strongly suggest to read also the interesting material published by the Committee on Publication Ethics (COPE, <http://publicationethics.org/>). COPE is a forum for editors and publishers of peer-reviewed journals to discuss issues related to the integrity of work submitted to or published in their journals. It supports and encourages editors to report, catalogue and instigate investigations into ethical problems in the publication process.

2. REVIEWERS

Reviewers are crucial for the peer-review process. The first time you are invited to do a review, you probably feel you have reached another goal in your career (to have been considered able to judge the colleagues' work). And here starts the problem. In fact, often you will have to do this work with little guidance, usually starting from the basis on having undergone the process yourself. In this way, your judgement could come out

as valuable or biased according to your own experience, character, and interests. If you previously met helpful reviewers, who allowed you to improve the quality of your paper, it is probable you will try to adopt the same strategy. On the other hand, if you previously had awful experiences of rejection without any explanation (maybe a long time after submission), you could be in a mood that does not facilitate good reviews.

Conversely, before starting this process it would be extremely useful for a reviewer to read some basic interesting publications on this topic (3, 10, 13, 21, 22). In general, ability to express balanced judgements, keep a sense of proportion, know the journal's standards and communicate complex ideas in a clear and concise way is not widespread and easy-to-learn. Moreover, humility and real interest in science can help in doing a good review.

2.1 General responsibilities and attitudes

Reviewers should accept an invitation to review an article only if they judge the topic as of interest and within their expertise, and can commit the appropriate time (10).

The main responsibilities of a peer reviewer are as follows:

- to treat the manuscript as confidential;
- to inform editors - before agreeing to review a submission - about any potential competing interest;
- to maintain objectivity;
- to comment on ethical questions and possible research misconduct raised by submissions (e.g., unethical research design; insufficient detail on patient consent or protection of research subjects, including animals; plagiarism, falsification or redundant publications; etc.);
- to provide within the set time limit a constructive and professional review, avoiding personal comments and assessing clarity, importance, quality and credibility of the paper.

The reviewer should identify and comment at best on major positive and negative characteristics of study design and methods, and on the authors' interpretation of data (including acknowledgement of the data's limitations). To do so, the reviewer should have (22):

1. motivation, i.e. sense of responsibility towards one's colleagues and strong conviction that a high standard in peer review is critically important for the progress of science;
2. scientific expertise, i.e. awareness of the literature and mastery of the relevant science (assistance can be asked for specific topics);
3. helpful and respectful attitude, i.e. ability to read with patience, objectivity, and openness towards new ideas and approaches, and then to report with complete clarity and without summarily closing off debate;
4. time, because for many manuscripts a few hours are not enough for a careful and helpful review, and a complex and potentially important paper can certainly take a full working day.

2.2 Starting the review

During the first reading, the abstract should preliminarily suggest what is important to look for in the text, while the text demonstrates the general level of ‘scientific thinking’ of the authors. If the manuscript shows unclear or troubling points, probably the authors failed to put their work into a fair and full perspective (22). Typical problems with the presentations span from redundancies, irrelevancies and unnecessary excursions, to unclear definitions of terms or concepts. More generally, authors may not be able to optimally lead the reader through their scientific thinking. Thus, the art of scientific writing is overlooked, and the sentences are not focused, concise, informative and well-organized, as expected. On the other hand, violations of logic or of common sense sometimes exist (e.g. unwarranted conclusions, incorrect attribution of causation, inappropriate extrapolations, etc.), or there is a lack of application of some basic scientific principles. Attention should also be paid to citations, which must be complete and worthwhile.

At this point, the reviewer should make preliminary notations on the text regarding what can be improved or is missing. Then, it is wise to put the manuscript aside for a day or more. If specific problems are identified, often there is need to thoroughly check the literature, consult a colleague, or think calmly about the possible main sources of the drawback and how they could be solved (if possible).

When the reviewer returns to the manuscript for a second reading, it is time to judge in detail the quality of scientific reasoning, and the novelty and interest of the methodology, results and conclusions. Of course, there is no need to comment point-by-point the style (even if suggestions for improving clarity, accuracy and utility of the submission are always welcome), but it is our thinking that if the reviewer has difficulty on some point, after careful reading, so too will many other readers.

2.3 Writing the response to editors and authors

Reviewers have the responsibility to give a proper response to the editors as well as to the authors. Failing to do this to one of the two makes the review flawed and not very useful (see also the paragraph ‘The editor decision’). The reviewer response is usually structured in two parts: one for the authors, the other for the editor.

The reviewer’s comments to the authors should be:

- clear. Each criticism/question must be explained and never left as an unsupported statement. Often it is important to add a short example or reference to clarify a remark;
- constructive and supportive (but avoiding useless praise);
- respectful, and thus without any disapproval or offence. Some sentences should reflect suspended judgment, pending a response from the authors, who thus have the opportunity to respond, even in the case of a potential fatal flaw.

The first paragraph should identify the topic of the study, indicate the basic approach, and summarize the main findings and conclusions of the study. Then, reviewers should provide a short paragraph of general comments, summarizing the strengths and weaknesses of the manuscript. The 'major comments' should clearly state the points requiring a compulsory revision. The final section is usually dedicated to 'minor specific comments', referring to redundancies, misprints, inappropriate symbols and the like. The comments should focus on what needs to be addressed based on the existing methods and data. Usually, it is not realistic to ask for more substantial changes (e.g. suggest that the study might have been conducted differently). However, authors can be requested to rewrite parts of the manuscript for clarity, to further detail the methods, to perform more statistical analyses, to consider additional literature, to enlarge/reduce discussion and conclusion, and so on. Comments should be listed in order of appearance in the text, numbered and identified by page and line.

The (confidential) comments to the editor should:

- list the most relevant general criticisms (in descending order of importance), indicating the importance of each item and what the authors should probably do in response;
- give a recommendation as to the manuscript's fate (e.g. to accept, accept pending revisions, reconsider after major revisions or reject), that is in line with the expressed criticisms.

Some sentences can also reflect suspended judgment pending a response from the authors, who thus have the opportunity to provide an explanation, even in the case of a potential fatal flaw.

The type of review that is most helpful to the editor is one that shows that the reviewer performed a close reading of the manuscript, thought carefully about its most important sections, provided constructive criticisms for the authors, and gave a recommendation/rating adequate to the comments expressed in the review.

2.4 Reviewer's point of view and rewarding

A recent survey conducted on more than 4,000 reviewers (the Peer Review Survey 2009 - http://www.elsevier.com/wps/find/reviewershome.reviewers/ru_issue3a) showed the following main preliminary findings:

- 90% say they review because they believe they are playing an active role in the community; 84% believe that without peer review there would be no control over scientific communication, but only a third (32%) think it is the best that can be achieved;
- 91% say that the last paper they submitted was improved through peer review, the discussion being the biggest area of improvement;
- whilst 86% enjoy reviewing, 56% say there is a lack of guidance on how to review; 68 percent think formal training would help;
- 58% would be less likely to review if their signed report was published;
- 76% favour the double blind system where just the editor knows who the reviewers are.

Another topic connected with the reviewer's work is the answer to the question proposed by an Editorial in *Cortex* (23): "*Why should reviewers put more effort into a process which carries no fame, no money, very little extra knowledge, and for which they bear no direct responsibility?*".

At present, peer reviewing is usually rewarded by naming the reviewer in the journal (once a year or less), writing letters of thanks (not often) or (very rarely) paying the reviewers. Some journals offer to the reviewers a free 1-year (usually online) subscription and provide a journal activity certificate, that keeps track of the work reviewers do for the journal and enables them to print off a certificate.

These kinds of benefits should be more widely adopted, considering the shortage of good reviewers and the importance of a correct acknowledgment of their activity.

In the Elsevier website, there is a section dedicated to the reviewer community, where comments on developments in peer review are included (http://www.elsevier.com/wps/find/reviewershome.reviewers/ru_most/). There, some measures of the annual importance of reviewers to their field have been proposed, e.g. journals could list, certificate and make publicly available the number of peer reviews done by each reviewer.

Certainly, it would be very helpful if - in an increasing number of countries - the professional service done by peer review activity could be also rewarded by academic recognition and/or by the granting of CME credits.

3. AUTHORS

"Peer review produces more angst than an Ingmar Bergman movie" (24).

First of all, authors have to acknowledge that the journal choice can strongly influence reviewers' answers and publication success (25). For this reason, some journals have also produced checklists to help authors decide whether that specific journal is the right one for a given work (<http://resources.bmj.com/bmj/authors/authors/checklists-forms/is-bmj-the-right-journal-for-my-research-article>). In fact, journals can reject papers simply because they consider the manuscript outside their field of interest, and some editors who receive a great number of submissions have to reject also some 'respectable' manuscripts simply due to the overload of incoming material.

Of course, the term 'authors' includes a wide range of people, ranging from beginners to experienced scientific writers. The beginners in scientific writing have very limited experience of peer review and really merit help during the publication process. We strongly suggest to these authors, before submitting a paper, to read at least one academic booklet on peer review, such as 'How to survive peer review' (3), that aims at explaining "*just enough about peer review to enable you to survive and benefit from it,*

and to be a competent reviewer". Papers can be rejected for multiple reasons that go beyond the research contents, and range from incorrect journal selection to difficulties with scientific writing (lack of clarity in contents, poor organization of the material and so on). From the authors' perspective sometimes a rejection is seen as a catastrophe and thus a negative response must be carefully handled (explaining in detail how to proceed and avoid future mistakes). Being rude to these people would be a loss for science.

Conversely, the experienced writers are persons who have already published some papers and are well aware of the publication process. Usually, they carefully chose the journal to which to submit their paper, and are conscious of procedures, timing and possible answers from reviewers. Among them, there are also authors who apply the so-called 'publish or perish' law: they are interested more in publishing in itself, than in the real quality of the content of what they do. Their motivation is often to increase their publication record to promote their academic career, achieve a PhD degree, or justify research funds. This leads to publication strategies such as 'salami slicing' (http://en.wikipedia.org/wiki/Least_publicishable_unit), where data gathered through one research project are separately reported in multiple (or 'spin-off') publications.

In general, authors acknowledge the assistance received by constructive reviewers. But there are also authors who regard the peer-review process as a series of hurdles to vault before their work gets published, and they find it hard to see reviewers as people offering help (22).

As a rule, authors should be aware that if they haven't received the first reply within three months after submission, it is reasonable for them to contact the journal office to politely inquire about the status of the reviewing process.

In the final analysis, authors' satisfaction is rarely associated with quality of the review of the manuscript but rather with the final acceptance of the manuscript. This dramatic influence of the editorial decision on authors' perception of the peer-review process may be disappointing but is not surprising (26): we are human beings! Thus, after a manuscript's refusal or request for a major revision, authors can react differently. There will be those who are so struck as to feel themselves inadequate to the process: this is a big problem if they are young inexperienced researchers who could be stunted in their growth, and it must be appropriately faced during the review process, making it as fair, educative, polite and understandable as possible. The sequence begins with a reaction of hurt, which progresses to anger and frustration (24). It has been ironically described that these authors can be seized by the notion that the 'fools' offering suggestions (i.e. reviewers and editors) did not understand what they wrote, know little about research, spent too little time examining relevant issues, were confused by their own biases, and other more offensive variants.

Only after a (short or long) period of cooling down do most authors enter the phase of reasoned acceptance of the comments (even if some hidden

resentment could remain) (24). In this phase they recognize that in the peer-review process it is almost inevitable to receive criticism, and the comments can provide useful suggestions for manuscript improvement. Thus, after setting spontaneous reactions aside, the authors try to understand how their work can be improved, based on the feedback received.

As Eva says in his editorial (27), the reviewers are “*always right*”, because if they have misunderstood something, this means that the author has failed to make the point clearly, even if he/she can point to a paragraph in the text that explicitly and literally counters the reviewer’s claim. Of course, reviewers can misinterpret things, but they tend to read manuscripts much more carefully than the general readership of any journal does. Thus, if the reviewer has missed something, then the odds are high that some other reader will do the same. As a consequence, what has not been properly caught by the reviewer is really worth correction if the authors do really want to reach their audience and make it understood. Rejected manuscripts still have a reasonable chance of being published within a reasonable time frame by another journal.

On the other hand, it is always possible that the reviewers really did not understand the points raised by the authors. For example, if a paper on ‘postural’ rehabilitation were sent to different journals (respectively dealing with orthopedic surgery, manual therapy, neurological sciences, rehabilitation medicine, and physical therapy), it is highly probable that the response will differ greatly, and rejection probability and type of criticism will differ as well. This happens mainly because reviewers reflect the readership of the journal and thus if the reviewers do not understand the paper, the readers of that journal would not understand it either.

For this reason, the selection of the best journal for submitting a specific research is a delicate responsibility of the authors (25), and problems with reviewers can also underlie an inappropriate journal selection.

3.1 Authors’ response

Authors should ensure they follow the editor’s revision submission instructions. Many editors will require a submission that includes a point-by-point response to the editor’s and the reviewers’ comments, as well as the revised manuscript (with changes highlighted). Authors should try to make as many of the changes requested by the editor and reviewers as they can, but they should not feel obliged to make changes that they believe are not warranted. A proper response should be complete, polite, and based on evidence. If authors do not agree with a comment, they should ensure their refutation is carefully justified.

To help ensure a complete reply, authors should prepare a list of editor’s and reviewers’ comments, with their response written immediately below each point raised. This format makes it easy for the editor and reviewers to see that the authors have considered every comment and minimizes the risk that authors will inadvertently miss a comment.

In order to give an evidence-based response, authors may find it helpful to search the literature again to identify additional, and possibly more recent, supportive publications. If further evidence is not available, authors should explain why the evidence already cited is sufficient. The additional citations and explanations do not necessarily need to be incorporated into the revised manuscript, but they should be included in the response document.

Authors are responsible for deficiencies in presentation. Before any submission, they have to obtain a meticulous critical review from expert colleagues and - if they write in a foreign language - to seek also a language revision by a native-tongue expert.

On the other hand, peer review is usually based on a mix of evidence and opinion. Editors aim for a good and fair appraisal, but appreciate that peer review does not necessarily always lead to the best decision. For this reason, serious appeals are welcome (and some of them succeed), even if this may be a waste of time if the editors think that your work is really not suited to the journal.

CONCLUSIONS - TOWARD AN EBM PEER REVIEW PROCESS

Over the past, to select the best quality evidence from the literature, Evidence Based Medicine has developed careful methodological tools to judge published papers and consequently extract the actual state-of-the-art research to be proposed to clinicians. Gradually they have been also adopted by researchers to produce good papers, and by reviewers to check the studies they were reviewing.

Originally these tools were studied for Randomised Controlled Trials (e.g. CONSORT guidelines), where obviously the biggest efforts were placed, since from these papers best evidence can be gathered: variations of this checklist have been proposed for pragmatic and cluster RCTs as well as for non-pharmacological ones (this is very relevant for PRM). Along the same lines was the QUORUM statement for meta-analysis of RCTs, which has recently developed into the PRISMA checklist (see chapter by Liberati & Taricco, in this book).

Nevertheless it is very well known that in many fields of the literature the only available evidence comes from studies of a lower methodological quality than RCTs, and as a result these were considered as well. As a consequence, today a series of checklist have been produced by STROBE, MOOSE, RATS and STARD commissions. All these checklists are collected in a website by the Equator initiative, where they are constantly updated (<http://www.equator-network.org/resource-centre/library-of-health-research-reporting/reporting-guidelines/>).

In conclusion, while the peer review process is fairly subjective, for the most part it is effective, and it is unlikely that at present there are other viable alternatives. All the methodological improvements proposed in the literature should gradually move into the peer review process to guarantee its gradual quality increase.

Editors feel more comfortable in their decisions when they are well informed by expert opinion, and are aware that their journal's prestige rests to a great extent on the thoroughness and expertise of the reviewers. Also readers can only stand to gain from a peer review process carefully carried out. In addition, in this chapter we have focused on the advantages for authors as well, which are important even if sometimes neglected. The key to the success of the whole process are the only ones who apparently work for free: the reviewers. Nevertheless they are very well aware of the advantages of the process, since they are among the best authors and most attentive readers. All these points have been focused on this chapter.

At the same time, we wish to remind you that scientific journals are "*archives of work done, not of revealed truth*" (9) and we should not confuse efforts to improve quality with the granting of a mark of perfection (10).

REFERENCES

1. International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals. © ICMJE 2009 <http://www.icmje.org/>
2. Wagner AK, Boninger ML, Levy C, Chan L, Gater D, Kirby RL. Peer review: issues in physical medicine and rehabilitation. *Am J Phys Med Rehabil* 2003; 82: 790-802.
3. Wager E, Godlee F, Jefferson T. How to survive peer review. London: BMJ Books, 2002. Available at: <http://resources.bmj.com/bmj/pdfs/wager.pdf>
4. Jefferson T, Rudin M, Brodney Folse S, Davidoff F. Editorial peer review for improving the quality of reports of biomedical studies. *Cochrane Database Syst Rev* 2007 Apr 18; (2): MR000016.
5. DeMaria AN. Peer review: better than the alternatives. *J Am Coll Cardiol* 2002; 40: 1017-8.
6. White C. Little evidence for effectiveness of scientific peer review. *BMJ* 2003; 326(7383): 241.
7. Rennie D. Editorial peer review: its development and rationale. In: Godlee F, Jefferson T, eds. *Peer review in health sciences*. 2nd ed. London: BMJ Books, 2003: 1-13. Available from: <http://resources.bmj.com/bmj/pdfs/rennie.pdf>
8. Ware M. Peer review in scholarly journals: Perspective of the scholarly community - an international study. *Information Services & Use* 2008; 28: 109-112. Full report available from: <http://www.publishingresearch.net/documents/PeerReviewFullPRCReport-final.pdf>
9. Tobin MJ. Rigor of peer review and the standing of a journal. *Am J Respir Crit Care Med* 2002; 166(8): 1013-4.
10. Bordage G, Caellegh AS, Steinecke A, et al.; Joint Task Force of Academic Medicine and the GEA-RIME Committee. Review criteria for research manuscripts. *Acad Med* 2001; 76: 897-978.
11. Woolley KL, Barron JP. Handling manuscript rejection: insights from evidence and experience. *Chest* 2009; 135: 573-7.
12. Committee on Publication Ethics (COPE). Best practice guidelines for journal editors. <http://publicationethics.org/guidelines>
13. Rockwell S. Ethics of peer review: a guide for manuscript reviewers. Available at: http://www.wame.org/ethical_issues_in_peer_review.pdf
14. Jefferson T, Alderson P, Wager E, Davidoff F. Effects of editorial peer review: a systematic review. *JAMA* 2002; 287: 2784-6.
15. Opthof T, Coronel R, Janse MJ. The significance of the peer review process against the background of bias: priority ratings of reviewers and editors and the prediction of citation, the role of geographical bias. *Cardiovasc Res* 2002; 56: 339-46.

16. Jefferson T, Wager E, Davidoff F. Measuring the quality of editorial peer review. *JAMA* 2002; 287: 2786-90.
17. Callaham ML, Knopp RK, Gallagher EJ. Effect of written feedback by editors on quality of reviews: two randomized trials. *JAMA* 2002; 287: 2781-3.
18. Wade D, Tennant A. An audit of the editorial process and peer review in the journal *Clinical rehabilitation*. *Clin Rehabil* 2004; 18: 117-24.
19. Rothwell PM, Martyn CN. Reproducibility of peer review in clinical neuroscience. Is agreement between reviewers any greater than would be expected by chance alone? *Brain* 2000; 123: 1964-9.
20. Dye L. A review of the review process. *J Med Toxicol* 2007; 3: 143-5.
21. Provenzale JM, Stanley RJ. A systematic guide to reviewing a manuscript. *AJR Am J Roentgenol* 2005; 185: 848-54.
22. Hoppin FG Jr. How I review an original scientific article. *Am J Respir Crit Care Med* 2002; 166: 1019-23.
23. Della Sala S, Grafman J. Refereeing mortuus est, vivat refereeing. *Cortex* 2002; 38: 269-71.
24. Rothstein JM. The ugly side of peer review. *Phys Ther* 1995; 75: 582-4.
25. Thompson PJ. How to choose the right journal for your manuscript. *Chest* 2007; 132: 1073-6.
26. Weber EJ, Katz PP, Waeckerle JF, Callaham ML. Author perception of peer review: impact of review quality and acceptance on satisfaction. *JAMA* 2002; 287: 2790-3.
27. Eva KW. The reviewer is always right: peer review of research in Medical Education. *Med Educ* 2009; 43: 2-4.

APPENDIX

PEER-REVIEW: A PRACTICAL CHECKLIST

We will focus here on specific issues that must regularly be faced by reviewers during peer-review. Authors writing a paper can also consider all these issues since, if adequately tackled, the peer review process is more likely to be smooth and rapid, without problems arising. The material is organised in the form of a checklist, and could be appropriate for some Journals, while not for others.

Before accepting to perform the review

The process starts even before accepting to review the paper, with some key questions that need to be answered.

- Do I have time to perform accurately this review?
- Does my expertise adequately cover the topic of this review?
- Do I know sufficiently the journal requesting my review, so to be able to correctly target my work?
- Do I have any conflict of interest with the contents of this paper?

Before writing the review

Some answers should be given already before writing the review:

- Was this research worth being performed?
- Does the research address an important and useful problem (relevant to journal's audience)?
- Has the manuscript been previously published?
- To which manuscript category does this manuscript best conform?
- Are there any potential biases in reviewing this manuscript?
- Is the manuscript of appropriate length (relative to information content)?

The title

The title is the first reason why readers access a manuscript. In the era of Medline and Internet, average readers perform searches on what they are interested in, retrieve a huge number of manuscripts and, first of all, review all titles (sometimes also looking at the journal that published the paper and the authors' names). If a title is able to attract their attention, then the paper is accessed. Consequently, the questions to be answered include:

- Is the title informative about the contents of the paper?
- Does the title contain the key information on the manuscript (topic, methodological strengths, main results)?
- Is the title realistic and correct with regard to the contents of the paper?

The abstract

- The abstract is the second main reason why a paper is accessed, after the title. Consequently the abstract must be informative and well written, to let the reader understand if the paper is really worth his attention. Methods and main Results must be well focused, as well as the main research question and conclusions.
- Does the abstract respect the rules of the journal?
- Does the abstract appropriately summarize the manuscript?
- Are there discrepancies between the abstract and the remainder of the manuscript?
- Can the abstract be understood without reading the manuscript?

The introduction

- Is the Introduction concise?
- Is the purpose of the study clearly defined?
- Do the authors provide a rationale for performing the study based on a review of the medical literature? If so, is it of the appropriate length?
- Do the authors define terms used in the remainder of the manuscript?
- If this manuscript is Original Research, is there a well-defined hypothesis?

The methods section

- Could another investigator reproduce the study using the methods as outlined or are the methods unclear?
- Do the authors justify any choices available to them in their study design (e.g. choices of imaging techniques, analytic tools, or statistical methods)?
- If the authors have stated a hypothesis, have they designed methods that could reasonably allow their hypothesis to be tested?

The results section

- Are the results clearly explained?
- Does the order of presentation of the results parallel the order of presentation of the methods?
- Are the results reasonable and expected, or are they unexpected?
- Are there any results presented that are not preceded by an appropriate discussion in the Methods section?

The discussion section

- Is the discussion concise? If not, how should it be shortened?
- If a hypothesis was proposed, do the authors state whether it was confirmed or disproved? Alternatively, if no hypothesis was proposed, do the authors state whether their research question was answered?
- Are the authors' conclusions justified by the results found in the study?
- If there are unexpected results, do the authors adequately account for them?
- Do the authors note limitations of the study? Are there additional limitations that should be noted?

Figures and graphs

- Are the figures and graphs appropriate and are they appropriately labelled? Would a different figure better illustrate the findings?
- Do the figures and graphs adequately show the important results?
- Do the figure legends provide a clear explanation that allows the figures and graphs to be understood without referring to the remainder of the manuscript?

Tables

- If there are tables, do they appropriately describe the results?
- Should one or more tables be added?

References

- Does the reference list follow the format for the journal? Does it contain errors?
- Have the authors appropriately represented the salient points in the articles cited as references? Alternatively, have the authors misquoted the references?
- Are there important references that are not mentioned that should be noted?
- Are there more references than are necessary?

TOWARDS AN EPIDEMIOLOGY OF FUNCTIONING

¹ Swiss Paraplegic Research (SPF),
Nottwil, Switzerland

² Department of Statistics, Swiss Federal
Institute of Technology, Zurich,
Switzerland

³ Department of Health Sciences
and Health Policy, University
of Lucerne and SPF, Nottwil,
Switzerland

⁴ ICF Research Branch of WHO FIC CC
in German (DIMDI) at SPF Nottwil,
Switzerland

INTRODUCTION

Epidemiology of Functioning

Leon Gordis (1) characterises *epidemiology* as “the study of how disease is distributed in populations and of the factors that influence or determine this distribution”. In particular, he states five specific aims of epidemiology:

- 1) The determination of the extent of a disease found in the population.
- 2) The identification of the etiology or the cause of a disease and its risk factors.
- 3) The evaluation of existing and new therapeutic measures and modes of health care.
- 4) The study of the history and prognosis of disease.
- 5) The provision of a fundament for public policy and regulatory decisions relating to environmental problems.

The study of the genome allowed a shift from an analysis of the disease phenotype to the disease genotype. This area of research has nowadays become widely known as genetic epidemiology (2). In a similar fashion, the large interest in quality of life research reflects a shift from a pure biomedical perspective towards a more patient-oriented understanding of disease, symptoms, func-

tional and subjective outcomes (3). With the World Health Organization's International Classification of Functioning, Disability, and Health (ICF) (4), a framework for a structured study of functioning and disability has been established. In this chapter, we describe an *epidemiology of functioning* that aims to better understand a population's functional characteristics and options to cope with disability. We will use data from populations of patients experiencing a spinal cord injury (SCI) as a case-in-point. Suitable methods for the analysis will be thoroughly described in the following sections. Finally, applications will be shown and linked to the questions outlined above.

Human Functioning and Disability of People with Spinal Cord Injury

The complexity of human functioning and disability (3-5) in relation to various environments poses major challenges to scientific understanding and public health interventions. In SCI, functioning ranges from involuntary functions of the human organism to activity and participation in society (6-8). According to the ICF, differential functioning levels result from the interaction between health condition, personal and environmental factors (4). At the country level, economic resources and their distribution including investments in infra- and support structures set the scene for individual functioning (9-10). This may relate to inequalities in population levels of functioning between countries with differential economic resources and resource distribution. While this has been relatively well demonstrated for mortality (11-12) as well as subjective health of the general population (13-14), no evidence exists regarding functional outcome and the coping with disability.

The ICF Core Set Development Project for Spinal Cord Injury

The ICF Core Set development project for SCI (6-8, 15) was an effort to assess the burden of people with SCI and will provide an exemplary data basis throughout this chapter. 16 study centres in 14 countries collected data of 1048 adults with SCI due to an injury or disease with an acute onset. The sample was constructed to represent all WHO regions as well as different phases of SCI rehabilitation (6, 15). The functional problems of the study participants were assessed using a Case Record Form (CRF) comprising 264 ICF second-level categories. Information on the ICF components *body functions* and *body structures* was mostly based on the health professionals' assessment, information on the ICF component *activities and participation* was, depending on the ICF category, either reported by the professionals or the participants. The assessment of environmental factors was based solely on the patients' perception (15). We binary coded each category of *body functions*, *body structures*, and *activities and participation* with respect to the presence or absence of a problem. Each category of *environmental factors* was re-coded into two dummy variables, i.e., whether an environmental ICF category was

named a facilitator, a barrier or both (7). For a more in-depth discussion on (alternative) operationalisation strategies of the ICF and a discussion on rater agreement, see for example: (4, 16).

The study further examined the following injury-related and demographic variables: level of SCI (para- vs. tetraplegia), severity of SCI (complete vs. incomplete), time to injury in years, situation (early post-acute vs. chronic), years of formal education, gender, and age at interview. Macro-economic level information on the study countries' per capita gross domestic products (GDP) and Gini coefficients is available from World Bank (17). The former is a measure for the wealth of a country, whilst the latter describes its distribution among the people. For some analyses, we divided the population into higher- and lower-resource countries based on their per capita GDP (see figure 1).

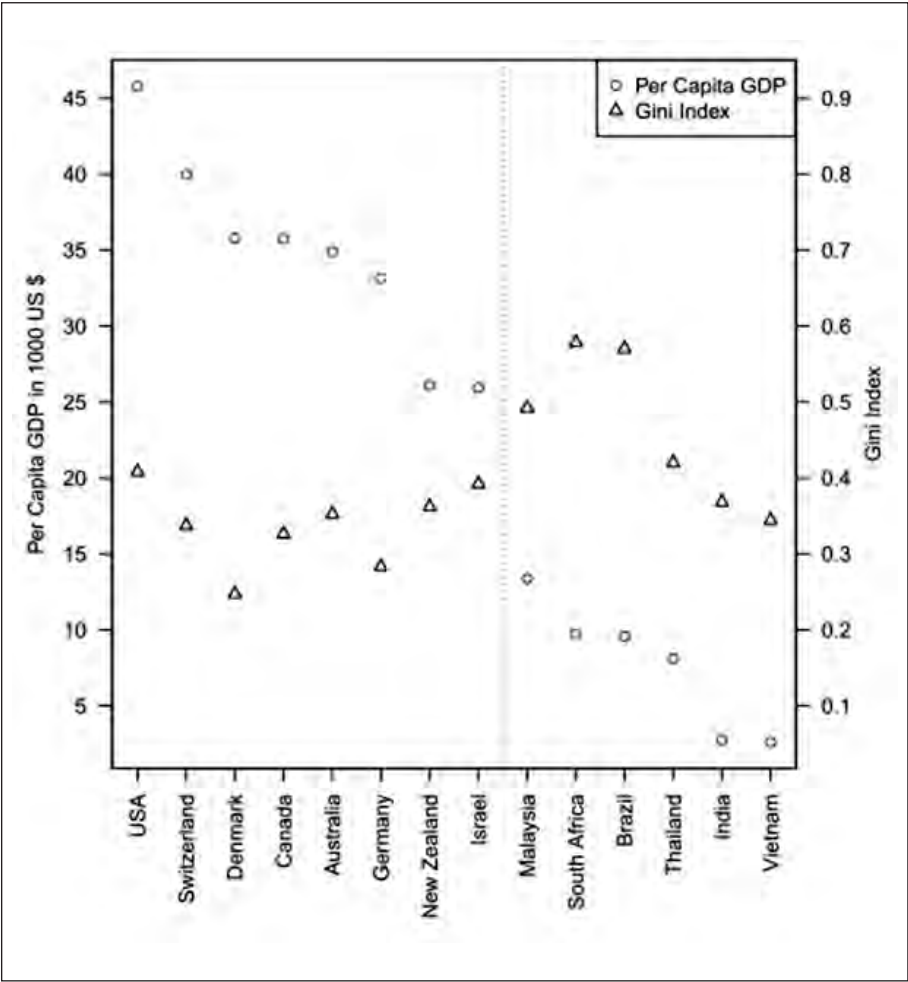


FIGURE 1. Per capita GDP and Gini-coefficients of the countries under study. The dashed line corresponds to the threshold between high (left) and low (right) resource countries.

FUNCTIONING PROFILES AND GRAPHICAL MODELS

In this section we present methods well known from genetic epidemiology and which we believe to be very useful in an epidemiology of functioning: Functioning profiles and graphical models. Functioning profiles are widely applied in the study of microarray data and pose a visual alternative to frequency tables. The term “graphical model” does refer to a whole array of statistical methods. Graphical models can be separated into undirected and directed graphical models. Whilst the former is useful to visualise the association structure among variables, the latter is concerned with the recovery of the directionality of effects among variables.

Functioning Profiles

Gene maps have their origins in systems biology where they are frequently applied to study whether “knocking out” a particular gene has a facilitating or inhibiting effect on the production of a particular substance (18). The same methodological concept, which we call henceforth functioning profile, can easily be applied to represent the distribution of functional problems in a study population. For each ICF category (apart the facilitator dummies from the environmental factors domain) a “1” denotes a problem whereas a “0” denotes no problem (vice versa for the facilitator dummies, where a “1” indicates this category means a facilitating effect on functioning). This allows counting the relative percentage of “1”s among the study population for each ICF category (“0”s for environmental facilitators). We can then assign a grey colour scale ranging from white (0-10% frequency) up to black (91-100% frequency). The same technique can also be used to describe the distribution among a certain sub population (e.g., only participants from high resource countries).

Undirected Graphical Models

An (undirected) graphical model can be used to visualize associations among the ICF categories via a network of nodes and edges (18). *Nodes* represent variables (the ICF categories). An *edge* between two nodes represents an association between these two ICF categories (whilst controlling for non-ICF variables such as gender or age). In general, we call a set of edges that connect two variables X and Y (possibly through various more nodes) a *path* from X to Y. If two nodes are only connected through a third (or even more) nodes, we say these two ICF categories are conditionally independent given any of the ICF categories (nodes) lying on this path. For example, if A and B are only linked via a path that goes through node C, any effect A might have on B is already contained in C. Thus A has no direct effect on B (and vice versa). One then says “A is conditionally independent of B given C” (see figure 2).

An algorithm to estimate the association structure of ICF categories has recently been outlined by Strobl et al. (19). The estimation of the association structure in the graphical model is based on the Least Absolute Shrinkage and Selection Operator (LASSO) (20). The LASSO can be thought of as an extension of the simple backward selection regression scheme. Unlike the backward selection regression, which can only handle, at most, as many variables as study participants, the LASSO does not have such

a restriction. For both regression approaches, the outcome is similar: Variables which are regarded as not important for the dependent variable have an estimated parameter equal to 0. Thus, if the reader feels uncomfortable with the deep mathematical roots of the LASSO regression, it is just as well possible to replace the LASSO by the backward selection regression in the algorithm by Strobl et al. We will thus refer to both LASSO regression and backward selection regression as “regression” henceforth.

It is also possible to control for additional variables such as gender or age which take the role of additional (non-ICF category) variables in the model. These additional variables vary in the sense that they are fix, i.e., the regression will not remove them from the model (thus ensuring that we always control for them). Also, since we consider their purpose to control the association structure for certain external effects, neither of them will be shown in the graph (although technically possible). Throughout this chapter, each regression was controlled for the patient’s gender, situation (post-acute vs. chronic), age, years of formal education, level of SCI and completeness of injury.

Following the algorithm by Strobl et al., the graph can then be constructed as follows. For each ICF category a regression is performed with all the other ICF categories as independent variables. We thus perform in total as many regressions as we have ICF categories. From these regressions we know for each ICF category the important predictors (ICF categories with non-zero parameter). Consequently, an undirected graph contains an edge between any two ICF categories if and only if the corresponding parameter in the regression was non-zero.

To enhance the stability of the identified edges, the regression approach can be combined with additional statistical methods. Here, we will describe the so-called bootstrap aggregation (or, in short, bagging) (21). This method is based on the idea that an analysis is repeated a certain number of times on different subsamples of the study participants and

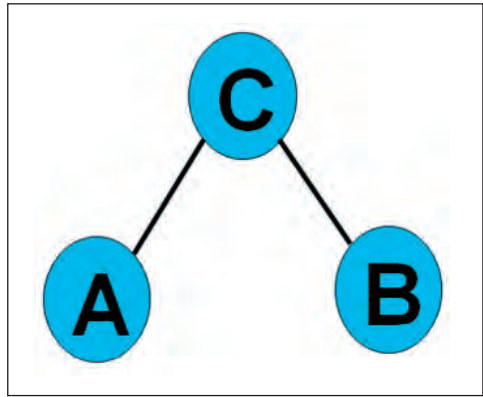


FIGURE 2. Nodes A and B are only associated with each other through node C. Thus, A does not have a direct effect on B and its effect is already contained in C (and vice versa).

then aggregated into one final solution. Repeating the analysis on different subsamples introduces additional variation. Aggregating the different solutions into one combined solution (for example via averaging or taking the most frequently appearing solution) thus corresponds to tackling the question “How would my results have changed if my sample had been slightly different?”. This approach is also referred to as sensitivity analysis. The generation of a suitable subset is straight forward: From a study population of n subjects we randomly sample n participants with replacement. Thus, in this subset a particular study participant might appear several times whilst another participant does not appear at all. We have thus generated an artificial sample that is slightly different from the original sample. For a more in-depth treatment of the bootstrap and bagging, we like to refer the reader to the work of Leo Breiman (21). All undirected graphs presented in this chapter have been aggregated over 500 bootstrap samples. Only edges which showed up in 100% of the samples were considered in the final aggregated graph. To improve the readability of the graphs, we only show nodes which have at least one edge to another ICF category.

The algorithm has been implemented based on the *penalized* package for LASSO-type regressions in the statistical software package R (22).

Directed Graphical Models

An example of a directed graphical model is the Directed Acyclic Graph (DAG) model. DAGs consist of nodes and arrows connecting the nodes. As a further restriction, the arrows must be directed in a way, so that it is not possible to trace a circle when following the arrowheads (otherwise a variable points towards itself which contradicts a causal framework).

The interpretation rule is d-separation (23), which is closely related to conditional independence for undirected graphs discussed above (see figure 2). However, the directionality of the arrowheads makes this case slightly more complicated: The arrowheads in a DAG between two variables X and Y can result in either active paths, i.e., paths from one variable to another that carries information, or inactive paths, i.e., paths which do not carry information (in the sense that it is blocked by some variable(s) on the path). A direct path from $X \rightarrow Y$ is always active. Paths with two or more edges, however, can behave differently. A path $X \rightarrow Z \rightarrow Y$ is active if we do not know Z , but inactive if we condition on Z , since in the latter case all information contained in X is also contained in Z . ICF-based example: s7502: structure of ankle and foot, d4500: walking short distances, d4600 moving around within the home. On the path $s7502 \rightarrow d4500 \rightarrow d4600$ the knowledge that a person has trouble walking short distances already incorporates the information about an injury in his ankle or foot with respect to the ability of moving around within the home.

The opposite is true for a path $X \rightarrow Z \leftarrow Y$. Here, the path is to be read that X does not carry any information about Y if Z is unknown. The

path from X to Y becomes active once we condition on Z as the additional knowledge now also tells us something about Y. ICF-based example: d440 fine hand use \rightarrow d345 writing messages \leftarrow d3601 using writing machines. Formally, we say X and Y are d-connected if there exists at least one active path between them. If no such path exists, we say X and Y are d-separated.

DAG models are particularly useful for estimating intervention effects due to the directed nature of the graph. Imagine that a causal system is represented by a DAG: Nodes represent (observable) variables (in our case ICF categories) and arrows represent direct causes. Now assume that we gather data from the causal system by observing it many times in different states and recording the values of all involved variables. The observed data will entail some dependence information among the variables. Since every DAG on these variables also entails dependence information via d-separation, we could find the DAG that fits the dependence information in the data best. It is a basic fact of DAG models, that we usually won't be able to identify a unique DAG that fits the data best. Rather, we will find several DAG models that fit all equally well. These DAG models are called "equivalent". The DAGs of the equivalent DAG models have a noteworthy property: When ignoring the arrowheads, they look the same. But if the arrowheads are considered, they might point into different directions, i.e., the direction of some edges is ambiguous. It was shown in (24), that under certain technical assumptions (which hold for the model discussed here) the unique arrows, i.e., arrows which are the same among all equivalent graphs, coincide with the true arrows in the underlying causal system. Thus, by estimating a DAG model, we can get information about the underlying causal structure. This information is contained in the unique arrows of the DAG model. However, the ambiguous arrows don't contain direct information on the underlying causal structure. Hence, estimating a DAG model from observational data gives insight into some aspects of the underlying causal structure, but other aspects will remain hidden. For this reason, it is in general only possible to estimate bounds (and not precise values) on causal effects from observational data.

For estimating DAG models from observational data, we used the PC-algorithm, which is explained in detail in the literature (25-27). The PC-algorithm makes a series of conditional independence tests to estimate the undirected fundament of the graph (i.e., when arrowheads are ignored). This undirected graph is only a by-product of the PC-algorithm, and not to be confused with the undirected graphs we have described above and which are generated using a regression approach. Subsequently, the PC-algorithm identifies the unique arrows to find the equivalence class of DAG models that fit the data equally well. The PC-algorithm is available in the package *pcalg* for R.

Several recent papers have tried to improve the performance of the PC-algorithm (see for example (28) and (29)). Each of these algorithms can be used as an alternative to the PC-algorithm. Further below we will discuss a strategy to derive theoretical intervention effects from such a DAG model.

APPLICATIONS

Now that we have gained a better understanding of graphical models and their estimation, we like to discuss their applicability linked to the aims of epidemiology as outlined in the beginning of the chapter.

Aim 1 - The Determination of the Extent of a Disease Found in the Population

Although graphical models are a very flexible approach, they focus more on the effect and/or interplay of variables than on the extent of a disease. Here, functioning profiles are a much more eligible tool. Figure 3 shows a

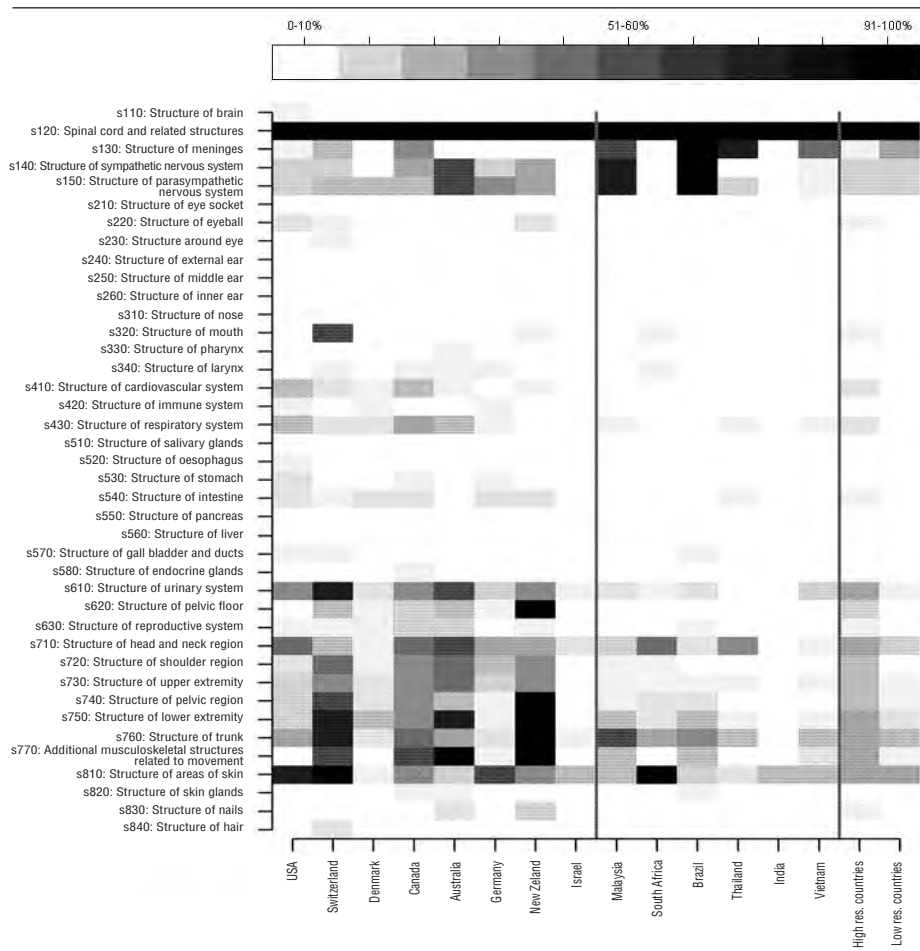


FIGURE 3. The figure shows the percentage of reported problems for various ICF categories from the component body structures. White indicates few/no study participants (0-10%) reported an impairment in this category, whereas black indicates that most/all study participants reported an impairment (91-100%). The first solid line corresponds to the threshold between high (left) and low (right) resource countries. The second solid line separates the country-specific frequencies from the frequencies for high and low resource countries.

functioning profile of all ICF categories from the body structures component for the SCI study population.

Not surprisingly, “s120: Spinal cord and related structures” was reported as a major impairment among all study participants, regardless of study country of the participant (black colour). Other structures, located above the position of a typical spinal cord injury, such as “s210: Structure of eye socket” or “s260: Structure of inner ear” are reported very rarely as an impairment (white colour). It is also interesting to note, that some differences show up between high- and low-resource countries. For example, “s410: Structure of cardiovascular system” is reported frequently in the high-resource countries but very rarely in the low-resource countries. It is tempting to speculate this might be related to more advanced medical equipment in high resource countries or better medical provision, but of course to confirm or reject such a hypothesis requires a much more in-depth study of the issue in question.

Aim 2 - The Identification of the Etiology or the Cause of a Disease and its Risk Factors

Two aspects of graphical models are particularly useful to examine this question: Dimension reduction and intervention calculus.

Dimension Reduction

A (directed or undirected) graphical model may produce several independent clusters of ICF categories. Within a cluster, all ICF categories can reach each other following a path along the edges, whereas variables which cannot be reached are part of another cluster. Thus, when studying a particular ICF category, only those ICF categories in the same cluster provide additional insights about that category. Clusters may thus be interpreted as independent dimensions of human functioning.

In figure 4 we labelled various dimensions of functioning as proposed by the graphical models approach. The edges describe associations that were present in both high- and low-resource countries. The resulting clusters are thus very small, but might still be useful as a starting point to identify the core variables associated with a specific ICF category. For example, “d540: Dressing” is part of a cluster we labelled “self care” which further consists of the ICF categories “d520: Caring for body parts”, “d510: Washing oneself” and “d530: Toileting”. Problems with dressing thus appear to be further connected with these three ICF categories regardless of a study participant’s country of heritage. It is up to the examiner to decide whether such a cluster might provide additional aspects to be incorporated in a more in-depth study or whether one is not interested in these additional aspects of functioning which, though associated, might not be directly related to the research question.

Also, as with all statistical methods, such a cluster is only valid if the estimated graph is correct. When studying the dimensionality of a certain problem, one also might want to study the stability of the graph estimate (e.g., based on bootstrap subsamples from the study population). Appar-

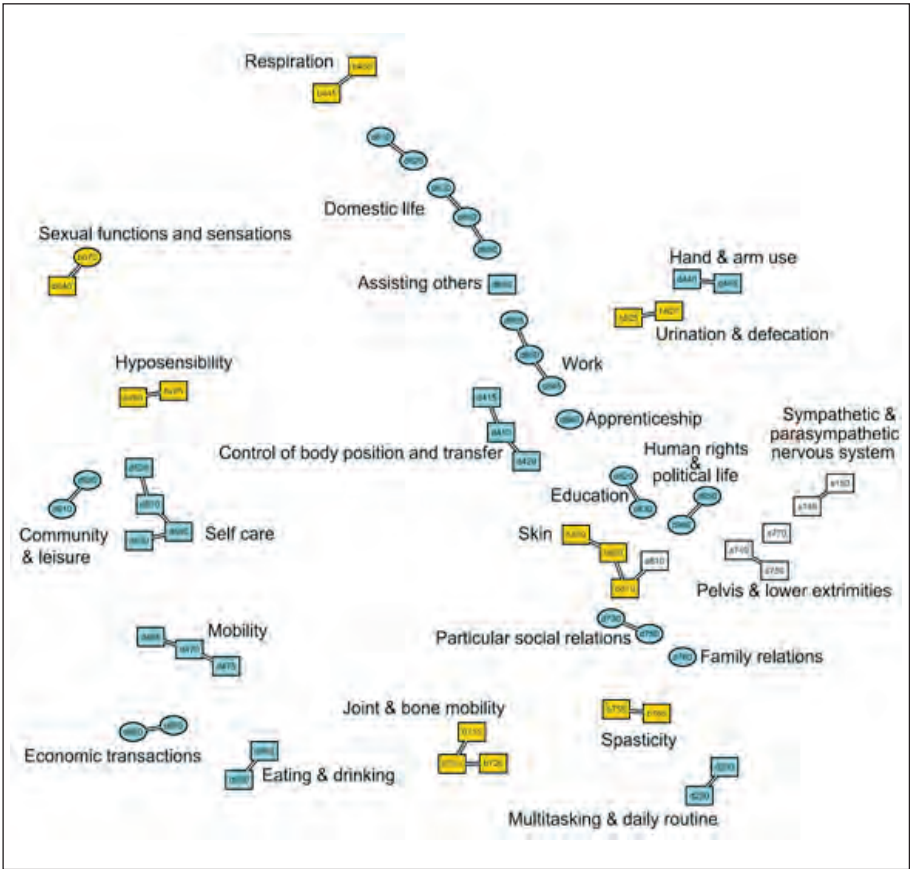


FIGURE 4. The figure shows associations among ICF categories, which remain stable across both high- and low-resource countries. Categories from the environmental factors domain are not shown.

ently, an unstable graph is untrustworthy. Thus its sensitivity to small changes in the study population should always be assessed.

Further note that the question of dimensionality is very often studied in the context of factorial analysis / principal components analysis. For a study of the overlap of the undirected graphical models approach with a principal components approach see, for example, Reinhardt and colleagues (30).

Intervention Modelling

Although it is a general postulate that causal effects can exclusively be estimated in experimental studies where an intervention is applied at random, in principle, and under certain assumptions (no unmeasured confounders, no measured selection variables and some more technical assumptions (24)) it is also possible to estimate bounds on intervention effects for Gaussian variables from observational data, even if the underlying causal structure is not known (31). Directed graphical models, and especially DAGs,

are a key element in this method. The result for Gaussian variables can be extended to binary variables thus providing a suitable approach for ICF data. Since, in general, intervention effects cannot be determined uniquely by observations alone (i.e., without an experimental design), one obtains a set of possible intervention effects which contain the true effect but might also contain wrong ones. This is a conceptual problem and occurs even if there are infinitely many observations. Since the estimated intervention effect for each ICF category need not be unique, we usually obtain several estimates of the intervention effect for each ICF category. We then aggregated the estimates for each ICF category using the mean over all estimates for one ICF category. We report on the ranking of the categories according to this measure of effect strength. We estimated the intervention effects using the R package *pcalg*.

Using the SCI data, we analyzed the question: “Which ICF category has the most positive effect on the dependent variable General Health Perception (ghp) if it is improved by external intervention (e.g., therapy)?” To this end, we estimated the intervention effect of each ICF category on ghp. More precisely, we studied both the patient assessment (figure 5a) and the health professional assessment (figure 5b) of ghp in relation to the various ICF categories. We can see that the two figures give quite a different picture. Only three categories (d540: Dressing, d410: Changing basic body position, b525: Defecation functions) show up in both perspectives. The health professionals seem to be more focused on the person’s ability to live a self-contained life (e.g., d570: Looking after one’s health, d560: Basic economic transactions). The patient’s perspective reflects a focus on more directly experienced factors (e.g., b280: Pain, s810: Skin (quite likely closely linked to bedsore), b550: Thermo-regulatory functions, b445 Respiratory

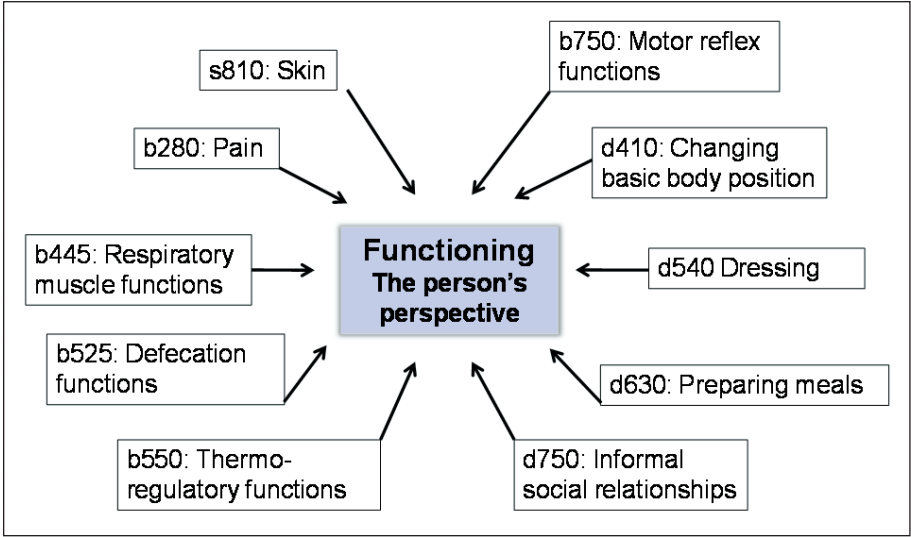


FIGURE 5a. The 10 most influential ICF categories with respect to the persons’ general health perception. This figure was estimated using the PC-algorithm.

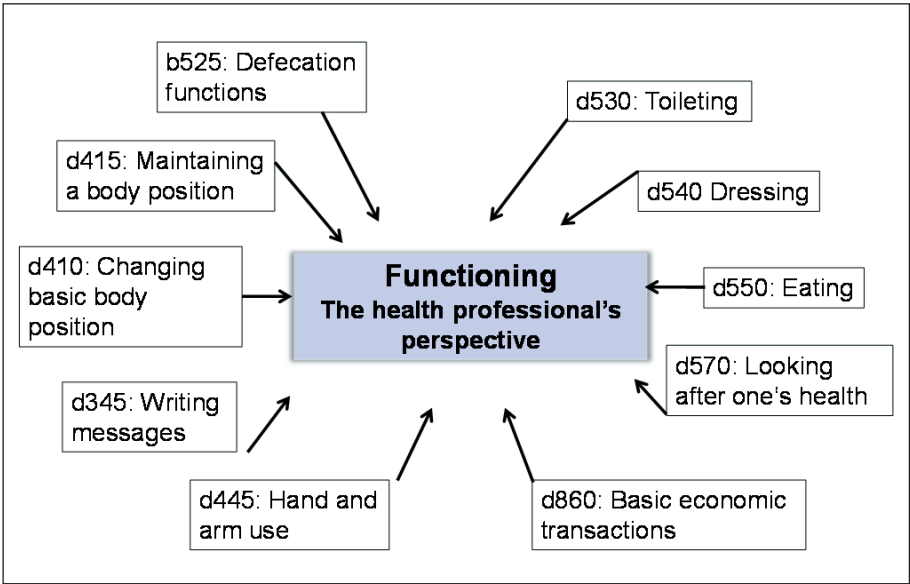


FIGURE 5b. The 10 most influential ICF categories with respect to the health professionals' belief about the person's general health perception. This figure was estimated using the PC-algorithm.

muscle functions etc.). But, even if the two sets of ICF categories seem to have some face validity, the approach cannot replace intervention studies. Intervention calculus should more be regarded as a tool to generate potential intervention targets from observational data for follow-up studies than an actual answer to the question of the best-possible interaction.

Aim 3 - The Evaluation of Existing and New Therapeutic Measures and Modes of Health Care

Comparison of Graphs

Graphs allow comparisons between two (or more) populations. The dissimilarity of two graphs can be quantified using the Structural Hamming Distance (SHD). The SHD between two graphs is defined as the number of edge insertions or deletions that are needed to transform one graph into the other. Using the bootstrap approach it is possible to derive a distribution of the SHD for each region as follows. Say, we generate 1000 bootstrap samples from an Asian population and another 1000 bootstrap samples from a European population. Then we estimate the corresponding 2000 graphs. We can then easily derive for both Asia and Europe the SHD between the graph estimated on the original data set and each of the 1000 respective bootstrap samples. The SHD ranges from 0 (when both graphs are identical) to the number of edges in the graph (the theoretical maximum). For the two distributions of the SHD from Asia and Europe one can now perform a simple two-sample t-test. See also Kalisch and colleagues (26) for a similar approach.

For the interested reader, we like to briefly outline a different approach, kin in nature to hierarchical multiple testing. Mansmann and colleagues (32) proposed an algorithm that follows a top-down testing strategy based on some kind of given hierarchy, in our case the ICF. The ultimate aim is to continue more-and-more specialised tests until the null-hypothesis “The two tested structures are equal” is rejected. First, a test on the global level is performed (one graph versus another). Then a test among the 4 ICF components (body functions in graph 1 vs. body functions in graph 2, body structures in graph 1 vs. body structures in graph 2, and so on). Next the algorithm performs tests on the ICF chapter level, then ICF subchapters. Thus, as a consequence of the hierarchy, this approach helps identifying differences in well-defined ICF structures. However, as tempting as testing for such a difference might appear, we would like to remind the reader of the complexity of both the underlying estimation and the test problem. The result is by no means “the final answer” and should always be critically perceived and be evaluated in follow-up studies.

Aim 4 - The Study of the History and Prognosis of Disease

Longitudinal graphical models are an extension of the static graphical models presented so far. They allow a visual display of changes in the association structure (undirected case) or causal structure (directed case) along the time axis. Due to the underlying time structure, this kind of graph is also known as chain graph. The principal idea is very common to the comparison of graphs for several populations (one for each point of time). However, one might want to incorporate additional restrictions, e.g., that estimated effect strength of an ICF category shall not vary drastically from one time point to another but rather in a smooth fashion. We refer to the literature for more information on this particular type of graphical model (see for example (23) and (33)).

Aim 5 - The Provision of a Fundament for Public Policy and Regulatory Decisions Relating to Environmental Problems

Health policy aims to identify modifiable intervention targets on the macro-, meso- and micro-level with major influence on desired outcomes and under the consideration of potential side-effects. Graphical models are hence a suitable method to discover influential variables (intervention calculus) and potential side effects in differential settings (comparison of graphs). Moreover, macro-environmental variables such as income distribution or health care expenditure can be modelled as nodes in a graph showing associated chains of functional categories and outcomes. We can, for instance, ask the question if and how the functioning of elder people with SCI may be enhanced when the income of a country would be distributed more equally. Since both intervention calculus and the comparison of graphs have been discussed earlier in this chapter, we relinquish a more in-depth (but conceptually similar) treatment in the health-policy context.

This concludes our overview of concepts eligible for an epidemiology of functioning. Though some of them have very mathematical roots, the visual element allows a rather intuitive interpretation, and we hope we could spark the interest of the reader. Of course, the methods described here are by no means complete, and might very well be combined with concepts outlined in other chapters of this book. After all, an epidemiology of functioning, or more general “epidemiology”, is a rapidly progressing field, and we ourselves look forward to thrilling developments in the future.

REFERENCES

1. Gordis L. Epidemiology. Philadelphia: Saunders Company; 2000.
2. Ziegler A, König I. A Statistical Approach to Genetic Epidemiology: Concepts and Applications, with an e-learning platform. Weinheim: Wiley-VCH 2010.
3. Wilson I, Cleary P. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *JAMA* 1995; 273: 59-65.
4. WHO. International Classification of Functioning, Disability and Health (ICF). Geneva: WHO Press; 2001.
5. Bickenbach J, Chatterji S, Badley E, Ustun T. Models of disablement, universalism and the international classification of impairments, disabilities and handicaps. *Soc Sci Med* 1999; 48: 1173-87.
6. Biering-Sørensen F, Scheuringer M, Baumberger M, Charlifue S, Post M, Montero F, et al. Developing core sets for persons with spinal cord injuries based on the International Classification of Functioning, Disability and Health as a way to specify functioning. *Spinal Cord* 2006; 44: 541-6.
7. Cieza A, Kirchberger I, Biering-Sørensen F, Baumberger M, Charlifue S, Post M, et al. ICF Core Sets for individuals with spinal cord injury in the long-term context. *Spinal Cord* 2010; 48: 305-12.
8. Kirchberger I, Cieza A, Biering-Sørensen F, Baumberger M, Charlifue S, Post M, et al. ICF Core Sets for individuals with spinal cord injury in the early post-acute context. *Spinal Cord* 2010; 48: 297-304.
9. Albrecht G, Bury M. The political economy of the disability market place. Albrecht G, Seelman K, Bury M, editors. Thousand Oaks (Ca): Sage; 2001.
10. Badley E. Enhancing the conceptual clarity of the activity and participation components of the International Classification of Functioning, Disability, and Health. *Soc Sci Med* 2008; 66: 2335-45.
11. Wilkinson R, Pickett K. Income inequality and socioeconomic gradients in mortality. *Am J Public Health* 2008; 98: 699-704.
12. Woolf S, Jones R, Johnson J, Phillips RJ, Oliver M, Bazemore A, et al. Avertable deaths associated with household income in Virginia. *Am J Public Health* 2010; 100: 750-5.
13. Pei X, Rodriguez E. Provincial income inequality and self-reported health status in China during 1991-7. *J Epidemiol Community Health* 2006; 60: 1065-9.
14. McGrail K, Doorslaer Ev, Ross N, Sanmartin C. Income-related health inequalities in Canada and the United States: a decomposition analysis. *Am J Public Health* 2009; 99: 1856-63.
15. Kirchberger I, Biering-Sørensen F, Charlifue S, Baumberger M, Campbell R, Kovindha A, et al. Identification of the most common problems in functioning of individuals with spinal cord injury using the International Classification of Functioning, Disability and Health. *Spinal Cord* 2010; 48: 221-9.
16. Grill E, Mansmann U, Cieza A, Stucki G. Assessing observer agreement when describing and classifying functioning with the International Classification of Functioning, Disability and Health. *J Rehabil Med* 2007; 39: 71-6.

17. The World Bank. World development indicators 2008. Washington DC: World Bank; 2008.
18. Friedman N. Inferring cellular networks using probabilistic graphical models. *Science* 2004; 303: 799-805.
19. Strobl R, Stucki G, Grill E, Müller M, Mansmann U. Graphical models illustrated complex associations between variables describing human functioning. *J Clin Epidemiol* 2009; 62: 922-33.
20. Meinshausen N, Bühlmann P. High-dimensional graphs and variable selection with the LASSO. *Ann Stat* 2006; 34: 1436-62.
21. Breiman L. Bagging predictors. *Mach Learn* 1996; 24: 123-240.
22. R Development Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2010.
23. Lauritzen S. Graphical models. Oxford: Oxford University Press; 1996.
24. Pearl J. Causality. Cambridge: Cambridge University Press; 2000.
25. Kalisch M, Bühlmann P. Estimating high-dimensional directed acyclic graphs with the PC-algorithm. *J Machine Learning Research* 2007; 8: 613-36.
26. Kalisch M, Fellinghauer B, Grill E, Maathuis M, Mansmann U, Bühlmann P, et al. Understanding human functioning using graphical models. *BMC Med Res Methodol* 2010; 11: 10-4.
27. Spirtes P, Glymour C, Scheines R. Causation, Prediction, and Search. Cambridge: The MIT Press; 2000.
28. Tsamardinos I, Brown L, Aliferis C. The max-min hill-climbing Bayesian network structure learning algorithm. *Machine Learning* 2006; 65(1): 31-78.
29. Ramsey J, Zhang J, Spirtes P. Adjacency-faithfulness and conservative causal inference. In *Proceedings of 22nd Conference on Uncertainty in Artificial Intelligence*. Oregon: AUAI Press, pp. 401-408.
30. Reinhardt JD, Fellinghauer BAG, Strobl R, Strucki G. Dimension reduction in human functioning and disability outcomes research: graphical models versus principal components analysis. *Disabil Rehabil* 2010; 32: 1000-10.
31. Maathuis M, Kalisch M, Bühlmann P. Estimating high-dimensional intervention effects from observational data. *Ann Stat* 2009; 37: 3133-64.
32. Mansmann U, Schmidberger M, Strobl R, Jurinovic V. Indirect comparison of interaction graphs. In: Kneib T, Tutz G, editors. *Statistical Modelling and Regression Structures*. Heidelberg: Physica-Verlag; 2010, 249-65.
33. Wermuth N, Lauritzen S. On substantive research hypotheses, conditional independence graphs and graphical chain models. *J R Stat Soc. Series B* 1990; 52: 21-72.

CARL V. GRANGER¹, MARSHA CARLIN¹
HARRY FELICIANO², SAM MARKELLO¹
LUIGI TESIO³

FUNCTIONAL ASSESSMENT AND OUTCOME MEASUREMENT: MODELS AND PERSPECTIVES ACROSS THE REHABILITATION CONTINUUM

¹ University at Buffalo, The State
University of New York, Buffalo,
New York, USA; Uniform Data
System for Medical Rehabilitation,
Amherst, New York, USA

² Palmetto GBA, LLC, Columbia,
South Carolina, USA

³ Università degli Studi and the
Department of Neurorehabilitation
Sciences, Istituto Auxologico
Italiano, IRCCS, Ospedale
San Luca, Milan, Italy

INTRODUCTION

Tesio (1) stated that in medicine *function* has two meanings. First, there is the exchange of energy or information from body part to body part. For example, the heart exchanges energy with the blood, but the information from the exchange is scant: the low frequency modulation of on-off beats. The second meaning, more relevant to physical medicine and rehabilitation, is the exchange of energy or information between the whole person and the environment, inclusive of other persons. This is *nature* in the sense of the Greek etymology φυσικς, from which *physical medicine* takes its name (2). Rising from a chair requires a considerable power output, yet the information content from the action is modest: sitting or standing. In order to facilitate the understanding of functional assessment methods, a useful distinction must be made between human biology and human behavior. Contrary to the function of organs, cells, and molecules, the whole person must be seen as an indivisible unit. While nerve fibers can make muscle

fibers contract, *I* am the actor in an exchange at the person-environment level (3). Functional measurement—documenting how much function there is—must be based on behaviors and perceptions (4).

An overview is given of the applications of the functional assessment philosophy to clinical measurement and decisions in rehabilitation medicine, including assessing individual cases, diagnostic groups, rehabilitation hospitals and lesser levels of care, the continuum of care, and national health care programs. Examples are taken from the US experience with particular reference to the most widely used instruments and systems developed decades ago in the US by the Uniform Data System for Medical Rehabilitation and the University at Buffalo, The State University of New York. In the US, classification and functional assessment research led to the nationwide implementation of classification and payment systems in acute-care hospitals, rehabilitation, and long-term care. Some of these same instruments and systems are used in numerous countries around the world. The discussion of functional assessment would not be complete without addressing the International Classification of Disease (5) and the International Classification of Functioning, Disability, and Health (6). These are described in terms of their roles in the larger national health care delivery systems.

WHAT IS FUNCTIONAL ASSESSMENT?

The restorative patient typically is in an intensive program of medical rehabilitation because of loss of functional ability subsequent to illness or injury, is near medical stability, and has potential for significant practical recovery of abilities. In the United States, rehabilitation hospitals, which are freestanding inpatient hospitals or in-hospital units, offer intensive rehabilitation treatment, while less intensive care is dispensed in such facilities as skilled nursing, sub-acute, outpatient, and adult day; and in the home. Rehabilitation patients receive physical, occupational, speech, vocational, psychological, and recreational therapies. Functional assessment documents and objectively measures levels of functional abilities in performing motor and cognitive activities of daily living and participation in relevant psychosocial activities. Clinicians in all rehabilitation venues conduct functional assessments at appropriate intervals to set and re-set therapy goals, change care venues, determine types and amounts of assistance needed from devices and helpers, and recommend environmental adaptations.

The most useful functional assessment tools are scientifically tested over time for their reliability, validity, responsiveness to change, feasibility for use, and meaningfulness in the clinical setting. Tools encompass several domains and are used with various disabilities and in numerous care settings. Table 1 contains select assessment tools used in medical rehabilitation, organized by functional domains (7). The medical rehabilitation field primarily uses tools that measure activities

TABLE 1. Select Rehabilitation Assessment Tools

Functional Domains	Tools
ADLs	Barthel Index (8) FIM® Instrument (9) Katz Index (10) LIFEware SM System (11)
Ambulation/ Locomotion	Dynamic Gait Index (DGI) (12) Functional Ambulation Profile (FAP) (13) Gait Abnormality Rating Scale (GARS) (14) Physical Performance Battery (15) Six Minute Walk (16) Timed Up & Go (17) Walking Speed (18)
Balance	Berg Balance Scale (19) Balance Self Perceptions Test (20) Functional Reach Test (21)
Cognitive Functioning	Mini-Mental State Exam (MMSE) (22)
Depression	Beck Depression Inventory (BDI) (23) Center for Epidemiologic Studies Depression Scale (CES-D) (24)
Executive Functioning	Stroop Test (25) Trails A & B Tests (26)
IADLs	Everyday Problems Test (EPT) (27) Lawton Index (28) LIFEware SM System (11) Pfeffer Index (29)
Memory	Wechsler Memory Scale (30)
Pain	McGill Pain Questionnaire (31) Visual Analog Scale (32)
Well-Being/ HRQOL	36-Item Short Form Health Survey (SF-36®) (33) Sickness Impact Profile (SIP) (34)

Note - ADLs: activities of daily living; IADLs: instrumental activities of daily living; HRQOL: health-related quality of life.

of daily living (ADL), such as eating, grooming, mobility, transfers, and some cognitive tasks. Less frequently used are tools for instrumental activities of daily living (IADL), such as cooking, shopping, and house-keeping; and health-related quality of life (HRQOL) tools, for general health and general functionality. The tools in Table 1 have been translated into multiple languages and have demonstrated validity and reliability. Most tools require administrators (in some cases, trained on the instrument) to observe and record patient performance of activities; while other tools are self-assessments of patient-perceived abilities in performing activities in their own environments. The scales of measurement (nominal, ordinal, interval, or ratio level data) are not the same across tools. A clinician must choose tools that match the

purpose of the measurement or intervention with the functional domains of interest.

THE FIM® INSTRUMENT: HISTORY, METRIC PROPERTIES, APPLICATIONS

Rehabilitation patients used to be classified and treated according to diagnostic disease categories; yet it was functionality more often than disease that affected patients' abilities to live as normally as possible in their own environmental and social circumstances. Some patients enter inpatient rehabilitation after being stabilized from such events as stroke and orthopedic surgery, while others utilize several levels of inpatient and outpatient rehabilitative care for many years due to chronic disability resulting from such illnesses as multiple sclerosis and cardiac dysfunction. Locally developed functional assessment instruments have been in use since rehabilitation medicine emerged, but were not optimal because they lacked scientific testing and standardization; had motor but not cognitive items; and rating scales were insufficient, using vague terms like mild, moderate, and severe. A US task force—representing physiatry, psychology, physical therapy, occupational therapy, rehabilitation nursing, rehabilitation counseling, and other disciplines—developed guidelines over 20 years ago for development of assessment tools (35). A critical guideline was that instruments must be scientifically tested for validity and reliability. Content validity, predictive validity, concurrent validity, and construct validity are the complex validity tests that instruments must pass. Reliability tests, which minimize error across test administrators, time, and settings, include inter-rater reliability and test-retest reliability (36). The FIM® instrument (9) was created to fulfill a long-standing need for standardized functional assessment, and has become the most widely used instrument for adults in US inpatient rehabilitation hospitals, and also is used internationally. Patients are assessed at admission, at discharge, during treatment, and at follow-up. The FIM® instrument (9) is administered in a half hour or less by trained clinicians observing the patient performing activities of daily living. It has 18 items—13 motor and 5 cognition—and a rating scale of 1-7, representing levels of dependence to independence. Item ratings are totaled and range from 18-126, representing worse to better function. The motor items are: eating; grooming; bathing; dressing upper body; dressing lower body; toileting; bladder management; bowel management; bed/chair/wheelchair transfer; toilet transfer; tub/shower transfer; walk/wheelchair; and stair climbing. The cognition items are: comprehension, expression, social interaction, problem solving, and memory. The 7 rating levels are well-defined and categorized into subgroups of complete independence, modified dependence, and complete dependence. Summed item ratings are used to select treatments and care venues; for payment; for research; and for estimating patient bur-

den of care, in hours and minutes per day, in performing activities of daily living in the home, once the patient is discharged from the rehabilitation hospital (37). Knowing this burden of care is important, as families or hired caregivers must provide care in the home, often for many years. In functional assessment, ratings indicate less or more of a trait. Because raw scores have unknown spaces between them, Rasch analysis (38) is used to build estimates of true intervals of item difficulty and subject ability, creating a hierarchy of easy to difficult items. Figure 1 shows this hierarchy of items for the FIM® instrument (9). The FIM® instrument (9) has been adapted for pediatric use as the WeeFIM® instrument (39, 40); and shortened as the AlphaFIM® instrument (41), which assesses adult functional status in the acute-care hospital prior to inpatient rehabilitation.

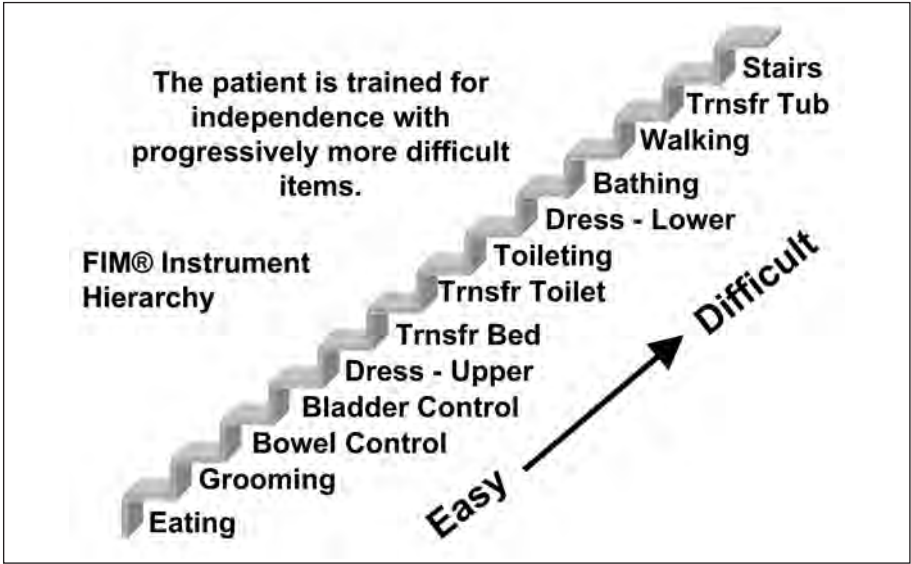


FIGURE 1. FIM® Instrument Hierarchy of Items.

FUNCTIONAL ASSESSMENT AS A BASIS FOR DECISION-MAKING IN REHABILITATION: FUNCTION BASED CLASSIFICATION SYSTEMS

Technical Issues and CART/Neural Networking: Automatic versus Human-Driven Interactive Methods

Functional assessment is key to classifying rehabilitation patients into subcategories for predictive purposes, using Classification and Regression Trees (CART) (42) and Artificial Neural Network (ANN) (43) methodologies.

CART may be viewed as a non-parametric algorithm designed to produce a decision-tree using binary recursive partitioning (42). The technique involves searching for a set of rules leading to a prediction of

dependent variables from known values of a set of predictor variables. For each patient in the primary dataset there are values for a set of one or more potential predictor variables, ranging from: functional motor and cognitive assessment; medical status; environmental factors; socio-demographics; resource utilization; treatment factors; treatment setting; and regional/geographical factors. Clinical plausibility, known relationships from previous published works, and caution against highly correlated variables (44) can serve as a guide to inclusion of appropriate predictors into the model application. Analysis begins with a population of subjects termed the root node. An example may be rehabilitation patients who have completed a course of therapy in a given type of setting, i.e., outpatient or acute inpatient rehabilitation. A root node is split into two groups, based on the strongest independent predictor of the outcome under investigation. Dichotomization continues automatically at each branch point, based on the strongest predictor, until such time that the predictive power of the model is saturated, or until a pre-defined number of branches has been achieved. To evaluate the final model it is important to determine the misclassification risk or proportion of cases classified incorrectly. A risk estimate of 0.025 would be considered low, indicating that only 2.5% of the cases are misclassified in the model.

In the US, Stineman (45) and colleagues developed for inpatient rehabilitation a number of function-based classification systems called FIM-FRG (function-related groups) modules, as a way to match patient complexity with payment. The classifying variables were: Rehabilitation Impairment Categories (RICs) (46); admission motor-FIM[®] instrument (9) rating; admission cognition-FIM[®] instrument (9) rating; and age. Modules served multiple purposes: predicting resource use (length of stay and costs); outcomes (gain from admission to discharge in motor- and cognition-FIM[®] instrument (9) ratings); and discharge to home and other destinations. Carter et al. (47) expanded on the FIM-FRG (45) module for resource use, employing CART (42), and went on to develop a Case Mix Group or CMG-based prospective payment system for Medicare patients in inpatient rehabilitation facilities in the US, which assigns patients into one of 92 intensive rehabilitation categories (48)¹. Prospective payment in the US is reimbursement to hospitals for Medicare patients using a pre-determined per-case payment rate. Key predictors of resource use—function at admission using the FIM[®] instrument (9); medical condition (rehabilitation impairment group, etiologic diagnosis, and co-morbidities); and patient demographics (specifically age)—are captured by the Inpatient Rehabilitation Facility-Patient Assessment Instrument (IRF-PAI) (46), which includes the FIM[®] instrument (9).

¹ Medicare in the US is a federally funded health insurance program for persons 65 years and older and persons with disabilities, and is part of the US Centers for Medicare & Medicaid Services.

Classification systems designed for payment purposes, such as those described above, were built using classification analysis (recursive partitioning) rather than multiple linear regression. Stineman et al. (49) posit that classification analysis is better suited to capture higher-order interactions among clinical variables, and yields clinically similar patient groups. Koss and Feinstein (50) suggest that the hierarchical nature of classification groups more logically delineates signs and symptoms, which in turn is better aligned with clinician thinking. Similarly, Temkin et al. (51) indicate that CART (42) trees are simple enough to be used in a clinical setting and offer predictions accurate enough for clinical utility. In contrast, multiple linear regression is more restrictive, attempting to impose linear relationships on clinical data where relationships may be non-linear.

CART (42) algorithms also have served to create expected outcomes for rehabilitation case groups, against which actual outcomes can be compared, thus allowing clinicians to monitor individual treatment progress and adjust modalities as needed (52). These authors created six distinct rehabilitation efficiency groups, for example, for stroke patients, expressed as FIM[®] instrument (9) point gain per day, calculated from admission functional assessment data: admission-FIM[®] instrument (9) motor rating; admission-FIM[®] instrument (9) cognition rating; and onset time (duration in days since the stroke event). Data were derived from stroke patients treated in inpatient rehabilitation. Similarly, the US Centers for Medicare & Medicaid Services publishes an annual update of expected lengths of stay (in days) by CMG (53) and co-morbidity subgroup (tiers, explained in next section) for Medicare fee-for-services patients treated in inpatient rehabilitation (54). Hospitals use these data as expectations to gauge their actual resource utilization performance.

In contrast to CART (42)—which models a linear relationship among a set of independent predictor variables and a single dependent variable and produces defined subgroups as output—an artificial neural network (ANN) (43) models complex, non-linear processes automatically, producing multiple outputs (44). One of the first applications of neural network methodology in inpatient rehabilitation was published by Grigsby et al. (55) and focused on predicting functional outcome, length of stay, and cost. Ohno-Machado et al. (56) describe a number of early rehabilitation studies and provide brief summaries concerning neural networks, limitations of conventional predictive models, and comparison with statistical regression models. There are numerous recent rehabilitation studies for predictive modeling or for comparing results of ANN (43) with CART (42) and/or multiple regression (57-64).

ANN (43) may be good for producing accurate and optimum predictive modeling results and require fewer assumptions to be met, but because of their need for relatively large computation resources, lack of transparency, and inability to query and quantify the contribution of individual predictors, tend to be difficult to interpret and less acceptable by clinicians (65). Tu (66) discusses the advantages and disadvantages of ANN (43) relative to other predictive models. As with any computational tool,

analytical goals must first be clearly defined, and the most appropriate tool selected that meets those needs, given the nature of the scientific question and complexity/nature of the inter-relationships between predictors and outcome(s). If the primary goal is to achieve reliable prediction, and the relationship between independent and dependent variables is non-linear, plus computational resources is not an issue, then preference might be given to ANN (43). If, however, the intent is to gain insight into a problem and easily quantify the causal contribution from the various components, logistic regression or multiple-linear regression may be preferred since the researcher has the ability to manipulate or control variables entered into the model and can test for interaction or effect modification (67).

FUNCTIONAL ASSESSMENT AS A BASIS FOR DECISION-MAKING IN REHABILITATION: FUNCTION-BASED CLASSIFICATION SYSTEMS

Function and Diagnosis

The patient classification system for US inpatient rehabilitation is largely function-based, however, it should be noted that the first level of classification is formed by the Rehabilitation Impairment Category (RIC), which describes the impairment or principal diagnosis for which the patient requires rehabilitation (45, 68). Clinical diagnosis also was eventually found to play a role in establishing medical complexity. In the early function-related group (FRG) work conducted by Stineman, (45) ICD-9-CM (69) diagnoses did not substantially improve the FRGs capacity to predict length of stay. However, Carter et al. (47) found a positive effect of co-morbidities or complications on cost-based FRGs, meaning medical complexity was shown to have an impact on cost for inpatient rehabilitation patients. This cost predictor was carried over into the current functional CMG-based payment system for US inpatient rehabilitation hospitals. CMGs are divided into four tiers, each with a relative weight and associated payment (54), based on specific co-morbid conditions using ICD-9-CM codes (69, 70).

The system used in the US since 2008 for Medicare prospective payment for acute inpatient rehabilitation hospitalization is based on Diagnostic-Related Groups (DRGs), refined and renamed Medicare Severity DRGs (MS-DRGs), which are based on a classification system using ICD-9-CM codes (69, 71, 72). This system better recognizes severity of illness and resource use based on case complexity. Originally, DRGs were developed not for payment purposes but to facilitate the needs of hospital management by providing a system to measure and evaluate hospital performance (73). The system was designed within a recursive partitioning computer program called AUTOGRP (74), which aids patient care management with rapid analyses and hypothesis testing. The dependent variable was length of stay; and classifying variables were primary and secondary diagnoses, surgical procedures, age, gender, and discharge status. A total of 475 DRGs resulted. Following nearly a decade of research,

DRGs became the basis for a Medicare acute-care hospital-based prospective payment system (75).

In the US, skilled nursing facilities provide a lesser level of rehabilitation care than inpatient rehabilitation hospitals, and also follow a prospective payment system for Medicare recipients. Payments are adjusted for case-mix using a classification system that accounts for the relative resource utilization of different patient types. Resource Utilization Groups (RUG-III) is the classification system for skilled nursing facilities. Using recursive partitioning, (76) 44 RUG-III categories were developed. Wage-weighted facility staff time was the dependent variable, and classifying variables were: a) special rehabilitation; b) extensive services; c) special care; d) clinically complex; e) impaired cognition; f) behavior problems; g) reduced physical functioning; h) affect (depression); and i) activities of daily living. Currently, the first seven independent variables serve as the first test of whether a skilled nursing facility resident qualifies for payment under RUG-III. Residents who qualify then are split into mutually exclusive groups, such as “Very High Rehabilitation A” and “Extensive Services 3” (77). Grouping assignment is based on assessment data collected for each resident using the Minimum Data Sets 2.0 (MDS 2.0) (78). This model explains 56% of the variance in total per-diem costs. White et al. (79) reported a much lower contribution to explaining variance when using total costs as the dependent variable versus staff-time costs, and also lower rates when focusing strictly on Medicare stays and non-therapy ancillary costs versus all stays and only staff-time costs.

Failure to appropriately compensate for non-therapy ancillary costs has been the focus of complaints from providers. CMS has taken steps to compensate for these failures, such as introducing RUG refinements in 2006, recalibration of case-mix indices for fiscal year 2010, establishing a revised case-mix methodology (RUG-IV) for implementation in fiscal year 2011, and a transition from MDS 2.0 to MDS 3.0 (80).

FUNCTIONAL ASSESSMENT FOR PREDICTION ACROSS POPULATIONS: CONTRASTING OBSERVED AND EXPECTED OUTCOMES

The FIM® Instrument and Managed Care in the US

For observed rehabilitation outcomes to generate useful information about the delivery of quality care there must be a comparative reference point depicting some agreed-upon, unbiased, predictable, minimum standard or expected level of the desired outcome. Such outcomes expectations are challenging to develop because rarely, if ever, is there a simple linear cause-effect relationship between a particular healthcare intervention and the desired outcome.

Within the model, factors such as baseline status, clinical status, and demographic/psychosocial characteristics have been termed *risk factors*, while treatment and setting are termed *treatment characteristics*.

To be able to truly determine if actual observed outcomes associated with treatment meet or exceed expectations, or meet or exceed those of a referent population, effects of these non-treatment-related risk factors must be controlled. This process is termed *risk adjustment*. Risk adjustment can be achieved by various methods, such as applying patient-specific factors to a previously developed and validated regression model to generate an expected outcome or a probability of achieving that outcome. Another approach involves subjecting key patient outcome predictors into a validated CART (42) regression model to classify a given patient on the basis of the outcome of interest. An example is FIM[®] instrument (9) point gain per day for stroke patients admitted to acute inpatient rehabilitation programs (52). Patients with varying functional status at admission, as determined by admission motor- and cognition-FIM[®] instrument (9) ratings, as well as the duration (in days) prior to rehabilitation admission, could be classified into one of six categories of an expected rate of functional improvement. Thus, when subjecting stroke rehabilitation patients to the model, the unique attributes of each patient will have been taken into account, and they will be classified accordingly. In effect, this is the same tact used in classifying patients into the various cost-based CMGs (predictors including impairment group, weighted admission motor-FIM[®] instrument [9] rating, admission cognition-FIM[®] instrument rating [9], and age) for the prospective payment system of acute inpatient rehabilitation in the United States (47). On average, patients classified into the stroke CMG 0104 category in Rehabilitation Hospital ABC would be expected to have similar average outcomes as those classified CMG 0104 in Rehabilitation Hospital XYZ, given that predictive factors in each case have been accounted for by the model.

Because each hospital treats a unique population of rehabilitation cases, with different proportions of cases among all the CMGs, a way of comparing Rehabilitation Hospital ABC's performance to Rehabilitation Hospital XYZ's performance would be to use a stratified comparison at the CMG level (i.e., comparing outcomes CMG by CMG). While possible, it is not always practical given that many CMG categories may have few cases. An alternative, with a large representative dataset containing all the types of rehabilitation cases treated in the US—such as the database kept by the Uniform Data System for Medical Rehabilitation (UDSMR), University at Buffalo, The State University of New York—is to create an outcomes expectation for each rehabilitation hospital based on the experience of the US as a whole². Comparing hospital performance requires use of an indirect standardization method, commonly used in epidemiologic population comparison studies (81). The method controls for the unique case-mix and

² The UDSMR is a not-for-profit program that began in 1987, and has the largest non-governmental database with outcomes on patients admitted to acute inpatient rehabilitation. The UDSMR provides outcomes services to over 850 out of a national total of nearly 1200 inpatient rehabilitation facilities in the United States.

severity of patients treated in each hospital. By applying national-level CMG-specific outcomes (external rates) to the hospital-level CMG case proportions, and summing the product terms (hospital CMG-specific proportion multiplied by national CMG-specific rate), an expectation X can be produced for each outcome of interest specific to that hospital. The way to compare actual and expected outcomes produced in this manner is to say: If US rehabilitation hospitals combined had Rehabilitation Hospital ABC's cases and its unique case-mix, the expectation of functional improvement would be X . If Rehabilitation Hospital ABC's functional improvement score for its program is below the expected value, this should spark interest in delving into the potential causes—ruling out internal data/measurement errors, followed by searching for unique uncontrolled factors next, and then searching for possible performance issues.

FUNCTIONAL ASSESSMENT FOR QUALITY CHECK OF CARE PROVIDERS

The Program Evaluation Model in the US

The Program Evaluation Model (PEM) in the US was created in 2006 by the UDSMR for US acute rehabilitation hospitals that subscribe to its rehabilitation outcomes services. The PEM was intended to serve three purposes: a) measure rehabilitation program quality and share results with individual subscribers to facilitate and motivate quality improvement in their overall program; b) identify consistently high performing IRFs to learn about and facilitate sharing their system processes that give rise to superior outcomes; and c) create a model that could serve the needs of payers, such as Medicare, for ranking and compensating rehabilitation hospitals in the near future using value-based purchasing, a mechanism commonly referred to as *pay-for-performance*.

To fulfill the first purpose required a decision on the definition of quality. This was easily met using the Institute of Medicine's six criteria for quality of healthcare delivered in the 21st century, namely, that care is effective, efficient, timely, safe, equitable, and patient-centered (82). Five outcome measures pertinent to acute rehabilitation healthcare were selected from the Medicare Inpatient Rehabilitation Facility-Patient Assessment Instrument (IRF-PAI) (46) that met these criteria. The IRF-PAI (46) is an assessment tool used in US rehabilitation hospitals (called Inpatient Rehabilitation Facilities or IRFs). The IRF-PAI (46) has functional, medical, and socio-demographic elements, and also contains the 18-item FIM® instrument (9) that every US rehabilitation hospital must complete on every rehabilitation patient covered by Medicare in order to submit data as part of the prospective payment system. UDSMR subscribers complete the IRF-PAI (46) on every patient, irrespective of payer source, and share their data with UDSMR for aggregation and performance comparison with national, regional, and state benchmarks.

Among the five outcome measures chosen for the PEM, three are case-level, and evaluated on each patient. These measures are: a) discharge-FIM[®] instrument (9) rating; b) FIM[®] instrument (9) gain, which is the difference between admission and discharge FIM[®] instrument (9) ratings; and c) length-of-stay efficiency, which is the FIM[®] instrument (9) gain divided by length of stay, in days. The other two measures are program-level: d) percent discharge to the community, a goal of rehabilitation; and e) percent discharge to an acute-care hospital (in the United States, a hospital for severe illness), which indicates significant functional decline. Discharge FIM[®] instrument (9) rating, FIM[®] instrument (9) gain, and community discharge rate are indicative of effectiveness, with the latter also indicative of patient-centeredness. Length-of-stay efficiency is indicative of efficiency and timeliness; while the acute-care hospital discharge rate is viewed as a component of safety. The PEM is case-mix and severity adjusted to account for and permit the inclusion of all cases treated and discharged by a rehabilitation hospital, with the exception of expired cases. The model is designed to account for varying proportions of cases from all rehabilitation impairment categories (RICs) (45, 68) and all levels of severity (case-mix groups, or CMGs) (54) within each RIC. On an aggregate basis, every rehabilitation hospital has a unique actual and expected performance score for each of the five outcome measures in the model. Annually, a single composite PEM score is calculated, representing a compilation of weighted performance ratios (actual, to a case-mix and severity adjusted expectation). Rehabilitation hospital-specific PEM scores are ranked and assigned a percentile. Each rehabilitation hospital then is given an annual, single-page score card containing not only PEM score and percentile ranking, but sufficient performance data for each component outcome measure to identify areas of strength and opportunities for improvement. The percentile provides each rehabilitation hospital with a sense of where it ranks. Rehabilitation hospitals at the 90th percentile and above are considered high performers. Those that have consistently ranked at or above the 90th percentile from 2006-2009 number 21 out of more than 800, and are worthy of study of system processes, thus meeting the second purpose of the PEM.

Regarding the third purpose, the PEM was fashioned after a similar model co-sponsored by the US Centers for Medicare & Medicaid Services, and designed to evaluate performance with built-in financial incentives among a group of acute-care hospitals (83, 84). This model showed that *pay-for-performance* has the potential to increase clinical quality and save lives. More than 100 pay-for-performance and incentive programs have been implemented in recent years in the private sector, with mixed results (85). As a result of US healthcare legislation of year 2010, Medicare will establish in hospitals in year 2012 value-based purchasing programs, and will include skilled nursing facilities, home health agencies, and ambulatory surgical centers (86). While rehabilitation hospitals are not mentioned specifically, they could be considered hospitals

in this policy. The PEM stands ready as a working tool for consideration and does not require collection of any new data since it was built on measures currently captured by the IRF-PAI (46) as part of the existing prospective payment system for US rehabilitation hospitals.

FUNCTIONAL ASSESSMENT ALONG THE CARE CONTINUUM: THE LIFEwareSM SYSTEM PARADIGM

The ideal of a continuum of care is following patients over time and rehabilitation settings to measure long-term outcomes. The FIM[®] instrument (9) for adult inpatients and the LIFEwareSM System (11) for adult outpatients (both part of the UDSMR) are joined in a continuum of care using Rasch analysis, (38) allowing both instruments to measure on a scale of 0-100, with higher ratings meaning better motor and cognitive function, less pain, and less psychological distress. The LIFEwareSM System (11) assesses functional status, through self-reporting, of persons with musculoskeletal, neurological, cardiac, and other conditions. There are 130 measures and scales, assessing the domains of physical functioning, affective sense of well-being or mood state, pain experience, satisfaction with life in general, and community role/participation. As a rule of thumb, ratings at 70 and above appear to represent the threshold of clinical significance, or the equivalent of a passing grade. Subscribing facilities customize forms to meet each patient's needs. For example, cardiac patients would be rated for fatigue, breathing, and chest discomfort; adult day care patients for memory, medication, and nutrition. Longitudinal records show: dates of assessments; measures with items beneath; and totals (Absolute) of item ratings, with minuses indicating item values below expected, based on thousands of cases in the database. Items rated at and above expected do not appear; while items that are below expected (for each item) do appear, signaling clinical attention is required. This functional information is used with medical information to create a comprehensive picture of health status. Assessment dates are on the far left; the measures run from left to right across the top, and the items appear beneath their measures.

The LIFEwareSM System longitudinal record, Figure 2, is an example of a stroke case. Selected for this patient were FIMALL, which is FIM[®] instrument (9) items adapted to the LIFEwareSM System (11); and the following measures and scales from the outpatient LIFEwareSM System (11): BMC (Body Movement and Control); LIMITATIONS (Physical Limitations); PAINSCALE (pain measure and the LIFEware[®] Visual Analog Scale [LVAS]; PLACID (emotional distress); SATISFACT, satisfaction with life in general; and COMMUN/SATISFACT, satisfaction with community participation and roles.

The patient is a 59-year-old woman, a homemaker residing with her spouse, in the United States. Her medical history includes diabetes controlled with oral medication, hypertension, and she is moderately over-

DATE	FIMALL	BMC	LIMITATIONS	PAIN-SCALE	PLACID	SATISFACT	COMMUN/ SATISFACT
07/06/05	33 Absolute -50 Transfers bed -48 Transfers toilet -44 Walk/ wheelchair -42 Transfers tub -35 Stairs -26 Bowel -24 Bladder -23 Dress up -21 Toileting -19 Dress low -19 Memory -9 Social interaction -8 Bathing -8 Problem solving -7 Grooming -1 Eating -1 Expression	Incomplete	Incomplete	Incomplete	Incomplete	Incomplete	Incomplete
08/15/05	53 Absolute -42 Transfers tub -10 Comprehension -9 Dress low -8 Bathing -7 Stairs -4 Bladder -3 Dress up -2 Transfers bed -1 Eating -1 Toileting	Incomplete	Incomplete	Incomplete	Incomplete	Incomplete	Incomplete
1/25/06	Incomplete	Incomplete -58 Kneeling -35 Getting up -35 Stairs -10 Reaching	62 Absolute -74 Left up limb -68 Fatigue -20 Bowel -16 Bladder -16 Left low limb	100 Absolute None below expected	23 Absolute -83 Blame -83 Morbid -68 Pessimistic -60 Irritated -58 Uptight -41 Panic	60 Absolute -2 Life satisfaction	Participation 80% Social Primary Role 80% Work Secondary Role 40% Home- making
4/10/06	Incomplete	Incomplete -45 Reaching -35 Getting up -35 Stairs -28 Standing	70 Absolute -74 Left up limb -68 Fatigue -16 Left low limb	100 Absolute None below expected	59 Absolute -60 Irritated -58 Uptight -21 Panic -15 Blame -15 Morbid	60 Absolute -2 Life satisfaction	Participation 80% Social Primary Role 80% Work Secondary Role 80% Home activities
9/18/07	76 Absolute -11 Eating	60 Absolute -44 Traveling -35 Stairs -35 Getting up -23 Lifting -20 Walking -10 Reaching	28 Absolute -88 Bowel -77 Vision -74 Left up limb -69 Right low limb -68 Fatigue -31 Bladder -16 Left low limb	100 Absolute None below expected	72 Absolute -60 Irritated -58 Uptight	60 Absolute -2 Life satisfaction	Participation 50% Social Primary Role 70% Work
10/08/08	90 Absolute None below expected	64 Absolute -40 Walking -35 Getting up -35 Stairs -10 Reaching -3 Lifting	46 Absolute -84 Bladder -77 Vision -74 Left up limb -35 Bowel -16 Left low limb -15 Fatigue	75 Absolute None below expected	72 Absolute -60 Irritated -58 Uptight	60 Absolute -2 Life satisfaction	Participation 60% Social Primary Role 50% Home- Making
Note - Absolute numbers in bold are below expected. Copyright 2010 - Uniform Data System for Medical Rehabilitation, a division of UB Foundation Activities, Inc., All Rights Reserved							

FIGURE 2. LIFEwareSM System (11) Longitudinal Record. Case Example.

weight. On 5/28/2005, she was admitted to the acute-care hospital for a stroke that caused severe left hemiparesis. Her vital signs remained stable; but she developed aspiration pneumonia followed by a urinary tract infection, both of which responded satisfactorily to antibiotics. She was transferred to the rehabilitation hospital on 7/6/2005. Her admission-FIM[®] instrument (9) rating, motor and cognition combined, was 51³. Applying Rasch analysis (38), her rating was 30. She received coordinated medical care, along with physical, occupational, and speech therapies, including training in swallowing. Over her 40-day rehabilitation hospital stay, she gained 29 FIM[®] instrument (9) points, bringing her discharge-FIM[®] instrument (9) rating to 80: an average improvement of 1.4 points per day. The gain was 21 Rasch (38) points, or an average improvement of 1.9 points per day. She was discharged to outpatient rehabilitation care on 8/15/2005, and followed periodically for more than 3 years. In that time, she recovered to 90 Rasch points. While her mood rating on the LIFEwareSM System (11) PLACID MEASURE increased from 23 to 72 Rasch points, she continued with significant limitations in bowel and bladder control and use of upper and lower limbs; while diminished vision and fatigue continued. Though rated for pain, no pain items were below expected values, therefore no pain items appear on the record. Clinicians would use this record and also refer to the patient's medical record to help determine why items appear, disappear, or change in values above or below expected.

USE OF ICF-BASED INFORMATION IN SUPPORT OF REASONABLE AND NECESSARY REHABILITATION SERVICES THROUGHOUT THE CONTINUUM OF CARE

Rehabilitation services are delivered in diverse settings and by interdisciplinary teams of healthcare professionals including physicians, nurses, physical therapists, occupational therapists, speech language pathologists, and respiratory care professionals. Healthcare systems look to the healthcare professionals' documentation to determine whether the rehabilitative services meet their coverage and payment standards. This may involve a review of the medical records to ensure that the services were reasonable and necessary with regard to amount, type, frequency, and duration. Despite the promise of electronic health records, healthcare systems have much work ahead of them in demonstrating their value in clinical practice.

Health status is a multi-dimensional variable that impacts the delivery of healthcare services and healthcare research. The relationship between the International Classification of Disease (ICD-10) (5) and healthcare outcomes offers an additional layer of specificity made possible by the International Classification of Functioning, Disability, and Health

³ The instrument rating range is 18-126, and as a frame of reference, a common rating at inpatient rehabilitation hospital admission is 60.

(ICF) (6). The taxonomy of the ICF (6), if applied in a manner supportive of healthcare professionals' tasks, has the potential to provide the framework for describing functional status from the acute-care hospital through to home health and even hospice and palliative care. The ICF (6), was created to supplement the ICD (5) with information on functioning and disability, and allows clinicians and researchers to address potential confounding variables through a comprehensive description of health status. The capacity for ICF-based (6) information to support patient-centered decision-making is helping to ensure that the right care is delivered to the right individual at the right time (87).

The economic and public health impact of chronic health conditions are being studied on a global level (88). The prevention, treatment, and palliation of these chronic health conditions and the management of their interactions with comorbid and related secondary conditions require a new approach to data collection at the point of service. As healthcare systems throughout the world prepare for the anticipated growth of the subpopulation 65 years of age and older, the utility of ICF (6) will become more evident, especially in the management of individuals with multiple, chronic conditions in need of rehabilitative services.

ACKNOWLEDGMENT

Luigi Tesio was supported by:

- Regione Lombardia (Lombardy Regional Government), Ricerca Finalizzata 2009
- Italian Ministry of Health, Ricerca Finalizzata 2009

REFERENCES

1. Tesio L. Functional assessment in rehabilitative medicine: principles and methods. *Eura Medicophys* 2007; 43: 515-23.
2. Tesio L, Franchignoni F. Don't touch the physical in "physical and rehabilitation medicine." *J Rehabil Med* 2007; 39(8): 662-3.
3. Tesio L. Bridging the gap between biology and clinical medicine. Some help from Rasch measurement theory. *J Appl Meas* 2004; 5(4): 362-6.
4. Tesio L. Measuring behaviours and perceptions: Rasch analysis as a tool for rehabilitation research. *J Rehabil Med* 2003; 35(3): 105-15.
5. World Health Organization. International classification of diseases, 10th revision. Geneva: World Health Organization; 2007.
6. World Health Organization. International classification of functioning, disability, and health. Geneva: World Health Organization; 2001.
7. Granger CV, Brownschidle CM, Carlin M, Graham JE, Malik C, Markello S, Niewczyk P, Ottenbacher K, Tesio, L. Functional assessment. In: Stone JH, Blouoin M, eds. International encyclopedia of rehabilitation. Available at: <http://cirrie.buffalo.edu/encyclopedia/article.php?id=44&language=en> Accessed March 17, 2010.
8. Mahoney FI, Barthel DW. Functional evaluation: the Barthel Index. *Maryland State Med J* 1965; 14: 63-5.
9. Uniform Data System for Medical Rehabilitation. Guide for the Uniform Data Set for Medical Rehabilitation (including the FIMTM instrument), Version 5.1, Buffalo, New York: Uniform Data System for Medical Rehabilitation; 1997.

10. Katz S, Ford AB, Moskowitz RW, Jackson BA, Jaffe MW. Studies of illness in the aged: the index of ADL, a standardized measure of biological and psychological function. *JAMA*. 1963; 185: 914-19.
11. Baker JG, Granger CV, Fiedler RC. A brief outpatient functional assessment measure: validity using Rasch measures. *Am J Phys Med Rehabil* 1997; 76: 8-13.
12. Jonsdottir J, Cattaneo D. Reliability and validity of the Dynamic Gait Index in persons with chronic stroke. *Arch Phys Med Rehabil* 2007; 88(11): 1410-15.
13. Nelson AJ. Functional ambulation profile. *Phys Ther*. 1974; 54(10): 1059-65.
14. Wolfson L, Whipple R, Amerman P, Tobin JN. Gait assessment in the elderly: a gait abnormality rating scale and its relation to falls. *J Gerontol* 1990; 45(1): M12-19.
15. Guralnik JM, Simonsick EM, Ferrucci L, Glynn RJ, Berkman LF, Blazer DG, Scherr PA, Wallace RB. A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. *J Gerontol* 1994; 49(2): M85-94.
16. Butland RJ, Pang J, Gross ER, Woodcock AA, Geddes DM. Two-, six- and twelve- minute walking test in respiratory disease. *Br Med J (Clinical Research ed.)* 1982; 284(6329): 1607-8.
17. Podsiadlo D, Richardson S. The timed "Up & Go": a test of basic functional mobility for frail elderly persons. *J Am Geriatr Soc* 1991; 39(2): 142-8.
18. Graham JE, Ostir GV, Kuo YF, Fisher SR, Ottenbacher KJ. Relationship between test methodology and mean velocity in timed walk tests: a review. *Arch Phys Med Rehabil* 2008; 89(5): 865-72.
19. Berg K, Wood-Dauphinee S, Williams JI, Gayton D. Measuring balance in the elderly: preliminary development of an instrument. *Physiother Canada* 1989; 41(6): 304-11.
20. Shumway-Cook A, Baldwin M, Polissar NL, and Gruber W. Predicting the probability for falls in community-dwelling older adults. *Phys Ther* 1997; 77(8): 812-19.
21. Duncan PW, Weiner DK, Chandler J, Studenski S. Functional reach: a new clinical measure of balance. *J Gerontol* 1990; 45(6): M192-M197.
22. Folstein MF, Folstein SE, McHugh PR. The Mini-Mental State. A practical method of grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975; 12(3): 189-98.
23. Beck AT, Steer RA, Garbin MG. Psychometric properties of the Beck Depression Inventory: twenty-five years of evaluation. *Clin Psych Rev* 1988; 8(1): 77-100.
24. Radloff LS. The CES-D Scale: A self-report depression scale for research in the general population. *Appl Psych Meas*. 1977; 1(3): 385-401.
25. Stroop JR. Studies of interference in serial verbal reactions. *J Exp Psych*. 1935; 18: 643-62.
26. Reitan RM. The relation of the Trail Making Test to organic brain damage. *J Consult Psych*. 1955; 19(5): 393-4.
27. Willis SL, Jay GM, Diehl M, Marsiske M. Longitudinal change and prediction of everyday task competence in the elderly. *Res Ageing* 1992; 14(1): 68-91.
28. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. *Gerontologist*. 1969; 9(3): 179-86.
29. Pfeffer RI, Kurosaki TT, Harrah Jr. CH, Chance JM, Filos S. Measurement of functional activities in older adults in the community. *J Gerontol* 1982; 37(3): 323-9.
30. Tulsky DS, Ledbetter MF. Updating to the WAIS-III and WMS-III: considerations for research and clinical practice. *Psych Assess* 2000; 12(3): 253-62.
31. Melzack R. The McGill Pain Questionnaire: major properties and scoring methods. *Pain*. 1975; 1(3): 275-99.
32. Revill SI, Robinson JO, Rosen M, Hogg MI. The reliability of a linear analogue for evaluating pain. *Anaesthesia*. 1976; 31(9): 1191-8.
33. Ware JR, Sherbourne CD. The MOS 36-Item Short-Form Health Survey (SF-36). I. Conceptual Framework and Item Selection. *Med Care* 1992; 30(6): 473-83.
34. Bergner M, Bobbitt RA, Carter WB, Wilson BS. The Sickness Impact Profile: development and final revision of a health status measure. *Med Care* 1981; 19(8): 787-05.
35. Johnston MV, Keith RA, Hinderer S. Measurement standards for interdisciplinary medical rehabilitation. *Arch Phys Med Rehabil* 1992; 73(suppl): S3-S23.

36. Granger CV, Black T, Braun SL. Quality and outcome measures for medical rehabilitation. In: Braddom RL, ed. *Physical Medicine & Rehabilitation*. 3rd ed. Elsevier; 2006.
37. Heinemann AW, Kirk P, Hastie BA, Semik P, Hamilton BB, Linacre JM, Wright BD, Granger C. Relationships between disability measures and nursing effort during medical rehabilitation for patients with traumatic brain and spinal cord injury. *Arch Phys Med Rehabil* 1997; 78(2):143-9.
38. Linacre JM, Wright BD. *A user's guide to Winsteps*. Chicago: MESA Press; 2000.
39. Uniform Data System for Medical Rehabilitation. WeeFIM System® Clinical Guide, Version 5.01. Buffalo, New York: Uniform Data System for Medical Rehabilitation; 1998, 2002.
40. Uniform Data System for Medical Rehabilitation. WeeFIM® II System Clinical Guide. Buffalo, New York: Uniform Data System for Medical Rehabilitation; 2004.
41. Uniform Data System for Medical Rehabilitation. The AlphaFIM® Instrument Guide, Version 2.1., Buffalo, New York: Uniform Data System for Medical Rehabilitation; 2005.
42. Breiman L, Friedman JH, Olshen RA, Stone CJ. *Classification and Regression Trees*. Belmont, Calif.: Wadsworth; 1984.
43. Kröse B, van der Smagt P. *An Introduction to Neural Networks*. 8th ed. The University of Amsterdam; 1996.
44. Segal ME, Philip H, Goldstein R, et al. The accuracy of artificial neural networks in predicting long-term outcome after traumatic brain injury. *J Head Trauma Rehabil* 2006; 21(4): 298-314.
45. Stineman MG. Function-related groups 101: a primer. *Crit Rev Phys Rehabil Med* 1998; 10: 319-58.
46. US Department of Health and Human Services, Centers for Medicare & Medicaid Services, Baltimore, Maryland. Inpatient Rehabilitation Facility-Patient Assessment Instrument (IRF-PAI). OMB-0938-0842 2002.
47. Carter G, Relles D, Buchanan J, et al. A Classification System for Inpatient Medical Rehabilitation Patients: A Review and Proposed Revisions to the Functional Independence Measure-Function Related Groups [PM-682-HCFA]. Santa Monica, Calif.: RAND/UCLA/Harvard Center for Health Care Financing Policy Research; 1997.
48. MedPAC. Rehabilitation facilities (inpatient) payment system. Available at: http://www.medpac.gov/documents/MedPAC_Payment_Basics_08_IRF.pdf Accessed April 27, 2010.
49. Stineman MG, Kallen MA, Thompson C, Gage B. Challenges in paying for effective stays. *Med Care* 2005; 43(9): 841-3.
50. Koss N, Feinstein AR. Computer-aided prognosis: development of a prognostic algorithm. *Arch Intern Med*. 1971; 127: 448-59.
51. Temkin NR, Holubkov R, Machamer JE, Winn HR, Dikmen SS. Classification and regression trees (CRT) for prediction of function at 1 year following head trauma. *J Neurosurg* 1995; 82(5): 764-71.
52. Granger CV, Ottenbacher K. An analysis of the rate of functional gain that may be expected during inpatient rehabilitation for patients with stroke. *Proceedings of the 2nd World Congress, International Society of Physical and Rehabilitation Medicine*; May 18-22, 2003. Prague, Czech Republic: ISPRM; 37-40.
53. Stineman M. Case mix measurement in medical rehabilitation. *Arch Phys Med Rehabil* 1995; (76)12: 1163-70.
54. US Department of Health and Human Services. Centers for Medicare & Medicaid Services. Medicare Program; Inpatient Rehabilitation Facility Prospective Payment System for Federal Fiscal Year 2010; Final Rule (CMS-1538-F). Update to the case-mix group (CMG) relative weights and average length of stay values for FY 2010. *Federal Register* 74(151) of August 7, 2009 (42 CFR Part 412); 39765-39773.
55. Grigsby J, Kooker R, Hershberger J. Simulated neural networks to predict outcomes, costs, and length of stay among orthopedic rehabilitation patients. *Arch Phys Med Rehabil* 1994; 75: 1077-81.
56. Ohno-Machado L, Rowland T. Neural network applications in physical medicine and rehabilitation. *Am J Phys Med Rehabil* 1999; 78: 392-8.

57. Lang E, Pitts L, Damron S, Rutledge R. Outcome after severe head injury: an analysis of prediction based upon comparison of neural network versus logistic regression analysis. *Neurol Res* 1997; 19: 274-80.
58. Oczowski WJ, Barreca S. Neural network modeling accurately predicts the functional outcome of stroke survivors with moderate disabilities. *Arch Phys Med Rehabil* 1997; 78: 340-5.
59. Walters DKW, Linn RT, Kulas M, Cuddihy E, Wu C, Granger CV. Selecting modeling techniques for outcome prediction: comparison of artificial neural networks, classification and regression tress, and linear regression analysis for predicting medical rehabilitation outcomes. *Proceedings of the American Medical Informatics Association Symposium*, 1999: 1187.
60. El-Sohl A, Saltzman S, Ramadan F, Naughton B. Validity of an artificial neural network in predicting discharge destination from a postacute geriatric rehabilitation unit. *Arch Phys Med Rehabil* 2000; 81: 1388-93.
61. Ottenbacher KJ, Smith PM, Illig SB, Linn RT, Fiedler RC, Granger CV. Comparison of logistic regression and neural networks to predict rehospitalization in patients with stroke. *J Clin Epidemiol* 2001; 54: 1159-65.
62. Ottenbacher KJ, Linn RT, Smith PM, Illig SB, Mancuso M, Granger CV. Comparison of logistic regression and neural network analysis applied to predicting living setting after hip fracture. *Ann Epidemiol* 2004; 14: 551-9.
63. Eftekhari B, Mohammad K, Ardebili H, Ghodsi M, Ketabchi E. Comparison of artificial neural network and logistic regression models for prediction of mortality in head trauma based on initial clinical data. *BMC Med Inform Decis Mak* 2005; 5(3): 1-8.
64. Kiani K, Snijders C, Gelsema E. Recognition of daily motor activity classes using an artificial neural network. *Arch Phys Med Rehabil* 2009; 79(2): 147-54.
65. Hart A, Wyatt J. Evaluating black-boxes as medical decision aids: issues arising from a study of neural networks. *Med inform* 1990; 15: 229-36.
66. Tu JV. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *J Clin Epidemiol* 1996; 49: 1225-31.
67. Greenberg RS, Kleinbaum DG. Mathematical modeling strategies for the analysis of epidemiologic research. *Ann Rev Public Health* 1995; 6: 223-45.
68. US Centers for Medicare and Medicaid Services 2009. Available at: <http://www.cms.hhs.gov/InpatientRehabFacPPS/LSIRF/list.asp#TopofPage>. Page 3. Transmittal 938. Related links inside CMS. Transmittal 938. R938CP [PDF, 187KB]. Accessed April 27, 2010.
69. World Health Organization. International Classification of Diseases, 9th Clinical Modification. Geneva: World Health Organization; 1977.
70. US Centers for Medicare and Medicaid Services 2009. Available at: http://www.cms.hhs.gov/InpatientRehabFacPPS/07_datafiles.asp#TopOfPage. FY 2009. Data files. 508. Compliant version of appendix C list of comorbidities.pdf. Updated August 7, 2009. Accessed April 27, 2010.
71. US Department of Health and Human Services. Centers for Medicare & Medicaid Services 2007. Medicare Program; Proposed changes to the Hospital Inpatient Prospective Payment Systems and fiscal year 2008 rates; final rule (CMS-1533-P), Federal Register 72(162) of May 3, 2007 (72 FR 24680).
72. Wynn BO, Scott M. Evaluation of severity adjusted DRG systems. Addendum to the interim report. RAND Corporation 2007; WR434/1-CMS.
73. Fetter RB, Brand DA, Gamache D. DRGs: Their design and development. Ann Arbor, Mich: Health Administration Press; 1991.
74. Fetter RB, Shin Y, Freeman JL, Averill RF, Thompson JD. Case-mix definition by diagnosis related groups. *Med Care* 1980; 18(Suppl): 39.
75. Mayes R. The origins, development, and passage of Medicare's revolutionary prospective payment system. *J Hist Med Allied Serv* 2006; 62(1): 21-55.
76. Fries BE, Schneider DP, Foley WJ, Gavazzi M, Burke R, Cornelius E. Refining a case-mix measure for nursing homes: resource utilization groups (RUG-III). *Med Care* 1994; 23: 668-685.
77. Federal Register: Medicare Program; Prospective Payment System and consolidated billing for skilled nursing facilities; Final rule. 63FR(91): 26098-26147 (May 12, 1998).

78. US Centers for Medicare & Medicaid Services. Minimum Data Sets 2.0. Available at: <http://www2.cms.gov/MinimumDataSets20/> Accessed May 3, 2010.
79. White, C. Pizer, SD, White, AJ. Assessing the RUG-III Resident Classification System for skilled nursing facilities. *Health Care Financing Rev* 2002; 24(2): 7-15.
80. US Centers for Medicare & Medicaid Services: Office of Public Affairs. CMS announces more accurate payment rates for Medicare skilled nursing facilities in fiscal year 2010: case-mix adjustment recalibrated. Friday, July 31, 2009. Available at: <http://www.cms.gov/apps/media/press/release.asp> Accessed May 3, 2010.
81. Selvin, S. *Statistical Analysis of Epidemiologic Data*. 2nd ed. Oxford University Press, New York, NY: 1996
82. Institute of Medicine. *Crossing the quality chasm: a new health system for the 21st century*. Washington, DC: National Academy Press; 2001.
83. Premier Inc. Centers for Medicare & Medicaid Services (CMS)/Premier Hospital quality incentive demonstration (HQID) project: project overview and findings from year one. Charlotte, NC: Premier Inc.; 2006.
84. Premier Inc. Centers for Medicare & Medicaid Services (CMS)/Premier Hospital quality incentive demonstration (HQID) project: findings from year two. Charlotte, NC; 2007.
85. Institute of Medicine (IOM). *Rewarding Provider Performance: Aligning Incentives in Medicare*. The National Academies Press, Washington DC; 2007.
86. The Henry J. Kaiser Family Foundation. Focus on health reform. Available at: http://www.kff.org/healthreform/upload/healthreform_tri_full.pdf Accessed May 4, 2010.
87. Palmetto GBA Going Beyond Diagnosis®: valuable information (Revised February 2008). Available at: [http://www.palmettogba.com/palmetto/providers.nsf/docsCat/Regional%20Home%20Health%20Hospice%20Intermediary%20\(RHHI\)~Learning%20Education~Going%20Beyond%20Diagnosis%20Series](http://www.palmettogba.com/palmetto/providers.nsf/docsCat/Regional%20Home%20Health%20Hospice%20Intermediary%20(RHHI)~Learning%20Education~Going%20Beyond%20Diagnosis%20Series) Accessed April 18, 2010.
88. Global Initiative for Chronic Obstructive Lung Disease. Global strategy for the diagnosis, management, and prevention of COPD. Updated December 2009. Available at: <http://www.goldcopd.com/> Accessed May 3, 2010.

ATILLA H. ELHAN¹
AYŞE A. KÜÇÜKDEVECİ²
ALAN TENNANT³

THE RASCH MEASUREMENT MODEL

¹ University of Ankara, Faculty
of Medicine, Department
of Biostatistics, Turkey

² University of Ankara, Faculty
of Medicine, Department
of Physical Medicine and
Rehabilitation, Turkey

³ University of Leeds, Department
of Rehabilitation Medicine,
Faculty of Medicine and Health,
United Kingdom

INTRODUCTION

Assessment and outcome measurement is an essential part of medical practice. The term “assessment” includes techniques and procedures for classification and measurement of a variable pertaining to a person (1). Measurement is quantification of an observation by a standard unit. An assessment becomes a potential “outcome measure” when it is associated with the result of an intervention of some kind. In other words, “outcome” is defined as change in a state or situation that arises as a result of some process or intervention (2). A wide range of assessments are used in physical and rehabilitation medicine across a variety of clinical or community settings. These assessments might be at the levels of body functions or structures, activities and participation, or quality of life and might differ from a laboratory test to a specific joint examination or to a questionnaire evaluating activities of daily living. They can be performed for various reasons such as clinical decision-making in individual patients, clinical audit, policy-making or research purposes and can be undertaken by a wide range of professionals and, in some circumstances may be self completed by patients.

The type of information obtained from assessments differs in important ways, such that they can be categorical, ordinal, interval or ratio. Many assessment tools used in every day clinical practice are ordinal scales which are based upon a score derived from a set of tasks or questions. In this scale type, the numbers assigned to tasks or questions are representative of rank ordering. Consequently ordinal scales order people by the magnitude of the construct under consideration, for example their level of dependency as in the Barthel Index (3), or their physical functioning as in the Health Assessment Questionnaire (4), or their level of pain as in the McGill Pain Questionnaire (5). A key characteristic of these scales is that the distances between the raw score points are unequal and mathematical calculations such as change scores are invalid (6-8). Generally different sets of statistical procedures (non-parametric) are available for use with ordinal scales (8). In contrast an interval scale has equal interval units which can support mathematical operations such as the calculation of a change score, or an effect size. As most interventional studies in rehabilitation and health in general wish to calculate a change score, or use values from ordinal scales in procedures such as ANOVA and regression, then the challenge is to provide a transformation of such scale to the interval level. Recently Rasch analysis has gained widespread use in this context, opening up a wider range of available statistical procedures (9-10).

WHAT IS RASCH ANALYSIS?

Rasch analysis is the formal testing of an assessment, or an outcome measure against a mathematical measurement model developed by the Danish mathematician Georg Rasch (9). Widely used in education (student ability and test item difficulty), the model first made an appearance in the health literature in the late 1970s (11-12). Its use accelerated after the seminal paper by Wright and Linacre in 1989 (13), and then again in the later 1990s with the increasing availability of user-friendly software. Consequently 89% of all Rasch articles indexed on MEDLINE have been published from 1998 onwards, and half since 2006.

The Rasch model shows what should be expected in responses to items if interval scale measurement (at the metric level) is to be achieved. These responses are tested against what is expected by the model, which is a probabilistic form of Guttman scaling (14). Guttman scaling has a deterministic pattern that expects a strict hierarchical ordering of items (e.g., from low to high levels of activity limitation) such that if (in the dichotomous case) a patient has affirmed an item representing a task of average difficulty, then all the items below that task on the scale (i.e., easier tasks) should also be affirmed. The Rasch model relaxes this to say that if a harder task is affirmed, then there is a high probability that easier tasks will also be affirmed.

Consequently the Rasch model asserts that the easier the item the more likely it will be passed, and the more able the person, the more likely

he will pass an item compared to a less able person. Formally the probability that a person will affirm an item (in its dichotomous form) is a logistic function of the difference between the person's ability (θ) and the difficulty of the item (b) (i.e. the ability required to affirm item i) and only a function of that difference

$$\ln \frac{P_{ni}}{1 - P_{ni}} = \theta_n - b_i$$

where P_{ni} is the probability of person n affirming item i , θ_n is person ability, b_i is the item difficulty of item i (expressed in the logit form).

From this, the expected pattern of responses to an item set is determined given the estimated θ and b . When the observed response pattern coincides with or does not deviate too much from the expected response pattern then the items constitute a true Rasch scale. The model can be extended to cope with items with more than two response categories (polytomous items) and this involves an explicit 'threshold' parameter, where the threshold represents the equal probability point between any two adjacent categories within an item. Both the rating scale model (RSM) and the partial credit model (PCM) can be used for the polytomous items. The RSM includes an additional parameter τ_k , which represents the threshold location on the underlying construct being measured by the scale. The RSM differs from the PCM in that in the former the distance between threshold parameters from category to category within each item is the same across all items (15). The constraint that a fixed set of threshold parameters are used for the entire item set requires the item formats to be similar throughout the scale. The RSM equation, in the logit form is

$$\ln \frac{P_{nik}}{1 - P_{nik}} = \theta_n - (b_i + \tau_k)$$

where P_{nik} is the probability of person n affirming category k in item i , compared with an adjacent category ($k-1$); θ_n is person ability, b_i is the difficulty of the i^{th} item and τ_k is the k^{th} threshold which is the probabilistic midpoint (i.e., 50/50) between any 2 adjacent categories.

The PCM is one in which each item can have a different distance between thresholds. The PCM can accommodate both dichotomous and polytomous items at the same time (i.e. items with different response categories) (16). The PCM equation, in the logit form is:

$$\ln \frac{P_{nik}}{1 - P_{nik}} = \theta_n - b_{ik}$$

where P_{nik} is the probability of person n affirming category k in item i , compared with an adjacent category ($k-1$); θ_n is person ability, b_{ik} is the difficulty of the k^{th} threshold which is the probabilistic midpoint (i.e., 50/50) between any 2 adjacent categories in item i .

When data fit the Rasch model, because of the separation of parameters ($\theta_n - b$), item difficulties (i) are estimated independent of the person ability (θ), and vice versa, as either can be conditioned out. It is this aspect, the separation of parameters, which is a necessary condition to satisfy the requirements for constructing interval scale measurement (17). Consequently the Rasch measurement model transforms the ordinal raw scores derived from an assessment or outcome measure into interval measures which are objective, fundamental, and linear (18-19).

WHAT SOFTWARE IS AVAILABLE TO PERFORM RASCH ANALYSIS?

In the medical outcomes literature, most Rasch analysis is undertaken with proprietary software. The most commonly used packages are WINSTEPS, RUMM2030, and ConQuest, but more are available (20-22). WINSTEPS, RUMM2030 and ConQuest estimate person and item parameters using joint maximum likelihood estimation, pairwise conditional estimation and marginal maximum likelihood estimation methods, respectively. Each report findings in a slightly different way, although the basic premise is to test whether the response pattern observed in the data matches the theoretical pattern expected by the model (i.e., the probabilistic form of Guttman scaling). This difference (between observed and expected) is at the heart of the statistics used to test if the data fit the model.

WHAT IS THE PROCESS OF RASCH ANALYSIS?

Thus the process of Rasch analysis involves testing the assumption of stochastic ordering (probabilistic Guttman pattern) in the data, and also that other assumptions such as local independence and unidimensionality are satisfied. Depending upon the type of data to be analysed, a certain analytical strategy can be followed which is illustrated below using the RUMM2030 software (with occasional reference to relevant commands and output in WINSTEPS and ConQuest).

The threshold ordering of polytomous items

Before evaluation of item fit, where polytomous items are involved, the response categories should be examined for correct ordering. This involves the examination of the threshold pattern, the threshold being the transition point between adjacent categories. This ordering of thresholds is graphically demonstrated using the category probability curves (Figure 1). For an item with an appropriate ordering of thresholds, each response option would demonstrate the highest probability of endorsement at a specific range of the scale, with successive thresholds found at increasing levels of the construct being measured. Specifically, the transition between

categories (called thresholds) should show an increase along the underlying trait. Thus in Figure 1 the transition between categories 1 & 2 is higher on the underlying trait than the transition between categories 0 & 1, as would be expected.

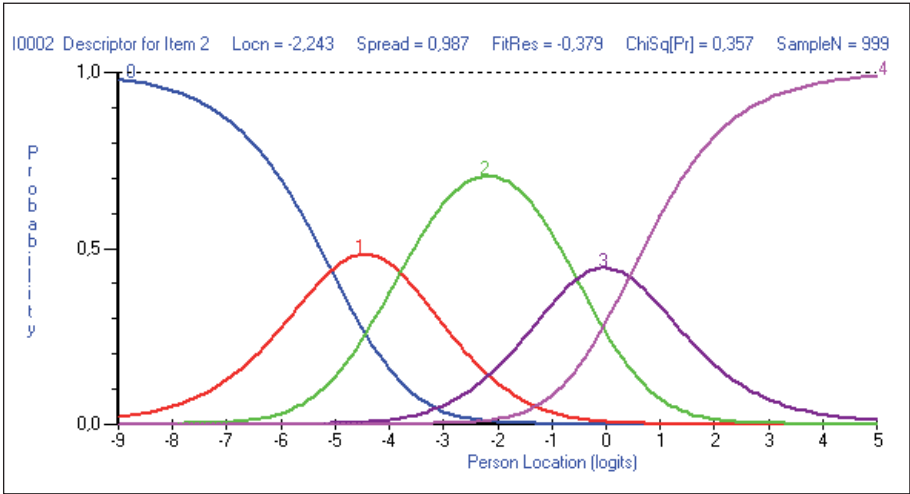


FIGURE 1. An example of ordered thresholds for a polytomous item with five response categories.

One of the most common sources of item misfit concerns respondents' or raters' inconsistent use of these response options resulting in what is known as 'disordered thresholds'. In these circumstances the expected order is not found. In Figure 2 the transition between categories

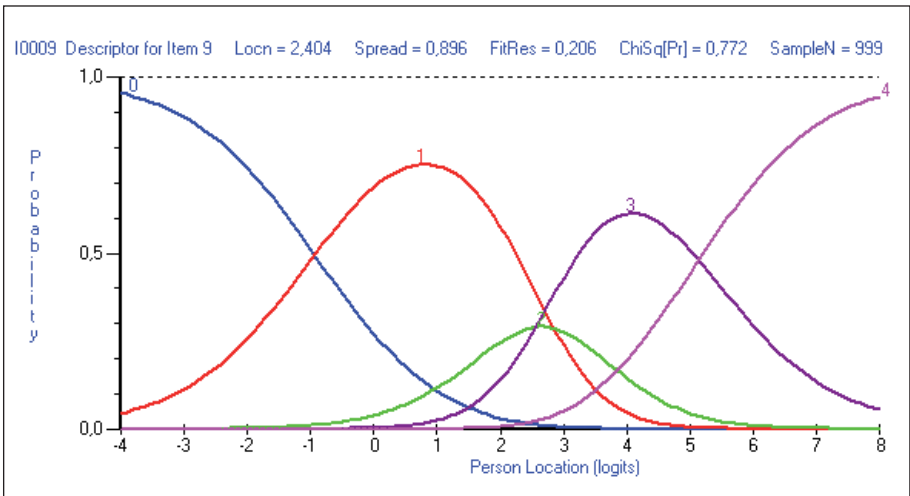


FIGURE 2. An example of disordered thresholds for a polytomous item with five response categories.

2 & 3 is lower on the trait than the transition between categories 1 & 2, which would not be expected. Often the collapsing of categories (i.e. grouping them together) where disordered thresholds occur improves overall fit to the model (23).

Once thresholds are found to be ordered (which is not required for dichotomous data where each item has only one threshold), the analysis can proceed to testing the various assumptions.

Local Independence

Local independence means that the response to any item should be uncorrelated to any other, once the main (Rasch) factor of interest has been conditioned out. This analysis is undertaken by a correlation of the item residuals, where high positive residuals (>0.3) would indicate a breach of local dependency (24). In RUMM this is shown by the residual item correlation matrix and in WINSTEPS by the ICORFILE command. A breach in the assumption of local independence of items can be found in 2 ways, through response dependency and multidimensionality (see below) (25). Response dependency is where items are linked in some way, such that the response on one item will determine the response on another. The classic example of this is where two stair climbing items are included in a scale. If a person can climb several flights of stairs, they must be able to climb a single flight of stairs. Such sets of items inflate classic reliability and affect parameter estimation in Rasch analysis. They are identified through the residual correlation matrix and dealt with by combining the items into a testlet (26). In the stair example it would be the equivalent of making one stair climbing item with response options that relate to how many flights of stairs a person can climb. It is important to deal with local dependency at the outset, as its presence may affect the item fit statistics.

Tests of fit to the model

Examining the fit of the data to the Rasch model tests the assumption of the stochastic ordering of items, that is, they have a probabilistic relationship to one another (measuring from less to more of the construct) in a way that satisfies the model expectation. Consequently, most of the fit statistics are based on the difference between the observed response and that expected by the model. Unfortunately these differ across software, and there is little overlap. In RUMM2030 (and its earlier versions) there are summary and item- and person-specific fit statistics. In other words, it is possible to see if both items and persons are consistent with model expectations.

Overall item fit statistics, overall person fit statistics and item-trait interaction statistics are presented. If the data accord to the model expectation, the mean of the overall item and the overall person fit statistics should be close to 0 and their standard deviation close to 1. A third summary fit statistics is an item-trait interaction statistic reported as a Chi-

Square, reflecting the property of invariance across the trait. A significant chi-square indicates that the hierarchical ordering of the items varies across the trait, compromising the required property of invariance. Individual item chi-square statistics are also available, giving detailed information about item-deviation from model expectations. An individual item residual fit statistic is also calculated, based on the standardised residuals (differences between the observed and expected responses divided by square root of variance and calculated for each patient for a given item). To obtain an overall statistic for an item, the standardised residuals are squared and summed over the patients. The individual item fit statistic is calculated by transforming this overall statistic to make it more nearly approximate a standard normal deviate under the hypothesis that the data fit the model. Thus, it is concluded that the deviations between the responses and the model are no more than random errors. Residuals between ± 2.5 are deemed to indicate adequate fit to the model. A person fit statistic is constructed for each person in a way similar to that of each item.

In WINSTEPS the fit statistics are called INFIT and OUTFIT statistics, and in ConQuest the same statistics are called the Weighted and Unweighted fit. INFIT takes particular note of the difference between observed and expected response for those items that have a difficulty level near the person's ability level. OUTFIT includes the differences for all items, irrespective of how far away the item difficulty is from the person's ability. Thus INFIT is a weighted fit statistic in that it gives greater weight to responses to items close to the person's ability level. WINSTEPS and Conquest also have some standardized fit statistics (reported as ZSTD and t respectively) and these are similar to the RUMM residual statistic, which is the standardized sum of all differences between observed and expected values summed over all persons. There is a lack of consensus over appropriate ranges for these statistics. Typically, two approaches can be found in the literature; those that espouse a range such as 0.7-1.3 as the acceptable range for the INFIT and OUTFIT Mean Square fit statistics, and those that argue that the critical interval ratio of these statistics vary by sample size, and thus adjust the range accordingly (27).

Figure 3 illustrates a typical output of individual item fit from the RUMM2030 programme. These data were simulated to fit the Rasch model, and thus all residual values are within the range ± 2.5 , and all the p values obtained from chi-square fit statistics are above 0.05 (or Bonferroni adjusted value) (10).

It is important to note that a few respondents with atypical response patterns (identified by high positive or negative residuals) may seriously affect fit at the item level. Such aberrant response patterns may be due to a variety of causes, for example 'block responding' (where a subject simply scores all items the same), or co-morbidity such as cognitive deficits. Therefore, where some respondents misfit in this way, removal from the analysis may make a significant difference to a scale's internal construct validity, while at the same time raising questions about the external construct validity of the scale with the particular patient group.

INDIVIDUAL ITEM-FIT for Analysis Name DENEEME1 - Serial Order														
	Seq	Item	Type	Location	SE	FitResid	DF	ChiSq	DF	Erob	R-stat	DF-1	DF-2	Prob
1	1	I0001	Poly	-2,929	0,054	-0,344	895,20	9,806	9	0,383328	1,146	9	989	0,326942
2	2	I0002	Poly	-2,243	0,051	-0,379	895,20	9,923	9	0,356752	1,214	9	989	0,282073
3	3	I0003	Poly	-1,625	0,053	-0,112	895,20	3,537	9	0,939178	0,416	9	989	0,927276
4	4	I0004	Poly	-0,989	0,047	-0,306	895,20	8,068	9	0,527261	1,042	9	989	0,404038
5	5	I0005	Poly	-0,389	0,046	-1,396	895,20	12,512	9	0,185983	1,710	9	989	0,082296
6	6	I0006	Poly	0,299	0,045	-1,048	895,20	15,230	9	0,084818	2,012	9	989	0,035054
7	7	I0007	Poly	0,992	0,045	-0,593	895,20	8,013	9	0,532833	0,993	9	989	0,443819
8	8	I0008	Poly	1,664	0,052	-0,641	895,20	6,154	9	0,724403	0,736	9	989	0,675857
9	9	I0009	Poly	2,404	0,054	0,206	895,20	5,677	9	0,771713	0,699	9	989	0,710158
10	10	I0010	Poly	2,616	0,059	1,137	895,20	9,136	9	0,424857	0,937	9	989	0,491334

FIGURE 3. Fit statistics for a 10-item polytomous scale using the RUMM2030 programme.

Unidimensionality

The Rasch model is a unidimensional measurement model, therefore the assumption is that the items summed together form a unidimensional scale. This is a requirement for summing together any set of items (28), and in the literature it can be considered as part of the local independence assumption (see above), or more usually is dealt with separately. There are various ways to test this assumption, but the Rasch programs usually provide a principal component analysis of the residuals. The absence of any meaningful pattern in the residuals will support the assumption of unidimensionality. This analysis is handled differently in the different software. In RUMM2030 a test for unidimensionality, proposed by Smith EV (29), takes the patterning of items in the residuals, examines the correlation between items and the first residual factor, and uses these patterns to define two subsets of items (i.e., the positively and negatively correlated items). These two sets of items are then used to make separate person estimates, and, using an independent *t*-test for the difference in these estimates for each person, the percentage of significant different *t*-tests outside the range -1.96 to 1.96 should not exceed 5%. A confidence interval for a binomial test of proportions is calculated for the proportion of observed number of significant tests, and the lower bound should overlap the 5% expected value for the scale to be unidimensional. Given that the differences in estimates derived from the two subsets of items are normally distributed, this approach is robust enough to detect multidimensionality (30) and appears to give a test of strict unidimensionality, as opposed to essential unidimensionality (31). In the latter case a dominant factor occurs, and although other factors exist, they are not deemed to compromise measurement. In the current example, the proportion of significant tests is just 5% (Figure 4).

Testing for differential item functioning

Another important aspect of the internal integrity of a scale is the absence of item bias or differential item functioning (DIF) (32). At a given level of disability it is important that the response to any item is unaffected by group membership. For example, *at the same level of disability it*



FIGURE 4. Test for unidimensionality.

is important that both young and old have the same probability of affirming an item. If not, then the scale works in different ways by age, rendering comparison between age groups difficult. Two types of DIF may be identified. One is where the group shows a consistent systematic difference in their responses to an item, across the whole range of the attribute being measured, which is referred to as uniform DIF (32). When there is non-uniformity in the differences between the groups (e.g., differences vary across levels of the attribute), then this is referred to as non-uniform DIF. Figure 5 shows an item where there is no DIF present by gender. Note that in this case the Item Response Function Curves for each gender closely match one another whereas if DIF were present they would show a greater separation. A statistical test is available to determine if the difference is significant.

In RUMM2030, DIF is assessed by a two-way analysis of variance of the residuals (i.e., the difference of expected scores and actual scores)

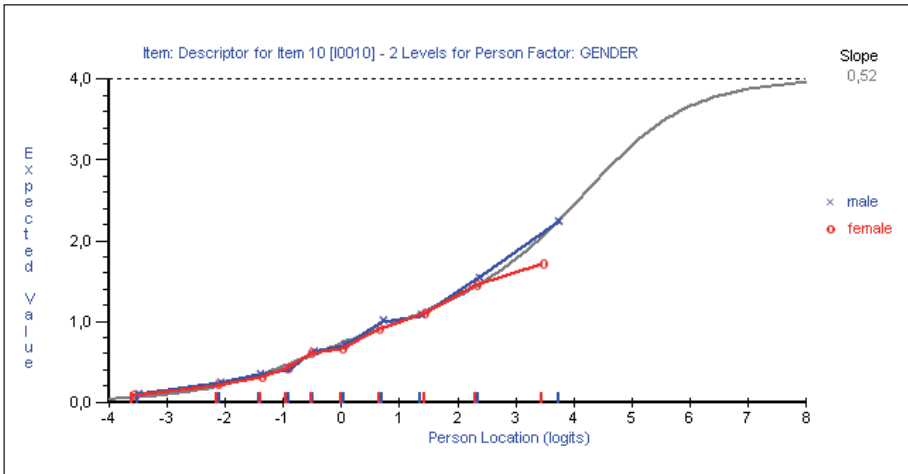


FIGURE 5. Differential Item Functioning by gender.

where subjects are grouped into classes of ± 50 persons each the continuum of the construct being measured, each group being defines as a 'class interval'. This provides one factor, while the second factor is represented by the groups for which DIF is to be investigated, such as age or gender.

In WINSTEPS, the analysis proceeds as though all items and persons, except the currently targeted item or person group, are anchored at the measures from the main analysis (estimated from all persons and items, including the currently targeted one). The item or person measure for the current group is then computed, along with its standard error. Mathematically, it is unlikely that no bias effects will be observed, or that bias sizes will cancel out exactly. The DIF contrast is the difference between the DIF sizes, and is a log-odds estimate, equivalent to a Mantel-Haenszel DIF size. The t is the DIF contrast divided by the joint standard error of the two DIF measures. It is equivalent to the Mantel-Haenszel significance test (20).

Thus DIF tests for the property of invariance by group membership. While age and gender are common factors assessed by DIF, this type of analysis is also essential to confirm the invariance of a scale by different cultures, and the analysis of cross-cultural validity by DIF analysis within the framework of the Rasch model has become commonplace (33-34).

Sample size and targeting of persons and items

A minimum sample size of 243 subjects is required to provide accurate estimates of item and person locations irrespective of the scale targeting (35). This is much smaller when subjects are well targeted to the scale (e.g. 108 for perfect targeting) and it has also been argued that 50 well-targeted cases are sufficient to see if the scale is displaying serious misfit to model expectations (36). Most Rasch software will provide a

graphical display of the targeting, which is the association between persons and items. As the scale is always centred on zero logits, the average person measure gives an indication of how well targeted the scale is to the current sample. In Figure 6, the average person logit is 0.062, (i.e. close to zero) which is reflective of the fact that the items are almost perfectly targeted, covering a broad range of the construct where the majority of the sample is also located.

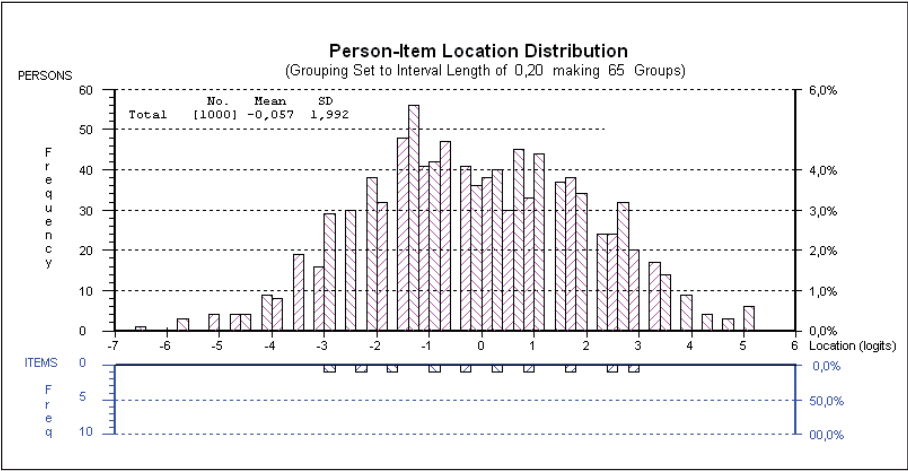


FIGURE 6. Targeting of scale to persons.

Reliability

An estimate of the internal consistency reliability of the scale is also available, based on the Person Separation Index (PSI), where the estimates on the logit scale for each person are used to calculate reliability. In RUMM2020 and earlier versions this can be interpreted as equivalent to Cronbach's alpha coefficient (37) but has the linear transformation from the Rasch model substituted for the ordinal raw score (38). Cronbach's alpha coefficient can be calculated with complete data only, while the PSI can be calculated with random missing data. With substantial missing data, the values for the two indices might be different. In the simulated data set the classical Cronbach's alpha was 0.926; whereas the PSI was 0.933. This level of reliability is consistent with use of scales at the individual level, whereas a value of 0.7 is considered the minimum for group use (39).

This convergence between the PSI and Cronbach's alpha will hold when data are distributed normally. However in the presence of skewed data, in RUMM2030 the PSI value may diverge considerably from the Cronbach's alpha, reflecting the lack of true discrimination of the item set, which Cronbach's alpha is unable to identify (40). Consequently in the presence of skewed data, the PSI, which can be used to assess how many distinct groups of persons can be differentiated, this value may indicate a significant lack of discrimination (38, 40).

WHY SHOULD RASCH ANALYSIS BE USED?

Patient-reported outcomes and standardised assessments are used widely throughout Physical and Rehabilitation Medicine, both in clinical and research contexts. Rasch analysis can support this process by providing confirmation of the internal construct validity of such assessments, and providing a transformation of an ordinal score into a linear, interval-level variable. Thus Rasch analysis allows for a unified approach to the construction and the internal construct validity of such assessments through the testing of its assumptions, and additional aspects such as category ordering (whether or not the category ordering of polytomous items are working as expected) and invariance of items across groups (DIF).

WHEN SHOULD RASCH ANALYSIS BE USED?

There are different occasions when Rasch analysis would be applied. Firstly, it can be used for the development of a new scale by making a substantial contribution to this process where it is possible to design the item set to fit the model expectations from the outset (41). Secondly, it can be used in reviewing the psychometric properties of existing ordinal scales, including testing their dimensionality. While factor analytic techniques are commonly used for this purpose, several aspects of ordinal scaling give rise to factors which may confound such analysis. For example, item difficulty clusters, DIF and local dependency can all give rise to apparent multidimensional structures (42-43). Thus post-hoc testing of unidimensionality, after such aspects have been accommodated, may offer a better solution. However, where the dimensional structure of the data is uncertain from the outset, exploratory factor analysis may provide initial guidance as to potential domains.

Thirdly, Rasch analysis can also be used in constructing item banks as the basis of computer adaptive testing (CAT) (44-47). With calibrated item banks (where the difficulty of items has been previously established on a single metric), it is possible to use computer algorithms to present items to patients in such a way that their level on the construct to be measured can be determined by just a few questions. Finally, Rasch analysis would be applied whenever change scores need to be calculated from ordinal scales (13). The data must be shown to meet model expectations so that an interval (logit-based) estimate can be derived.

CONCLUSION

The assessment and measurement of outcome in Physical and Rehabilitation Medicine often involve the use of instruments that deliver ordinal scaling. There are many appropriate statistical techniques to accommodate this type of scaling. However, more often than not, parametric techniques

are preferred, including the use of change scores. The emergence of Rasch analysis as a technique for converting ordinal scores to interval level measures has opened up and legitimised the use of parametric techniques, at the same time providing the type of scaling that will support mathematical operations such as the calculation of a change score. Furthermore, it provides a comprehensive diagnostic framework for the quality of such instruments, including the properties of invariance. It also provides a sound scientific framework to support the emergence of new approaches to health outcome assessment, such as computer adaptive testing.

REFERENCES

1. Tesio L. Functional assessment in rehabilitative medicine: Principles and methods. *Eura Medicophys* 2007; 43: 515-523.
2. Wade DT. Outcome measurement and rehabilitation. *Clin Rehabil* 1999; 13: 93-95.
3. Mahoney FI, Barthel DW. Functional evaluation: the Barthel index. *Maryland State Med J* 1965; 14: 61-65.
4. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980; 23: 137-145.
5. Melzack R. The McGill Pain Questionnaire: major properties and scoring methods. *Pain* 1975; 1: 277-299.
6. Forrest M, Andersen B. Ordinal scale and statistics in medical research. *BMJ* 1986; 292: 537-538.
7. Merbitz C, Morris J, Grip JC. Ordinal scales and foundations of misinference. *Arch Phys Med Rehabil* 1989; 70: 308-312.
8. Svensson E. Guidelines to statistical evaluation of data from rating scales and questionnaires. *J Rehabil Med* 2001; 33: 47-48.
9. Rasch G. Probabilistic models for some intelligence and attainment tests. Chicago: University of Chicago Press; 1960.
10. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Rheum* 2007; 57: 1358-1362.
11. Bech P, Allerup P, Rosenberg R. The Marke-Nyman temperament scale. Evaluation of transferability using the Rasch item analysis. *Acta Psychiatr Scand* 1978; 57: 49-58.
12. Haracopos D, Kelstrup A. Psychotic behavior in children under the institutions for mentally retarded in Denmark. *J Autism Child Schizophr* 1978; 8: 1-12.
13. Wright BD, Linacre JM. Observations are always ordinal; measurements, however, must be interval. *Arch Phys Med Rehabil* 1989; 70: 857-60.
14. Guttman LA. The basis for Scalogram analysis. In Stouffer SA, Guttman LA, Suchman FA, Lazarsfeld PF, Star SA, Clausen JA, editors. *Studies in social psychology in World War II: Vol 4. Measurement and Prediction*. Princeton: Princeton University Press; 1950, pp. 60-90.
15. Andrich D. Rating formulation for ordered response categories. *Psychometrika* 1978; 43: 561-573.
16. Masters G. A Rasch model for partial credit scoring. *Psychometrika* 1982; 47: 149-174.
17. Luce RD, Tukey JW. Simultaneous conjoint measurement: A new type of fundamental measurement. *J Math Psychol* 1964; 1: 1-27.
18. Perline R, Wright BD, Wainer H. The Rasch model as additive conjoint measurement. *Appl Psych Meas* 1979; 3: 237-256.
19. Van Newby VA, Conner GR, Grant CP, Bunderson V. The Rasch model and additive conjoint measurement. *J Appl Meas* 2009; 10: 348-354.
20. Linacre JM. A User's Guide to WINSTEPS/Ministeps. Rasch-model computer programs. Program Manual 3.70.0; 2010.

21. Andrich D, Lyne A, Sheridan B, Luo G. RUMM 2030. Perth: RUMM Laboratory; 2009.
22. Wu ML, Adams RJ, Wilson MR. ConQuest: Generalised item response modelling software (computer program). Camberwell, VIC, Australia: Australia Council for Educational Research; 1998.
23. Pallant JF, Tennant A. An introduction to the Rasch measurement model: an example using the hospital anxiety and depression scale (HADS). *Br J Clin Psychol* 2007; 46(Pt 1):1-1.
24. Andrich D. Rasch models for measurement. Newbury Park CA: Sage; 1988.
25. Andrich D, Marais I. Formalizing Dimension and Response Violations of Local Independence in the Unidimensional Rasch Model. *J Appl Meas* 2008; 9: 200-215.
26. Wainer H, Kiely G. Item clusters and computer adaptive testing: A case for testlets. *J Educ Meas* 1987; 24: 185-202.
27. Smith RM. Fit analysis in latent trait measurement models. *J Appl Meas* 2000; 2: 199-218.
28. Thurstone LL. Attitudes can be measured. *Am J Sociol* 1928; 33: 529-554.
29. Smith EV. Detecting and evaluation the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *J Appl Meas* 2002; 3: 205-231.
30. Tennant A, Pallant JF. Unidimensionality Matters. *Rasch Measure Trans* 2006; 19: 1048-1051.
31. Stout WF. A new item response theory modelling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika* 1990; 55: 293-325.
32. Teresi JA, Kleinman M, Ocepek-Welikson K. Modern psychometric methods for detection of differential item functioning: application to cognitive assessment measures. *Stat Med* 2000; 19: 1651-1683.
33. Küçükdeveci, AA, Şahin H, Ataman, S, Griffiths B, Tennant A. Issues in cross-cultural validity: Example from the adaptation, reliability and validity testing of a Turkish version of the Stanford Health Assessment Questionnaire (HAQ). *Arthritis Rheum* 2004; 51: 14-19.
34. Gilworth G, Emery P, Gossec L, Vliet Vlieland TP, Breedveld FC, Hueber AJ, et al. Adaptation and cross-cultural validation of the RA-WIS (Work Instability Scale). *Ann Rheum Dis* 2009; 68: 1686-1690.
35. Linacre JM. Sample Size and Item Calibration Stability. *Rasch Measure Trans* 1994; 7: 328.
36. Wright BD, Tennant A. Sample size again. *Rasch Measure Trans* 1996; 9: 468.
37. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951; 16: 297-333.
38. Fisher WP. Reliability statistics. *Rasch Measure Trans* 1992; 6: 238.
39. Nunally JC. Psychometric theory. New York: McGraw-Hill; 1978.
40. Wright BD. Local dependency, correlations and principal components. *Rasch Measure Trans* 1996; 10: 509-511.
41. Mills RJ, Young CA, Pallant JF, Tennant A. Development of a patient reported outcome scale for fatigue in multiple sclerosis: The Neurological Fatigue Index (NFI-MS). *Health Qual Life Outcomes* 2010; 8:22
42. Guilford JP. The difficulty of a test and its factor composition. *Psychometrika* 1941; 6: 67-77.
43. Gorsuch RL. Factor analysis (2nd ed.). Hillsdale, NJ: Erlbaum; 1983.
44. Gershon RC. Computer adaptive testing. *J Appl Meas* 2005; 6: 109-127.
45. Kopec JA, Sayre EC, Davis AM, Badley EM, Abrahamowicz M, Sherlock L, et al. Assessment of health-related quality of life in arthritis: Conceptualization and development of five item banks using item response theory. *Health Qual Life Outcomes* 2006; 4: 33.
46. Lai JS, Cella D, Chang CH, Bode RK, Heinemann AW. Item banking to improve, shorten and computerize self-reported fatigue: an illustration of steps to create a core item bank from the FACIT-Fatigue Scale. *Qual Life Res* 2003; 12: 485-501.
47. Elhan AH, Öztuna D, Kutlay Ş, Küçükdeveci AA, Tennant A. An initial application of computerized adaptive testing (CAT) for measuring disability in patients with low back pain. *BMC Musculoskelet Disord* 2008; 9: 166.

MAURO ZAMPOLINI¹, PATRIZIO SALE²
ALESSANDRO GIUSTINI³
MARCO FRANCESCHINI²

EXPERIMENTAL DESIGN IN PHYSICAL AND REHABILITATION MEDICINE

¹ Department of Rehabilitation,
ASL 3 Umbria, Italy

² Department of Neuro-Rehabilitation,
IRCCS San Raffaele Pisana,
Rome, Italy

³ Rehabilitation Hospital, San Pancrazio,
Arco (Trento), Italy

The first methodology for designing experiments was introduced by a Ronald A. Fisher, in his book 'The Design of Experiments' (1935), which effectively founded the field of experimental design (1).

The term experimental design refers to a plan for assigning experimental units to treatment conditions. Analysis of the design of experiments was built on the foundation of the analysis of variance, a collection of models in which the observed variance is partitioned into components due to different factors which are estimated and/or tested.

Analysis of the design of experiments was built on the foundation of the analysis of variance, a collection of models in which the observed variance is partitioned into components due to different factors which are estimated and/or tested.

1. CRITICAL ISSUES OF EXPERIMENTAL DESIGN IN PRM

Physical and rehabilitation medicine, compared with the other specialties lack the recognition they deserve in the scientific community. The reason is the lack of studies that support the evidence of the efficacy of rehabilitation medicine treatments (2).

We need to prove to the scientific community that our diagnostic and therapeutic tools are clinically effective and cost-efficient, to help recognize the worth of PRM as a fundamental medical specialty in tomorrow's increasingly competitive health care system. The best way to do this is to undertake well conducted clinical research, such as standard controlled randomized clinical trials. In PRM, however, there are some problems, in designing studies with a placebo group or blindness of treatment.

Another aspect which needs improving in Rehabilitation Medicine research is the use of measurement systems appropriate to quantify the improvement of functioning. Human functioning is characterized by a behavioural relationship between person and environment. The limit is that usually the measurement is performed with ordinal scales that lack linearity and with which we cannot apply robust statistical methods. We need to take full advantage of the scientific method and the Evidence-Based Medicine tools that require proper scientific measurement.

In this respect, Rasch Analysis may be the long-awaited tool by which Rehabilitation Medicine may bring its scientific credibility close to that of physical sciences, freeing itself forever from the unmerited role of the Cinderella of this research area and filling the otherwise widening gap existing between itself and the other branches of medicine.

2. VALIDITY OF EXPERIMENTAL DESIGN

2.1 Randomization

Randomization is a core principle in statistical theory, the importance of which was emphasized by Charles S. Peirce in 'Illustrations of the Logic of Science' (1877-1878) and 'A Theory of Probable Inference' (1883).

The use of randomization provides a basis for an assumption-free statistical test of the equality of treatments. Randomization-based inference is especially important in experimental design and in survey sampling. The principal goal of Random Allocation is a known chance of receiving a treatment but one cannot predict the treatment that will be given.

In the statistical theory of the design of experiments, randomization involves randomly allocating the experimental units across the treatment groups. In particular, it is important to eliminate the selection bias and to have similar treatment groups.

For example, if an experiment compares a new drug against a standard drug, then the patients should be allocated to either the new drug or to the standard control drug using randomization.

Randomized experimentation is not haphazard. Randomization reduces bias by equalising so-called factors (independent variables) that have not been accounted for in the experimental design.

We have different types of randomization: Simple Randomization (randomize each patient to a treatment with a known probability), Blocked Randomization (ensure the number of patients assigned to each treatment

is not far out of balance) and Stratified Randomization (randomize so that different levels of the factor are balanced between treatment groups).

2.2 Blinding

Blind experiments are an important tool of the scientific method, in many fields of research from medicine, forensics, psychology and the social sciences, to basic sciences such as physics and biology and to market research. A blind or blinded experiment is a scientific experiment where some of the persons involved are prevented from knowing certain information that might lead to conscious or unconscious bias on their part, invalidating the results. Blinding is a basic tool to prevent conscious and unconscious bias in research. For example, in open taste tests comparing different product brands, consumers usually choose their regular brand. However, in blind taste tests, where the brand identities of the products to be tasted are concealed, consumers may favour a different brand than the one they usually choose.

Similarly, when evaluating the effectiveness of a medical drug, both the patients and the doctors who administer the drug may be kept in the dark about the dosage being applied in each case - to forestall any chance of a placebo effect, observer bias, or conscious deception.

Blinding can be imposed on researchers, technicians, subjects, funders, or any combination of them. The opposite of a blind trial is an open trial. In some disciplines, such as drug testing, blind experiments are considered essential.

2.2.1 *Single Blind Test*

In a single blind experiment, the individual subjects do not know whether they are so-called 'test' subjects or members of an 'experimental control' group. Single-blind experimental design is used where the experimenters either must know the full facts (for example, when comparing sham to real treatment) and so the experimenters cannot themselves be blind, or where the experimenters will not introduce further bias and so the experimenters need not be blind. However, there is a risk that subjects are influenced by interaction with the researchers - known as the experimenter's bias. Single-blind trials are especially risky in psychology and social science research, where the experimenter has an expectation of what the outcome should be, and may consciously or subconsciously influence the behaviour of the subject.

2.2.2 *Double-blind trials*

In a double-blind experiment, neither the individuals nor the researchers know who belongs to the control group and the experimental group. Only after all the data are recorded (and in some cases, analyzed) do the researchers learn which individuals are which. Performing an experiment in double-blind fashion is a way to reduce the influence of the prejudices and unintentional physical cues on the results (the placebo effect, observer bias, and experimenter's bias). Random assignment of the subject to the experimental or control group is a critical part of double-blind research design. The key that identifies the subjects and which

group they belonged to is kept by a third party and not given to the researchers until the study is over.

Double-blind methods can be applied to any experimental situation where there is the possibility that the results will be affected by conscious or unconscious bias on the part of the experimenter.

2.3 Using variables

The level of measurement of a variable in mathematics and statistics describes how much information the numbers associated with the variable contain.

In statistics, the kinds of descriptive statistics and significance tests that are appropriate depend on the level of measurement of the variables concerned.

In practice the variables are things that change. There are two types of variable: independent and dependent. The independent variable is the variable that is purposely changed. It is the manipulated variable. The dependent variable changes in response to the independent variable. It is the responding variable.

2.4 Analyzing data of the research protocol

Data Management is the process by which the required data is acquired, validated, stored, protected and processed, and by which its accessibility, reliability and timeliness are ensured to satisfy the needs of the data users. Before starting a new scientific research project, the Principal Investigator (PI) and research team must address issues related to data management (Tab. 1).

TABLE 1. Issues related to data management that research teams must address

• Data Ownership - This pertains to who has the legal rights to the data and who retains the data after the project is completed, including the PI's right to transfer data between institutions.
• Data Collection - This pertains to collecting project data in a consistent, systematic manner (i.e. reliability) and establishing an ongoing system for evaluating and recording changes to the project protocol (i.e. validity).
• Data Storage - This concerns the amount of data that should be stored so that project results can be reconstructed.
• Data Protection - This relates to protecting written and electronic data from physical damage and protecting data integrity, including damage from tampering or theft.
• Data Retention - This refers to the length of time one needs to keep the project data according to the sponsor's or founder's guidelines. It also includes secure destruction of data.
• Data Analysis - This pertains to how raw data are chosen, evaluated and interpreted into meaningful and significant conclusions that other researchers and the public can understand and use.
• Data Sharing - This concerns how project data and research results are disseminated to other researchers and the general public, and when data should not be shared.
• Data Reporting - This pertains to the publication of conclusive findings, both positive and negative, after the project is completed.

2.5 Validity of the study

In science and statistics, validity has no single agreed definition but generally refers to the extent to which a concept, conclusion or measurement is well-founded and corresponds accurately to the real world. The word ‘valid’ is derived from the Latin *validus*, meaning strong. Validity of a measurement tool (e.g. test in education) is considered to be the degree to which the tool measures what it claims to measure.

In psychometrics, validity has a particular application known as test validity: “the degree to which evidence and theory support the interpretations of test scores” (as entailed by proposed uses of tests).

In the area of scientific research design and experimentation, validity refers to whether a study is able to scientifically answer the questions it is intended to answer.

In clinical fields, the validity of a diagnosis and associated diagnostic tests may be assessed.

It is generally accepted that the concept of scientific validity addresses the nature of reality and as such is an epistemological and philosophical issue as well as a question of measurement. The use of the term in logic is narrower, relating to the truth of inferences made from premises.

2.5.1 Internal validity

Internal validity is an inductive estimate of the degree to which conclusions about causal relationships can be made (e.g. cause and effect), based on the measurements used, the research setting, and the whole research design. Good experimental techniques, in which the effect of an independent variable on a dependent variable is studied under highly controlled conditions, usually allow higher degrees of internal validity than, for example, single-case designs. Eight kinds of confounding variable can interfere with internal validity (i.e. with the attempt to isolate causal relationships) (Tab. 2).

TABLE 2. Confounding variables can interfere with internal validity

1. History - the specific events occurring between the first and second measurements in addition to the experimental variables.
2. Maturation - processes within the participants as a function of the passage of time (not specific to particular events): e.g., growing older, hungrier, more tired and so on.
3. Testing - the effects of taking a test upon the scores of a second testing.
4. Instrumentation - changes in calibration of a measurement tool or changes in the observers or scorers may produce changes in the obtained measurements.
5. Statistical regression - operating where groups have been selected on the basis of their extreme scores.
6. Selection - biases resulting from differential selection of respondents for the comparison groups.
7. Experimental mortality - or differential loss of respondents from the comparison groups.
8. Selection-maturation interaction, etc. - e.g. in multiple-group quasi-experimental designs.

2.5.2 External validity

External validity concerns the extent to which the (internally valid) results of a study can be held to be true for other cases, for example for different people, places or times. In other words, it is about whether findings can be validly generalized. If the same research study was conducted in those other cases, would the same results be obtained? A major factor in this is whether the study sample (e.g. the research participants) are representative of the general population along relevant dimensions. Other factors jeopardizing external validity are:

- 1. Reactive or interaction effect of testing, a pretest might influence and increase the scores on a posttest.
- 2. Interaction effects of selection biases on the experimental variable.
- 3. Reactive effects of experimental arrangements, which would preclude generalization about the effect of the experimental variable upon persons being exposed to it in non-experimental settings.
- 4. Multiple-treatment interference, where effects of earlier treatments are not erasable.

2.6 Effectiveness vs. efficacy

Effectiveness means the capability of producing an effect. In medicine, effectiveness relates to how well a treatment works in practice, as opposed to efficacy, which measures how well it works in clinical trials or laboratory studies.

The word effective is sometimes used in a quantitative way, “being very or not very effective”. However, it does not inform on the direction (positive or negative) and the comparison to a standard of the given effect. Efficacy, on the other hand, is the ability to produce a desired amount of the desired effect, or success in achieving a given goal. Contrary to efficiency, the focus of efficacy is the achievement as such, not the resources spent in achieving the desired effect (Fig. 1). Therefore, what is effective is not necessarily efficacious, and what is efficacious is not necessarily efficient (3).

An ordinary way to distinguish among effectiveness, efficacy and efficiency:

- efficiency: doing things in the most economical way (good input to output ratio)



FIGURE 1. The clinical trials could range along a continuous variation ranging from ‘Efficacy Trials’ where the assessment of an intervention under highly controlled, experimentally near-ideal conditions and, on the other hand, ‘Effectiveness trials’ in which an intervention is assessed under real-world clinical conditions with relatively few restrictions.

- efficacy: getting things done, i.e. meeting targets
- effectiveness: doing the 'right' things, i.e. setting the right targets to achieve an overall goal.

2.7 Treatment theory

Treatment theories attempt to identify the features of interventions, recipients and their environment that comprise the causal sequence connecting interventions and outcomes (4).

In order to fully develop, such theories there is the need to include specifications of:

- the functional problems on which the intervention is intended to have an impact,
- characteristics of individuals that make them candidates for the intervention,
- critical features of the intervention that are ostensibly responsible for the intended outcomes,
- elements and contingencies in the causal chain connecting provision of the intervention with likely outcomes,
- expected changes in the recipients' status and in their environment that constitute those outcomes, both short and long term.

Such theories can help us to improve the efficacy of analysis of the active components and underlying mechanisms of interventions explaining failures to confirm the efficacy of other interventions.

2.8 Adherence

Another aspect to be considered is the means of assessing adherence in rehabilitation trials. One type of adherence regards participants' conformity with the regimens comprising interventions.

Interventions outside of a clinical setting (e.g., partly or wholly in people's homes) are especially problematic. The other kind of adherence regards participants' cooperation with the data-collection procedures (e.g., in connection with the post-discharge, long-term follow-up of an experimental treatment that was administered during inpatient rehabilitation).

Partial adherence with a treatment regimen results in underestimating its potential effects by virtue of participants receiving less than a complete 'dosage.'

Abandoning the trial may have a similar effect, especially if the data are analyzed according to the 'intention-to-treat guideline.' In order to avoid misinterpretation of the data, analyses should be based on the data of all participants who were randomized into the study, regardless of whether they completed it (5).

2.9 Treatment fidelity

The treatment fidelity aspect of RCTs is the appropriate manner to implement the features of experimental and control groups (6, 7).

The important aspects are:

- treatment integrity, an issue of whether the experimental intervention was implemented as intended
- if the experimental intervention and the intervention comprising the control condition differed in the intended manner.

Rehabilitation outcome studies typically describe the nature of the intended interventions but lack information about the actual administration.

If the intervention was inadequately administered we can have problems of internal validity that lead to negative results. A false positive effect may be attributable to contaminants that increase the potency of an otherwise ineffective intervention. The attempt to replicate a study whose actual interventions are only vaguely characterized or in considering how to apply those interventions in another setting could lead to problems related to external validity.

To minimise the potential problem of validity the practitioners who are administering the intervention could be better trained, more effectively supervised, supplied with guides or manuals that can be readily referenced, and if necessary, periodically retrained.

To establish the degree of treatment integrity that results, the investigator must document facts regarding who (delivering the intervention) did what (regarding the intervention's content) to whom (the participants) at particular points in time.

Another important aspect is the treatment differentiation through the control intervention to which the experimental treatment is being compared.

We should answer these questions:

- What defines that intervention and what are its critical features?
- With what fidelity was it administered?
- Did any contamination occur ?

Without clear answers the trial's findings are difficult to interpret.

3. TYPE OF EXPERIMENTAL DESIGN

3.1 The research designs

There are different dimensions of research designs. According to the purpose of the research we can distinguish:

1. Description
2. Analysis of relationships
3. Analysis of the differences (cause and effects)

From the timing point of view we can distinguish:

- Retrospective design: is the analysis of the past data usually collected from the clinical charts or other databases?
- Prospective design: are the data collected after the set-up of the research design?

In order to gain a better understanding of the differences described we can refer to the example of low back pain.

A first purpose could be to describe the variable such as intensity of pain, the mobility capacity, and the functional status during the recovery from low back pain (descriptive design).

Another possibility is to analyze the relationships between various predictive factors for the recovery from low back pain such as age, rest time and other variables (Exploratory design - Analysis of relationships).

A third possibility is to study the effect of the exercise on the recovery from acute low back pain. In this case we must design an experimental protocol dividing the patients into two groups: one that undertakes an individual rehabilitation program and another group that undertakes a generic exercise advice program (experimental design: cause ´ effect).

There is no hierarchical rank between the different research designs each of them could be useful according to the research question. The main difference is that if we wish to obtain the evidence of the effect of some particular intervention on the outcome of the patient we must use the third design, the experimental design.

One example of the different use of the different research paradigms are the trial phases of investigational studies, mostly used for the new pharmacological agents which, could also be used to explore and demonstrate the rehabilitative intervention (Fig. 2). Usually there is a preclinical phase, aimed at studying the effect of the intervention in animal models. The phase I is aimed at demonstrating whether the new therapy is safe with an analysis of the relationship between the intervention (drug or other) and any side effects in a small sample of subjects.

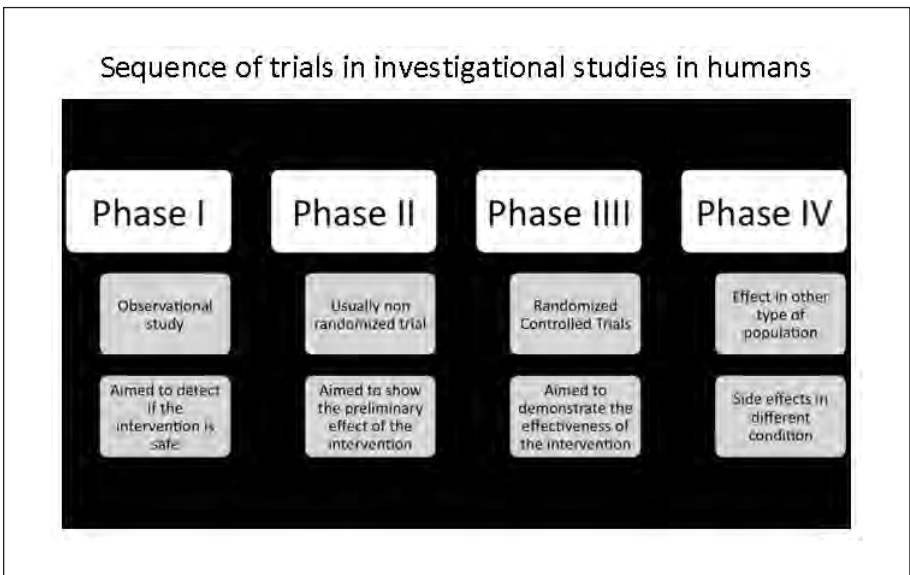


FIGURE 2. The researcher must choose the best clinical trial according to the best answer to the research question. In the figure there is an example.

Once the treatment has been shown to be safe there is the phase two that investigates the clinical effects of the drug, usually through a non randomized study. Once the effect of the treatment is promising the efficacy must be demonstrated with a Randomized Control trial. (phase III). The phase IV is aimed at testing the treatment in different conditions looking at the effects on the population.

An example, of this sequential methodology can be found is a recent study looking at the effects of Intrathecal Baclofen Therapy (ITB) carried out in Italy in acquired severe brain injury patients. The first study was a retrospective study aimed at looking at the effect of ITB in this group. The second study was a perspective study aimed at looking at the effect on spasticity before and after the treatment. From the data of these studies a randomised controlled trial was started, aimed at demonstrating the efficacy of ITB therapy in the subgroup of stroke patients.

Rehabilitation should move toward Randomized Controlled Trials to test the effectiveness of rehabilitative intervention in the meanwhile we should also consider the other designs according to the research question. The error to be avoided is to carry on with rehabilitative interventions without measuring anything and without considering studies to show the effectiveness of rehabilitation.

3.2 Experimental Design classifications

The experimental design is aimed at studying the cause-effect through a complete control of the dependent and independent variables that are influenced by some intervention. In this type of design at least two comparison groups are necessary. In these conditions, if all the characteristics are implemented, we are faced with the true experimental design.

The gold standard of the true experimental design is the Randomized Control Trial or Randomized Clinical Trial (RCT).

The comparison between two different groups is referred as 'Between-subjects Design'. Cases where the subject is the control of himself are referred as 'Within-Subject Design'.

The designs could be classified according to the number of independent variables (factors). In the case of one independent variable is referred as single factor (one way) design, in the case of multiple independent variables, multi-factors design.

The completely randomized design is probably the simplest experimental design, in terms of data analysis and convenience. With this design, participants are randomly assigned to treatments.

A completely randomized design relies on randomization to control for the effects of extraneous variables. The experimenter assumes that, on average, extraneous factors will affect treatment conditions equally; so any significant differences between conditions can be fairly attributed to the independent variable.

3.2.1 *Randomized Block Design*

With a **randomized block design**, the experimenter divides participants into subgroups called blocks, such that the variability within blocks is less than the variability between blocks. Then, participants within each block are randomly assigned to treatment conditions. Because this design reduces variability and potential confounding, it produces a better estimate of treatment effects.

This design ensures that each treatment condition has an equal proportion of men and women. As a result, differences between treatment conditions cannot be attributed to gender. This randomized block design removes gender as a potential source of variability and as a potential confounding variable.

A **matched pairs design** is a special case of the randomized block design. It is used when the experiment has only two treatment conditions; and participants can be grouped into pairs, based on some blocking variable. Then, within each pair, participants are randomly assigned to different treatments.

The **matched pairs design** is an improvement over the completely randomized design and the randomized block design. Like the other designs, the matched pairs design uses randomization to control for confounding. However, unlike the others, this design explicitly controls for two potential lurking variables - age and gender.

In order to improve the quality of presentation of randomized trial a consensus statement has been developed. The CONSORT (CONsolidated Standards of Reporting Trials) Statement helps researchers to improve reporting two-parallel design RCTs by using a checklist and flow diagram (8, 9).

3.3 Experimental design for independent groups

3.3.1 *One way design*

PRETEST-POSTTEST

This is the case of the study of the effect of an independent variable on the outcome dependent variables (one or more). A typical condition is the pretest-posttest control group design. In this case there are two groups, an experimental group where there is the intervention and a control group where there is no intervention or placebo. In order to show the effect of the intervention in both groups the independent variable is measured before and after the intervention (Fig. 3). In some cases the comparison is not with a placebo group but with another kind of intervention. This variation of the design is particularly important in rehabilitation. In fact if we wish to show the effect of a new technique it is not usually ethical to compare with no intervention because of the known effectiveness of the exercise whatever applied.

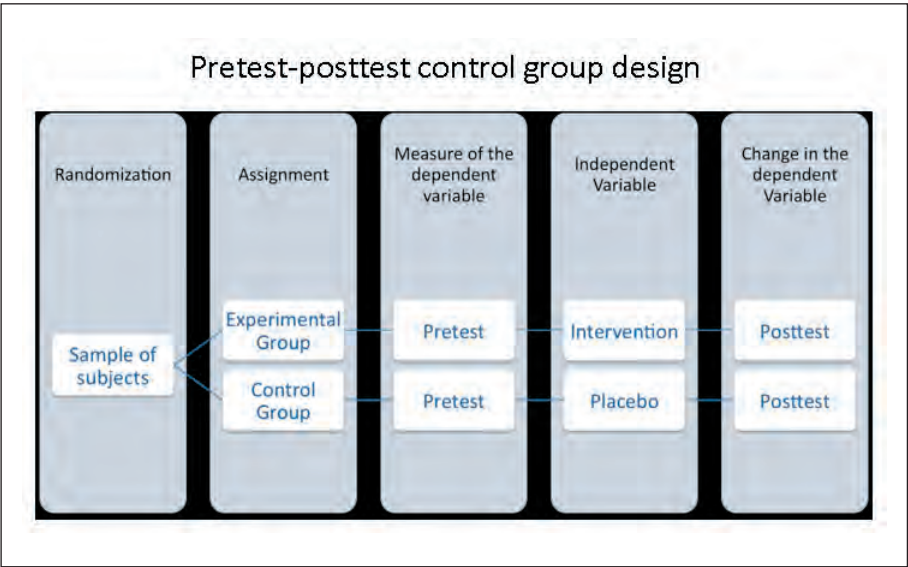


FIGURE 3. A typical experimental design Pretest-posttest Control Group Design.

A more complex design is the Multigroup pretest-posttest design where there is the comparison between more than two groups with the application of more than one intervention. In each group a different independent variable is applied maintaining the control group (Fig. 4).

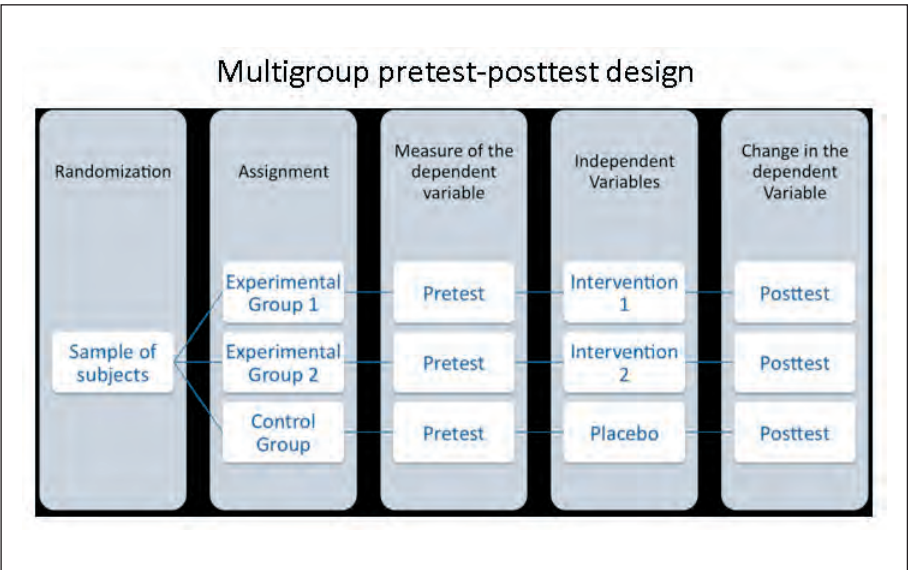


FIGURE 4. An example of experimental design with multiple experimental groups: multi-group pretest-posttest design.

PRETEST ONLY DESIGN

In this case there is not the pretest measurement in either group (Fig. 5). The potential bias is the non equivalence of the groups in pretest conditions. To limit this potential bias this design is appropriate with a large sample size.

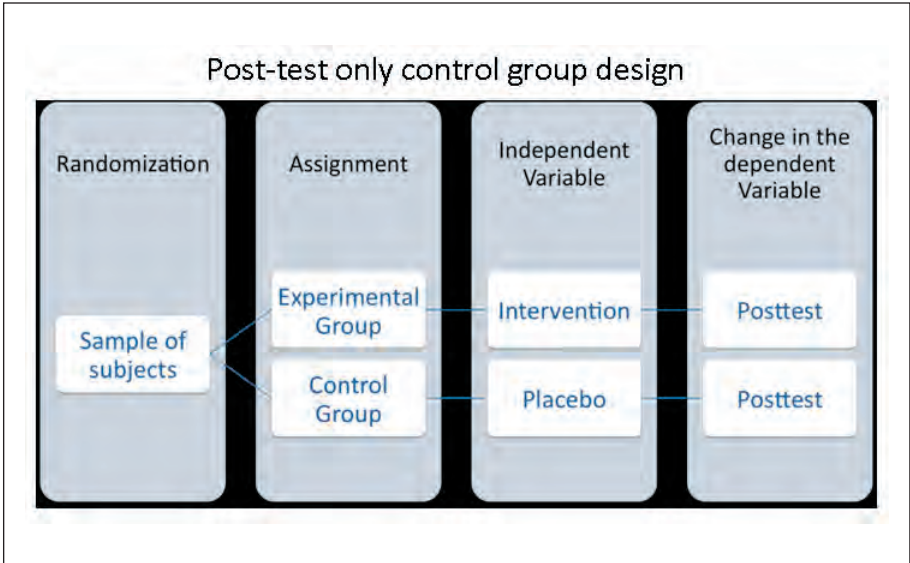


FIGURE 5. Experimental design lacking a pretest measure: post-test only control group design.

3.3.2 Statistical analysis

The statistical analysis appropriate for these types of design is the unpaired T-test in the case of interval-ratio data or, in the case of more than two groups the analysis of variance. In the case of ordinal data or not homogeneous distribution the Mann-Whitney U-test can be used to compare two groups or the Kruskal Wallis analysis of variance in the case of more than two groups.

3.3.3 Multifactor design

This type of design is widely used in rehabilitation because it allows researchers to analyze complex interactions between different factors.

For example, if we wish to study the effect of different rehabilitation approaches studying three groups: one where the independent variable should be home rehabilitation, another day hospital outpatient and the third hospital rehabilitation. In these conditions we would like to compare physiotherapy approach with occupational therapy approach. The aim is to verify the effect of different approaches in different settings. In this case we have two independent variables: exercise and the setting.

If there are two independent variables we define the design as a **two-way factorial analysis**. In the example we can answer the main

question related to the effect of the setting and the interaction effect related to the type of approach. The latter effect represents the main difference with the single factor design. For more complex research questions we can also use a **three way factorial design** where there are independent variables. Referring to the example we can increase the intensity of the exercises for further analysis. In this case we have multiple subgroups and to maintain an adequate impact of the study we should use a large sample size.

A further complexity could be added if we wish to know not only the generic setting but also the different type of setting such as public or private. Adding these subgroups the design is known as **nested design**.

3.3.4 Statistical analysis

In this type of design a two or three way analysis of the variance is most commonly used to analyse the main effect and the interaction between different factors. In the case of **nested design** we must apply very complex statistical analysis to be used in specific professional needs of analysis. (10).

3.4 Design for repeated measures

Some research questions are related to the effect of different interventions in the same group. In this case the subject is the control of himself and the measures are repeated after the application of different intervention (**within-subjects design**). In this case there is the lack of randomization between two or more groups. In order to maintain the paradigms of experimental design we should randomise the sequence of the application of interventions.

3.4.1 One way design

The simplest type of repeated measure design is a single factor design (one way) where a single group is exposed in sequence to different conditions of independent variable.

The sequence of the different interventions could be a bias. One solution of the **order effect** is to randomise the sequence of the application.

3.4.2 Cross-over design

When the repetition of an independent variable is two level (two time) to maintain the experimental design paradigm and overcome the potential bias we should use the **crossover design**. According to this design we create two different sequences A-B and B-A and they are randomly assigned. Sometimes the interventions have a cumulative effect and in this case we should include a wash out period in this sequence: A - washout - B.

When we use ordinal data the most appropriate analysis is with the Wilcoxon signed rank test.

3.4.3 Statistical Analysis

The appropriate analysis for one way design is the one way analysis of variance for repeated measures.

For cross-over design we can use a paired t-test or the two way analysis of variance with two repeated measures.

3.5 Multifactor design

Also with the repeated measures we can use more complex designs including two or three independent variables.

When we use both the between and within design we refer to the approach as a **mixed or split-plot design**.

4. QUASI-EXPERIMENTAL DESIGNS

In clinical settings, especially in rehabilitation settings, the paradigms of experimental design might not be feasible. The design is maintained but there is a lack of randomization or control group or both of them. The definition of **quasi-experimental design** is used when there is a lack of these criteria but the general design is to study the cause-effect.

In this type of design there is a decrease of the control and the research protocol needs to be well documented and possibly maintain as far as possible the blinding conditions.

The gold standard is the RCT but this design, taking into account the possible bias could also provide useful information.

4.1 Single group design

In this case the sequence is similar to the pretest-posttest experimental design. The difference is the lack of control group. The potential bias is the possibility that other variables, different from the independent variable, are involved in the improvement of the outcome. This study design could be useful when a previous study has already shown the efficacy of the treatment applied and we want to analyze the size of the effect better.

4.2 Repeated Measures Design Over Time

Also in this case there is the possibility to carry out a design with an independent variable taking the repeated measures over time. The possible bias is the lack of a control group. This design is appropriate when the time course of the disease is predictable (11).

4.2.1 Statistical analysis

The more appropriate statistical analysis is a paired T-test, in the case of ordinal data or small sample size a non parametric test such as the Wilcoxon signed ranks test can be used.

4.3 Time series design

This experimental design is useful for community of policy studies when we expect the intervention could change the trend of the measure-

ment over time. It could also be used for the single case study. The basic pattern is to measure the trend before the intervention and after with the repeated measurement at specific intervals. The multiple pretest and posttest measurements improve the control compared to the one sample pretest-posttest design.

4.3.1 Statistical analysis

The statistical analysis could be complex. There is not agreement on the methods to be applied. In some cases visualizing just the graphical trend is considered sufficient to demonstrate the effect. Other approaches are with the time-series analysis with multivariate methods.

4.4 Multiple group design

4.4.1 Non equivalent groups

The typical case is the multiple groups pretest-posttest design without randomization. The thread is the lack of the control of the potential bias of the different groups. A similar situation could be in the pretest-only control group design.

5. TYPES OF DESIGNS USEFUL FOR PRM INVESTIGATIONS

5.1 Observational studies

In a controlled experiment the researcher decides who gets assigned to which group. But there are many situations in which the researcher can just watch what happens.

Studies related to accidents or smoking are examples of observational studies. People are not usually willing to be randomized to smoke or have an accident just to participate in a study. Observational studies allow researchers to find evidence of association between a factor and a response. For example between smoking and lung cancer. The main goal of this type of experimental design is to find association between factors, but there may be hidden factors involved (e.g., other factors could make people smoke and also make them get sick). [N.B. Remember: association is not causation].

To reduce the effect of confounding factors in observational studies we have to make the control and the treatment groups as similar as possible, that is, we have to control for confounding factors. In the case of smoking, age and gender can be confounding factors. So the right thing to do is to compare subjects of the same age and gender who smoke and do not smoke.

The observational studies are part of the exploratory research and in part of the descriptive research.

Exploratory research can be carried out prospectively or retrospectively. Prospective research is more reliable than retrospective research. In the retrospective research the data collected are historical and it is difficult to document a temporal sequence of the experiment. Also the collection of

the data lacks a methodology and there is the possibility of missing data.

From a temporal point of view the research can be longitudinal when the researcher follows a cohort of subjects over time, with measurement at a defined time. Usually they involve large databases. One example is the Traumatic Brain Injury Model System (TBIMS) created by the National Institute on Disability and Rehabilitation Research (NIDRR) in 1987, showing the advantages of a well-coordinated rehabilitative system among various specialist centres (<http://www.tbindc.org/>) (12, 13).

Longitudinal studies have a potential problem for the internal validity because it is impossible to guarantee the compliance and the possible bias due to confounding factors.

Cross-sectional Research is used to explore a cohort of subjects at the same time. In this manner the problems of longitudinal variation are abolished but the threat is the historical effect, such as age that leads to a non homogeneous group of subjects (11).

In exploratory studies the researchers do not try to control or manipulate the variable but measure how they vary with respect to each other. The purpose is to describe the nature of existing relationships among variables.

Correlational studies are useful to discover the correlation between variables. A specific type of correlation is designed in the predictive correlational studies where the aim is to predict an outcome based on relationships between variables. An example is the analysis that predict the recovery in spinal cord injured patients that start the rehabilitation program with different delay periods following the lesional event (14).

5.1.1 Advantage of exploratory research

The exploratory studies are not able to demonstrate cause-effect relationships, however they play an important role in clinical research. The correlation analysis could be preliminary to an experimental design indicating the variables to be included.

Looking at the correlation could lead to several bias and misinterpretation of the complex nature of clinical phenomena. An apparent correlation between two variables could be influenced by latent variables that account for the real effect. In order to improve the validity of the conclusion it is important to use complex statistical analysis based in the multivariate methods.

5.1.2 Type of statistical analysis

In order to test the research hypothesis of the correlation a multivariate analysis is used. In the case of numerical variables the multivariate analysis is carried out with the regression model. A numerical dependent variable could be the gait velocity that could correlate to different independent variables (age, length of stay, delay of onset of rehabilitation program) that should give independent effect on gait velocity. With the multiple regression analysis it is possible to analyse the amount of the effect of these variables. It is important to look at the overall effect of these variables on the dependent variables. This effect is quantified with the R^2 .

A value of 0.50 means the variability of gait velocity is explained for 50% by the independent variables analysed. The lower the percentage the lower the overall effect (15).

In rehabilitation it is often useful to predict dichotomous outcomes, such as return to home after the rehabilitation program or achieving independence or not and analysing the factors that independently account for the different outcomes.

The general technique used to predict dichotomous outcomes is the logistic regression. In a cohort study we included 2626 subjects with severe brain injuries admitted to 52 Italian centres (GISCAR Study). We analyzed the factor that independently influence the dichotomous outcome (Discharged to go home or not). We found that the significant factors were: admitted with a vegetative condition, age and traumatic aetiology. This model seems to make sense but looking at the pseudo R^2 (the equivalent of R^2 in multiple regression) the value was 0.09 which means that the model explains only 9% of the causes of discharge to go home (personal data submitted for publication). The example underlines the importance of analysing not only in terms of significant variables but also the real effect of them. It is evident that in the example other factors contribute to explain the discharge to go home such as the type of house, the compliance of the family, and other social factors. This study suggests that in the case of a future RCT other independent variables should be collected.

5.1.3 Cohort study

A typical application of the observational models is the cohort study. In this case a group of patients is followed over time together. The advantage of the cohort study is the ability to determine the onset of the condition. A correlation between the exposure to risk factors and development of a particular outcome.

An example is a large Italian study GISEM where a cohort of consecutive patients with spinal cord injury was followed from the start of the rehabilitation program to the end of inpatient rehabilitation program. In this case the different analysis was possible selecting the subgroups of patients such as Spinal Cord Injuries vs Nontraumatic spinal cord lesions compared for specific outcome such as complications and length of stay (16).

5.1.4 Case-control study

The case-control study is used to verify the factors involved in determining a disorder. In order to obtain that usually the population is divided into two groups, one with a disorder (case) and another group without (control). Retrospectively one or more factors are analyzed to show whether there is a correlation with the development of the disorder.

The advantage of this type of study is that the patients are quite easy to gather. The case-control studies are very useful to study disorders that are relatively rare and for analyzing over a long latency period (11).

The problem with this study could come from a selection bias for both the classification of the subjects (case-control) and the assessment of exposure status.

In a recent review it has been estimated that 97% (83 out of 86 studies were mislabelled) were mislabelled case-control studies in rehabilitation journals. The issue is not simply a semantic problem but is likely to cause confusion in conducting, analyzing the study and in interpreting the findings (17).

The typical error was the confusion between the outcome and the exposure. The classification case-control should involve the outcome (to have or not have the disorder) and the independent variables should be the factors that could lead to the disorder. An appropriate classification is mandatory in order to avoid a selection bias. (18)

Another problem could be the **interviewer bias** when the interpretation of the interviewer is different for the cases compared to the control. A **recall bias** could be due to the different experiences to remember the history of exposure factors. In fact, the subjects with the disorder could be trained to pay attention to the factor influencing the health status better than the control (11).

For a correct methodological approach the case-controls are methodologically difficult to design and can be quite complex to analyze (18).

5.2 Single case design

They are quasi-experimental research designs that involve assessing change in a dependent variable on a single research subject. Single-case experimental designs or Single Subject Research Designs are designs that can be applied when the sample size is one or when a number of individuals are considered as one group. Single subject designs are often considered the design of choice when measuring behavioural change or when performing behavioural modification. Rather than comparing groups of subjects, this design relies on the comparison of treatment effects on a single subject or group of single subjects.

In single-subject designs, each participant serves as her or his own control, similar to a time-series design. Basically, the participant is exposed to a non-treatment and a treatment phase and performance is measured during each phase.

An important aspect of this type of study is the gathering of pre-test information, often called a baseline measurement. It is important to measure the dependent variable or behaviour prior to administering any treatment. Without this information, it is difficult, and likely impossible to determine if any change has occurred. Also often associated with this design are periods of measurement to determine not only a change but also the degree of change through the process of behavioural modification.

5.2.1 A-B design

The A-B Design is a two-phase design consisting of a no-intervention baseline phase (A) and an intervention phase (B). It allows for evaluation of pre-intervention and intervention problem status. It is a quasi-experimental design. The existence of a no intervention baseline allows for the establishment of a relationship between intervention and outcome. If a change in the dependent measure occurs at the onset of the intervention, a relationship between the independent and dependent variable has been established.

Since the presentation of the intervention is under experimenter control, the temporality (asymmetry) issue is covered. However, this design is susceptible to uncontrolled influences of extraneous variables (the no spuriousness issue in causal inference), especially the history threat to internal validity. Because this is a repeated measures design, there is some control for the threats of maturation, testing, instrumentation, and statistical regression. If these effects were operating, there is a reasonable chance that they would show up during baseline observations. The remaining threats would be irrelevant to a single subject design.

5.2.2 A-B-A-B Design

In order to improve internal validity of the design withdrawal models could be useful with a second baseline phase (A_2) and possibly a second exposure phase (B_2).

An example of A-B-A design has been a telerehabilitation study carried out to show the effect of a computerized system for upper limb rehabilitation taken at home. The study was a double design study: a RCT with the randomization in two groups (usual rehabilitation vs telerehabilitation) (19) and single case design for each patient exposed to the treatment with a baseline measure of function of the upper limb in stroke patients (A_1), after a month of exposure to the telerehabilitation system (B), a follow-up after a month without telerehabilitation (A_2). With a graphical representation we can verify the effect (Fig 6).

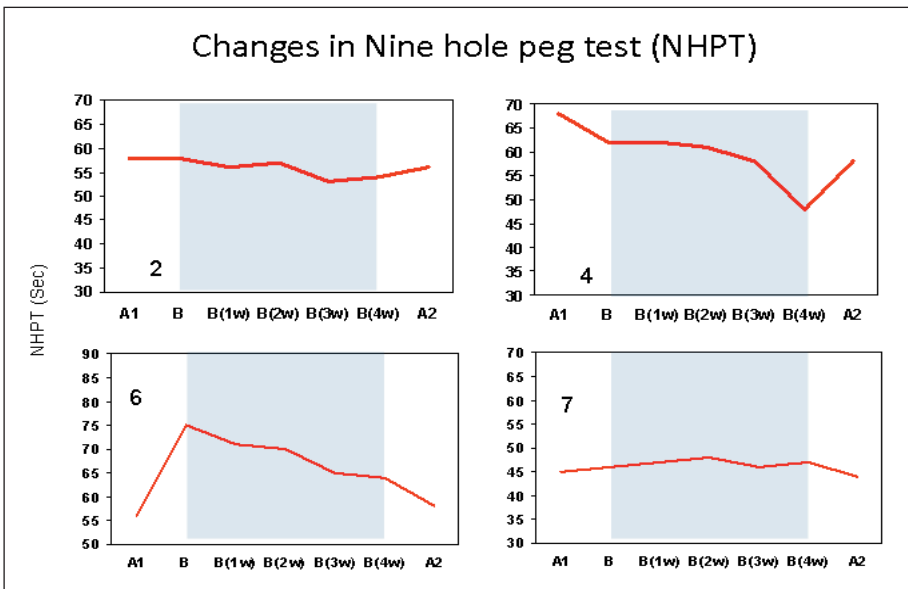


FIGURE 6. Practical application of a single case design ABA. The protocol is the study of the effect of telerehabilitation in single patients with moderate stroke. The measurement is the time to perform the nine hole peg test. The measurement is applied at A1, one month before the exposure to the rehabilitation program. B1 (w1....), the measurement during the month of exposure (one each week). A2 one month after the end of exposure. In each patient we can visually assess the effect of the treatment (personal data).

5.2.3 *Multiple treatment design*

The multiple baseline design allows for evaluation across clients, situations or problems. It is a true experimental design in that it allows for causal inference. It is extremely useful for evaluating situations where an intervention would be likely to bring about enduring changes in the dependent variable.

5.3 Type of analysis

In single case studies the aim is to demonstrate if an individual has changed over the time period. In order to show these differences different methods have been proposed considering that the usual statistical analysis is not applicable (20-22).

The common techniques are (15):

- Celebration line analysis
- Level, Trend and Slope
- Two standard deviation band analysis
- C statistics

6. APPLYING THE EXPERIMENTAL DESIGN TO REHABILITATION

According to contemporary medicine, the Physical and Rehabilitation Medicine ought to demonstrate that its practices are efficacious, effective and with a good cost-benefit ratio. For carrying out that challenge the researchers can utilize a variety of well tested research designs that includes, but is by no means limited to RCTs. (6).

The randomized clinical trials require in general, preliminary investigation to identify appropriate population and define an adequate sample size according to the power needed for the study. The clinical research needs to ensure appropriate prognostic balance between experimental groups or to adjust the minor imbalances statistically (23).

To cope with the methodological issue typically a multicenter clinic networks to recruit an adequate number of representative participants required. Rehabilitation researchers have specific problems compared with general clinical research. Disabled subjects receive a complex combination of treatments with a highly individualized goal.

Rehabilitation research often requires measurement of multiple outcomes and consequently deals with statistical problems. The need for studying a greater number of outcomes, in turn, requires understanding factors predictive of the prognosis for each of those outcomes.

The design and conduct of rehabilitation RCTs create some discrepancy between the ideal features of such studies and the realities of clinical investigation (Tab. 3).

TABLE 3. Methodological problems in rehabilitation research (6)

▪ Difficulties in blinding the administrators of interventions or the recipients.
▪ Candidate participants' resistance to being randomly assigned to experimental or control groups.
▪ The unacceptability, ethically or otherwise, of using control conditions that withhold or delay treatment.
▪ The extreme complexity of some interventions that makes it difficult to monitor the fidelity with which they are administered.
▪ An insufficiency of eligible participants in any one setting, necessitating difficult-to-administer, multisite trials.
▪ The relatively lengthy follow-up interval of some rehabilitation trials that makes them vulnerable to participant attrition.
▪ The unpreparedness of some federal agencies to fund the needed trials, especially large-scale ones.
▪ Other challenges pertain to interpreting the studies' findings. Included are the following: <ul style="list-style-type: none"> – Confounds between an intervention's effects and aspects of natural recovery processes. – Confounds between an intervention's effects and uncontrolled factors occurring during the intervention or follow-up periods. – Marked variability among participants on pre-existing characteristics that relate to the outcome variables, making it difficult to identify effects attributable to the intervention.
▪ The 'active ingredients problem' in interpreting the outcomes of studies with complex interventions (i.e., determining the relative causal contribution of interventions' various components).

The problematic aspects of RCTs are directly related to the question of their optimal role in rehabilitation outcomes research. There are individuals who want to decide about a treatment's effectiveness only on the base of RCTs. On the other hand other scientists insist that the only worthy outcomes research design is the RCT.

Perhaps most challenging is the fact that multifaceted non pharmacologic interactive treatments are more difficult to specify than pharmacologic or surgical interventions. Finally, in the current state of the field, a strong theoretical base to guide treatment research and to guide the design of comparison treatments is lacking.

The experimental design is the gold standard of the clinical research but, adapting the research to the possibility of the complexity of the rehabilitation, we have a series of instruments such as single case design and observational studies to be used to investigate the factors correlated with the recovery and with the outcome of the disabled people. Sometimes we should reach a compromise preserving the correct research methodology and the applicability in real conditions. We should avoid to renouncing carrying out the research because of the complexity of the variables involved and the apparent impossibility to use an RCT design.

Finally, we should draw attention to the choice of adequate outcome measurements (24) keeping in mind that the main outcome in rehabilitation is the return to participation in daily life with an as good as possible quality of life. Demonstrating the effect only in the impairment could be useless in terms of demonstration of real efficacy of rehabilitation program. The new concept of disability determined by the International

Classification of Functioning and Disability (ICF) of the World Health Organization (WHO) (25) could be a reference framework for the development of future research (26). Researchers in rehabilitation should start from the research question and then the optimal design should be selected from a wide repertoire of designs that provide the best evidence for the various purposes.

Researchers should not select necessarily RCT without ensuring that it is the appropriate design to answer the question posed according to the ICF model. The researchers should capitalize, rather than eliminate, natural variation in the populations of interest and using observational design on account of the multifactorial nature of disability and functioning (27).

On the other hand the improvements in methodology areas in PRM from psychometrics to meta-analysis, and from proper reporting of research to inputting missing data, are continuous, and even those investigators who have had extensive training in research methods and statistics as part of graduate and postgraduate study may need an expert to assist them when they are outside their areas of expertise (28).

REFERENCES

1. Fisher RA. The Design of Experiments. 1935.
2. Fregni FM, Imamura HF, Chien HL, Lew P, Boggio TJ, Kaptchuk et al. Challenges and recommendations for placebo controls in randomized trials in physical and rehabilitation medicine: a report of the international placebo symposium working group. *Am J Phys Med Rehabil* 2010; 89: 160-72.
3. Cochrane A. Effectiveness and Efficiency: Random Reflections on Health Services. London: Nuffield Provincial Hospitals Trust, 1972.
4. Keith R. The role of treatment theory. In: M. Fuhrer (Ed). *Assessing Medical Rehabilitation Practices: The Promise of Outcomes Research*. Paul H. Brookes Publishing: Baltimore. 1997; 257-74.
5. Piantadosi S. *Clinical Trials: A Methodologic Perspective*. New York: Wiley, 1997.
6. Fuhrer MJ. Overview of clinical trials in medical rehabilitation: impetuses, challenges, and needed future directions. *Am J Phys Med Rehabil* 2003; 82(10 Suppl): S8-15.
7. Leeuw M, Goossens ME, de Vet HC, Vlaeyen JW. The fidelity of treatment delivery can be assessed in treatment outcome studies: a successful illustration from behavioral medicine. *J Clin Epidemiol* 2009; 62: 81-90.
8. Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMC Med* 2010; 8: 18.
9. Antes G. The new CONSORT statement. *BMJ* 2010; 340: c1432.
10. Keppel G. *Design and Analysis: A Researcher's Handbook*. 4th ed. Englewood Cliff, NJ: Prentice Hall, 2004.
11. Portney LG, Watkins MP. *Foundations of Clinical Research: Applications to Practice*. 3th ed. Upper Saddle River, New Jersey: Pearson: Practice Hill, 2009.
12. Bushnik T, Hanks RA, Kreutzer J, Rosenthal M. Etiology of traumatic brain injury: characterization of differential outcomes up to 1 year postinjury. *Arch Phys Med Rehabil* 2003; 84: 255-62.
13. Bushnik T. Introduction: the Traumatic Brain Injury Model Systems of Care. *Arch Phys Med Rehabil* 2003; 84: 151-2.
14. Celani MG, Spizzichino L, Ricci S, Zampolini M, Franceschini M. Spinal cord injury in Italy: A multicenter retrospective study. *Arch Phys Med Rehabil* 2001; 82: 589-96.

15. Domholdt E. *Rehabilitation Research: Principles and Applications*. 3th ed. Saunders, 2005.
16. Pagliacci MC, Celani MG, Spizzichino L, Zampolini M, Aito S, Citterio A, et al. Spinal cord lesion management in Italy: a 2-year survey. *Spinal Cord* 2003; 41: 620-8.
17. Mayo NE, Goldberg MS. When is a case-control study not a case-control study? *J Rehabil Med* 2009; 41: 209-16.
18. Mayo NE, Goldberg MS. When is a case-control study a case-control study? *J Rehabil Med* 2009; 41: 217-22.
19. Huijgen BC, Vollenbroek-Hutten MM, Zampolini M, Opisso E, Bernabeu M, Van Nieuwenhoven J, et al. Feasibility of a home-based telerehabilitation system compared to usual care: arm/hand function in patients with stroke, traumatic brain injury and multiple sclerosis. *J Telemed Telecare* 2008; 14: 249-56.
20. Ottenbacher KJ. Analysis of data in idiographic research. Issues and methods. *Am J Phys Med Rehabil* 1992; 71: 202-8.
21. Nourbakhsh MR, Ottenbacher KJ. The statistical analysis of single-subject data: a comparative examination. *Phys Ther* 1994; 74: 768-76.
22. Bobrovitz CD, Ottenbacher KJ. Comparison of visual inspection and statistical analysis of single-subject data in rehabilitation research. *Am J Phys Med Rehabil* 1998; 77: 94-102.
23. Whyte J. Clinical trials in rehabilitation: what are the obstacles? *Am J Phys Med Rehabil* 2003; 82(10 Suppl): S16-21.
24. Bagiella E. Clinical trials in rehabilitation: single or multiple outcomes? *Arch Phys Med Rehabil* 2009; 90(11 Suppl): S17-21.
25. Stucki G, Ewert T, Cieza A. Value and application of the ICF in rehabilitation medicine. *Disabil Rehabil* 2002; 24: 932-8.
26. Stucki G, Reinhardt JD, Grimby G, Melvin J. Developing research capacity in human functioning and rehabilitation research from the comprehensive perspective based on the ICF-model. *Eur J Phys Rehabil Med* 2008; 44: 343-51.
27. Bartlett DJ, Macnab J, Macarthur C, Mandich A, Magill-Evans J, Young NL, et al. Advancing rehabilitation research: an interactionist perspective to guide question and design. *Disabil Rehabil* 2006; 28: 1169-76.
28. Dijkers M. Case-control studies false and true: mislabeled study designs in rehabilitation research, and a master class in case-control study design. *J Rehabil Med* 2009; 41: 289-90.

SOO JEONG HAN¹
MARTA IMAMURA²
FELIPE FREGNI³

NON-PHARMACOLOGICAL PLACEBO CONTROLLED RANDOMIZED TRIALS AND THE ROLE OF PRAGMATIC TRIALS IN PRM

¹ Department of Physical Medicine
& Rehabilitation, Spaulding
Rehabilitation Hospital,
Harvard Medical School.

Department of Physical Medicine
& Rehabilitation, Ewha Womans
University, School of Medicine

² Collaborative Professor. Division
of Physical Medicine.
Department of Orthopaedics
and Traumatology, University
of São Paulo School of Medicine

³ Department of Physical Medicine
& Rehabilitation & Department
of Neurology, Spaulding
Rehabilitation Hospital,
Massachusetts General Hospital
and Beth Israel Deaconess
Medical Center, Harvard
Medical School

INTRODUCTION: THE STATE OF CLINICAL RESEARCH IN PRM

Placebo randomized controlled clinical trials are the gold standard design in clinical research for the investigation of a new approach. Although this design is not always possible to be implemented due to ethical issues, challenges in blinding and costs, randomized controlled trials (RCTs) appear to be a progressive trend in the Physical and Rehabilitation Medicine (PRM) literature (1). Because PRM is also a specialty that uses several non-pharmacological interventions among its therapeutic tools such as acupuncture, physical training, devices (such as noninvasive brain stimulation), it is often complex to design RCTs (1). In this chapter, we will highlight most common difficulties in designing and conducting such clinical trials in PRM, discussing alternative methods such as pragmatic trials.

Clinical observation (naturalistic studies), uncontrolled observational studies, or clinical trials comparing two active interventions have been frequently used in medical trials; however, although useful,

these designs may introduce some risk of bias that can be avoided by placebo controlled trials. One of the issues is whether these designs provide sufficient proof of efficacy for novel interventions. Although recent meta-analyses have shown that observational studies can produce similar effect estimates as compared to randomized clinical trials (2, 3), observational studies are also not easy to be conducted and the results might vary significantly according to the analytic method used during data analysis. It is important to keep in mind, however, that the use of placebo or sham devices in placebo RCTs might not be practical or possible in some PRM clinical scenarios. Therefore, the understanding of placebo use in clinical research and the development of reliable placebos or sham interventions are necessary steps to improve the level of evidence in PRM. The ultimate goal is to improve patient care in PRM.

CURRENT STATE OF PLACEBO-CONTROLLED CLINICAL TRIALS IN PRM

In the recent years, the interest in randomized controlled studies in PRM has increased. In fact, the growing interest in controlled trials can be observed in PUBMED. There is an increase in the number of placebo and sham-controlled studies published since 1964 in rehabilitation, physical medicine, and physical therapy journals (1). In the last decade (1999-2008), there were 8680 placebo, or sham-controlled trials in PRM, which indicated an increase of 3743 from the prior decade (1989-1998). There is also an increasing interest in trials evaluating therapeutic effects of complementary medicine such as acupuncture and physical modalities including but not limited to electrical stimulation and application of heat and cold. The increase in the number of placebo-controlled studies in PRM in the last decade is a result of several factors:

1. The evidence-based medicine movement that started in the early 1990s, in which controlled clinical trials are used as best evidence to support novel interventions to improve the level of quality care.
2. An increased interest in rehabilitation modalities, with emphasis on conditions usually treated by PRM physicians, such as chronic diseases encountered in the aging population (e.g., arthritis and chronic pain).
3. Reimbursement decisions usually dependent on a high level of evidence. Although there is a clear increase in the number of placebo-controlled studies in PRM, there is still a number of important challenges that are responsible for the delay in the development of high-quality evidence in PRM specialty.

The Need for Sham or Placebo-Controlled Trials in PRM

In PRM, the most commonly used control group includes interventions as simple as non-participation (waiting list), usual care, or other

interventions. These control interventions can be associated with non-specific physiologic effects that can affect the results of a given trial. This is especially noted when patients and care providers are not blinded to these control interventions. Patients' expectation may also significantly interfere with treatment results (4), and nocebo effects might also play an important role (5).

Placebo-controlled trials ensure blinding and reduce the risk of potential biases that can interfere with the observed outcomes. There are two major types of biases commonly seen in studies: performance bias and detection bias. Performance bias is defined as systematic difference between groups in the care that is provided because of the exposure to factors other than the interventions of interest. Detection bias is defined as systematic differences between groups in how the outcomes are determined. These types of biases can be minimized by using adequate placebos (1, 2, 3, 6).

These biases may pose a special challenge because most PRM interventions, by their nature, are associated with the building of relationships among the patient, rehabilitation team members, family members, and care providers. In this scenario, well-designed clinical trials become critical for clinical trials as the ultimate goal of any clinical trial comparing two interventions of the effect of one intervention alone is to show that the observed differences are due to the intervention being tested and cannot be explained by other factors such as placebo effect, confounding or other biases.

Challenges of Placebo Use in PRM Clinical Trials

Recently, a working group was convened and discussed important challenges of use of placebo in PRM trials. Due to the relevance for RCTs, we discuss and summarize here the ten important challenges that were identified (1).

(i) *Development of Placebo and Sham Devices*. In most countries, devices have different and often less restrictive regulatory requirements than drugs, and development of effective sham methods is not sufficiently supported. (ii) *Lack of Standards in PRM Therapies*. Because many PRM therapies are difficult to be standardized, it is even more difficult to design a "standard" placebo therapy in these situations. (iii) *Treatment Heterogeneity Resulting from Therapist's Skill Differences*. To control for skills and levels of experience, it would be necessary to conduct multicenter studies with various levels of skills and experiences and perform multivariate analyses to adjust for these variables. In this scenario, a large number of patients would be necessary, increasing the difficulties to conduct such studies. (iv) *Issues with Adequate Masking*. If blinding is impossible, it does not make sense to design a placebo-controlled trial; however, alternative methods to analyze bias should be carried out. (v) *Personal Interaction Between Therapist and Patient*. In certain circumstances, the specific

and unspecific aspects of the intervention cannot be detangled, decreasing the chances of designing an effective placebo intervention. (vi) *Personal Beliefs, Previous Experience, and Motivation*. Because expectation plays a critical role for the effects of placebo, the placebo effect might be elevated in some treatments in PRM and therefore difficult to control for, such as in occupational therapy, home exercises, and cognitive-behavioral interventions (1, 7, 8, 9). (vii) *Small Effect Sizes*. Some of the PRM interventions induce small improvements (effect sizes), therefore studies to prove efficacy require large sample sizes, which implies increased costs. (viii) *Long Follow-Up*. Because many PRM interventions treat chronic conditions, the studies involve long follow-up periods which increase the costs and the complexity of the study. This makes the use of placebo more difficult because of ethical concerns of a long period of exposure to a placebo intervention. (ix) *Lack of Training to Conduct Clinical Research*. Many PRM practitioners lack appropriate training in clinical research; therefore, the importance and methods of placebo use are sometimes not appreciated. (x) *Use of Medical Devices*. The extensive use of medical devices in PRM may have an important impact on placebo use. This is due to the notion that sham devices would have a higher placebo response than placebo pills (10). However it is still debatable whether a sham device induces a larger response (11).

General Recommendations to Improve the Quality of Placebo and Sham-Controlled Clinical Trials in PRM

Some general recommendations may provide insights on how to improve the necessary non pharmacological interventions employed in PRM practice. We discuss some of them in this section.

First, it is important to develop a better understanding of placebos. It is useful to investigate the biological mechanisms of placebo effects to plan better placebo controlled interventions. In addition, studies should be conducted to find predictors for placebo response as to try to control for this effect at some extent. In fact, some of these studies might be based on secondary analyses of the placebo arm of previous RCTs in the field of PRM. In some cases, one placebo may have a larger effect than another. It is important, therefore, to learn more about the effect sizes of different placebos before designing a given trial.

Second, there are specific challenges particular to the population seen in PRM. A placebo treatment needs to be considered for a longer period of time. In addition, it is important that the researcher is aware of small effect sizes to detect a meaningful difference because improvement is less dramatic than with acute conditions. Also, patients in PRM trials might have had previous experiences with the intervention under study (such as receiving acupuncture treatment before participating in a new acupuncture trial) and therefore might have a different expectation regarding treatments that will influence placebo response.

Third, there are also challenges particular to the type of interventions commonly used in PRM. Efforts should be made to try to reduce the variability of interventions both in the real and in the placebo groups. The variability might be as a result of the (i) amount of time spent in each session, (ii) number and intensity of sessions, (iii) too many providers, (iv) different settings (private clinic, academic hospital, inpatient, outpatient, etc.), (v) verbal and nonverbal clues, (vi) educational materials or packages of information, and (vii) rewards or reimbursements. Therefore, controlling for these factors is recommended to decrease variability. When the investigation of the multidisciplinary treatments is needed, some alternatives can be considered: (i) consider a waiting list control group (when placebo and real interventions are similar); (ii) consider virtual reality in situations where designing a placebo is difficult and also to avoid provider-patient interaction, and (iii) use structural equivalence—a method useful in psychotherapy trials where the experimental and placebo groups have similar degree of therapeutic contact.

A better understanding of various mechanisms and types of exercises is necessary to allow the development of a standard sham exercise therapy. There are challenges for the development of sham devices. Having partnerships with engineers and with industry to develop valid and reliable sham devices could be helpful. There are also challenges for the development of sham injection techniques. Study designs should allow the comparisons of two variables (e.g., drug and the procedure). Use of subjective outcomes should be reduced, when possible. Some conditions in PRM such as chronic pain have subjective self-reported outcomes as the main end points. The use of functional outcomes or surrogate markers should be considered in these situations.

Fourth, bioethical considerations should also be considered. Beliefs regarding treatments vary across different cultures. This is important for global clinical trials. Therefore, it is critical to respect the specific cultural beliefs. In addition, patients need to be asked before they are randomized, and researchers need to be aware of potential biases of including subjects who have a strong belief in favor of one intervention. Also, patients' rights cannot be jeopardized by the scientific interest. One important issue when considering the ethics is the concept of equipoise (when there is uncertainty regarding the comparative therapeutic merits of two treatments) that is not usually appreciated by investigators.

THE ROLE OF RANDOMIZATION IN RANDOMIZED CONTROLLED TRIALS

A scientifically valid comparison between two treatment groups depends on the groups being alike as much as possible, with the only exception being the specific treatments under investigation. Without such standardization, one group of patients may be treated differently than another, and the, observed result would be biased rather than serve as a

valid comparison of the treatments. The best way to achieve such a balance is by the use of randomization in which a mechanism determines the treatment assignment by chance. Randomization ensures therefore that a specific treatment assignment is not known to either the clinician or the patient. The primary benefit of randomization is that it will eliminate both conscious and unconscious bias associated with the selection of a treatment for a given patient.

Although the majority of clinical investigators today are convinced of the benefits of randomization, some problematic issues exist. The action of randomization may interfere with the physician-patient relationship. In order to conduct an RCT, clinicians do not know if the studied intervention is best for a given patient, which can potentially compromise the physician-patient relationship. Furthermore, from an ethical perspective, a clinician should believe that these therapies are equivalent with respect to potential patient benefit.

Although randomization does not guarantee identically balanced treatment groups, it will tend to produce treatment groups that are similar on average. Additional protection against a possible imbalance, however, is preferred especially in small clinical trials. In common clinical research settings the difference in effect between treatments is usually small relative to the effect of prognostic factors, such as extent of disease and a patient's performance status. A concept useful for clinical trial design and analysis is that one is trying to detect a soft signal in a noisy environment. The soft signal is the effect of treatment and the noisy environment is patient variability caused by prognostic factors, referral patterns, adherence to therapy, or other factors. Careful study design can help improve the signal to noise ratio, which thereby more readily exposes any true difference in treatments.

Additional protection against a possible imbalance is easily obtained by the use of a stratified randomization. In stratification, patients are formed into risk groups (strata) based on one or more prognostic factors, and a separate randomization is conducted for each strata. When the treatment assignment groups are then summed over the various strata, the end result is a forced balance of these overall treatment groups according to the factors used to form the strata. Use of stratified randomization should be viewed as an insurance policy against a potential imbalance, and, because it has virtually no cost (ie, no increase in number of patients needed or additional administrative complexity), it can be routinely used in RCTs (12); however stratified randomization has its limitation especially for small studies where the number of strata needs to be small. In this case it is recommended to use no more than one risk factor in the stratified randomization.

IMPORTANCE OF NON-PHARMACOLOGICAL RCTs IN PRM

In the arena of non pharmacological interventions, such as with the use of medical devices, however, blinding and placebo (or sham) control

may sometimes be very difficult or even impossible to implement. Randomization, although arguably always possible in theory, is sometimes deemed as impractical for various reasons, including enrollment rate, investigator participation, potential noncompliance and ethical issues. Therefore, it is not uncommon for clinical studies of non pharmacological interventions usually depart from the paradigm of randomized, double-blinded, and well-controlled design that has become standard in the realm of drug regulation (13, 14).

Consequently, nonrandomized clinical studies play a substantial role in the evaluation of the benefits of non pharmacological interventions. The lack of randomization poses some challenges in the design of a nonrandomized study with a concurrent control arm. For example, without randomization it is difficult to predict the enrollment rates of the treatment and the control arms. Consequently, sample sizes from the intervention and the control arms are not reached at the same time. A best guess for the ratio of enrollment rates between the two study arms should be specified at the design stage, but there is no guarantee that it will be achieved. On the other hand, it is usually desirable that both arms finish enrollment at approximately the same time in order to avoid the artifacts of temporal bias. Another challenge of nonrandomized trials is the balance of baseline demographics and risk factors in both groups, without the help of randomization. Usually, the inclusion and exclusion criteria for the treatment and control arms are defined at the design phase of the study, but do not guarantee that imbalances will not occur.

In nonrandomized, confirmatory, medical device studies, direct treatment comparison without considering covariates is usually not sufficient to provide convincing evidence of safety and effectiveness, and a covariate-adjusted data analysis, using frequentist or bayesian paradigm (15), should be conducted. Details of the covariate-adjusted data analysis need to be specified a priori in the protocol, because any post-hoc data analysis could only provide supporting but not confirmatory evidence. In particular, it is important to also specify a priory as many clinically relevant baseline covariates as possible that will be collected in the study and used in the data analysis. Also, sensitivity analysis should be planned to estimate the effects of any unobserved covariates. It is crucial to select comparable patient populations, since no statistical method works well when there is severe imbalance in baseline covariates between the two treatment groups. However, it is sometimes impossible to assess in advance whether the patient population treated with a new device will be comparable to that represented by the control intervention, which is a challenge for nonrandomized studies. Consequently, a nonrandomized study, with all the complex issues in its design and analysis, may be more burdensome than a randomized trial. Investigators should be aware that a priori specification is essential in a regulatory setting and that direct treatment comparisons without taking covariates into account is usually not adequate.

THE ROLE OF PRAGMATIC TRIALS IN PRM

In the previous sections we discussed the challenges for conducting placebo controlled clinical trials in addition to the challenges to conduct non-pharmacological clinical trials. One potential alternative to address some of the aforementioned challenges is designing a pragmatic trial. In fact, aware of the limited applicability of the results derived from the artificial, 'laboratory' environment of a clinical trial, two French statisticians, Schwartz and Lellouch (16), proposed the interesting concept of differentiating them from trials that inform clinical health. They proposed a distinction between trials aimed at confirming a physiological hypothesis, precisely specified as a causal relationship between administration of an intervention and some physiological outcome (which they called 'explanatory') and the entirely different group of trials aimed at informing a clinical, health service or policy decision, where this decision involves the choice between two or more interventions (called 'pragmatic') (16, 17).

The greatest strength of pragmatic trials is that they can deliver evidence of effectiveness in everyday clinical practice (18). Such trials may be appropriate for complex interventions. They are especially useful when the use of a placebo control is problematic in terms of separating the specific from the non-specific effects. A good example to illustrate this scenario is when an investigator expects the positive synergistic effect between the non pharmacological intervention and the physician patient therapeutic relationship. The evaluation of the economic impact fits well with the pragmatic type of trial design. Pragmatic trials generate valuable evidence to health care purchasers, providers or patients who are making choices about treatment, whether on behalf of others or for themselves. The results help us understand more about the acceptability of the intervention to patients, and have potentially greater impact on decision-making, referral patterns and clinical guidelines. Hospital outpatient chiropractic care, for example, was proven successful in a pragmatic trial for low back pain (19). Because practitioners can have some flexibility in how they treat patients, they may be willing to be involved in a pragmatic trial. Likewise, patients are more likely to volunteer since they are not offered a "placebo" treatment.

Some features of explanatory and pragmatic trials can be presented archetypically as polar opposites. However, many randomized controlled trials lie somewhere along the spectrum between these two designs. In fact, if a pragmatic trial has a more tightly specified treatment protocol than allowing practitioners a free hand to treat normally, then the design is no longer fully pragmatic. For drug trials, the explanatory approach is advocated on the basis that new medication must be tested for efficacy prior to being available for general release. However, this argument does not hold for complementary medicine, which now is in widespread use. An estimated 50,000 complementary practitioners, for example are currently working in the UK providing treatment to around 5 million patients a year (20). As the House of Lords identified (21), this practice has now

become a public health issue because the putative benefits of these therapies are still unknown. One may argue that our research efforts should now focus on the development of meaningful evidence concerning routine care. Only after the overall beneficial effect is established, the effect of individual components of treatment should be identified. The shift towards evaluating practice in a pragmatic way is reflected in two large scale trials of acupuncture, one for back pain (22) and one for migraine (23), both funded by the UK government.

Whereas the explanatory trial must remain the foundation for assessing efficacy and for the extension of licensing into all age-groups, the pragmatic trial has much to offer to the prescribing clinician as it is able to address questions that the prescribing clinician needs to have answered. Future challenges will include the secure and independent funding of such 'real world' studies and in ensuring their generalizability to the patients and clinical scenarios typically seen in primary care and in hospital practice (24).

Both explanatory and pragmatic trials play an important role in the evaluation of health care interventions. However, they answer different research questions. Pragmatic trials are useful in answering questions about how effective a therapy is when compared to some standard or accepted treatment. Pragmatic trials also overcome the limitations of the complex packages of care in explanatory trials. Results of pragmatic trials can be generalized to wider clinical settings, providing evidence of the benefits of interventions used as alternatives or adjuncts to conventional interventions. Pragmatic trials facilitate the decision-making process about whether certain therapeutic modalities should be utilized more widely (25). Although this design has many advantages as shown here, it does not substitute explanatory clinical trials as any intervention has to undergo all the stages of clinical trial development and the art of clinical research is to identify when (and also how) each of the different designs should take place as to fully develop a medical intervention into clinical practice.

ACKNOWLEDGEMENTS. We would like to thank Kayleen Weaver, BA for helping to edit and proofread this chapter.

REFERENCES

1. Fregni F, Imamura M, Chien HF, Lew HL, Boggio P, Kaptchuk TJ, et al. Challenges and recommendations for placebo controls in randomized trials in physical and rehabilitation medicine: A report of the International Placebo Symposium Working Group. *Am J Phys Med Rehabil* 2010; 89: 160-172.
2. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med* 2000; 342: 1878-1886.
3. Bonovas S, Filioussi K, Tsavaris N, Sitaras NM. Use of statins and breast cancer: a meta-analysis of seven randomized clinical trials and nine observational studies. *J Clin Oncol* 2005; 23: 8606-8612.

4. Kalauokalani D, Cherkin DC, Sherman KJ, Koepsell TD, Deyo RA. Lessons from a trial of acupuncture and massage for low back pain: Patient expectations and treatment effects. *Spine* 2001; 26: 1418-1424.
5. Turner JA, Jensen MP, Warms CA, Cardenas DD. Blinding effectiveness and association of pretreatment expectations with pain improvement in a double-blind randomized controlled trial. *Pain* 2002; 99: 91-99.
6. Stewart LA, Parmar MK, Bias in the analysis and reporting of randomized controlled trials. *Int J Technol Assess Health Care* 1996; 12: 264-275.
7. Targino RA, Imamura M, Kaziyaama HH, Souza LP, Hsing WT, Furlan AD, et al. A randomized controlled trial of acupuncture added to usual treatment for fibromyalgia. *J Rehabil Med* 2008; 40: 582-588.
8. Kirsch I. Specifying nonspecifics: Psychological mechanisms of placebo effects. In: Harrington A, editors. *The Placebo Effect. An Interdisciplinary Exploration*. Cambridge: Harvard University Press; 1997.
9. Schwartz NA, Turturro MA, Istvan DJ, Larkin GL. Patients' perceptions of route of non-steroidal anti-inflammatory drug administration and its effect on analgesia. *Acad Emerg Med* 2000; 7: 857-861.
10. Walach H. The efficacy paradox in randomized controlled trials of CAM and elsewhere: Beware of the placebo trap. *J Altern Complement Med* 2001; 7: 213-218.
11. Brunoni AR, Lopes M, Kaptschuk TJ, Fregni F. Placebo response of non-pharmacological and pharmacological trials in major depression: a systematic review and meta-analysis. *PLoS One*. 2009;4(3):e4824.
12. Stanley K. Design of Randomized Controlled Trials. *Circulation* 2007; 115: 1164-1169.
13. Campbell G, Yue LQ, Pennello G, Barrick MK. (2003). Medical devices. In: Chow SC, editors. *Encyclopedia of Biopharmaceutical Statistics*. New York: Marcel Dekker; 2003, p. 574-585.
14. Campbell G, Irony T, Lao CS, Li H, Pennello G, Vishnuvajjala RL, Yue LQ. Medical devices. In: Armitage P, Colton T, editors. *Encyclopedia of Biostatistics*. John Wiley & Sons Ltd; 2005, p. 3119-3137.
15. US Food and Drug Administration (2006). Draft Guidance for Industry and FDA Staff - Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials. Rockville, MD: Center for Devices and Radiological Health, Division of Biostatistics, Office of Surveillance and Biostatistics.
16. Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. *J Chronic Dis* 1967, 20: 637-648.
17. Zwarenstein M, Treweek S, Gagnier J, Altman DG, Maclure M, Tunis S, et al. Improving the reporting of pragmatic trials: an extension of the CONSORT Statement. *BMJ* 2008, 337: a2390.
18. Medical Research Council (MRC). A framework for development and evaluation of RCTs for complex interventions to improve health. April 2000.
19. Meade T, Dyer S, Browne W, Townsend J, Frank A. Low back pain of mechanical origin: randomised comparison of chiropractic and hospital outpatient treatment. *Br Med J* 1990; 300: 1431-1437.
20. Budd S, Mills S. Professional organization of complementary and alternative medicine in the United Kingdom. A second report to the Department of Health. UK: University of Exeter; 2000.
21. House of Lords Select Committee on Science and Technology. 2000 Report on complementary and alternative medicine. London: Stationery Office.
22. Thomas KJ, Fitter M, Brazier J, MacPherson H, Campbell M, Nicholl JP, et al. Longer term clinical and economic benefits of offering acupuncture to patients with chronic low back pain assessed as suitable for primary care management. *Comp Ther Med* 1999; 7(2): 91-100.
23. Vickers A, Rees R, Zollman C, Ellis N. Acupuncture for migraine and headache in primary care: a protocol for a pragmatic, randomised trial. *Comp Ther Med* 1998; 7: 3-18.
24. Helms PJ. 'Real World' pragmatic clinical trials: What are they and what do they tell us? *Pediatr Allergy Immunol* 2002; 13: 4-9.
25. MacPherson H. Pragmatic clinical trials. *Comp Ther Med* 2004; 12: 136-140.

HOW TO DO AND REPORT SYSTEMATIC REVIEWS AND META-ANALYSIS

¹ Professor of Statistics and Clinical Epidemiology, School of Medicine - University of Modena and Reggio Emilia, Italy. Director of the Italian Cochrane Centre, Mario Negri Institute Milano, Italy.

² Director of the Rehabilitation Medicine Unit, Azienda Ospedaliera Universitaria S. Orsola Bologna, Italy.

Which type of organized care pathway should a hospital trust organize in order to improve the outcome of stroke patients? How should care be best organized in a geriatric ward in order to prevent the occurrence of pressure soars? How can we identify an effective management to reduce the loss of working days due to back pain for health professionals?

All the statements above are just examples of the hundreds of questions that arise in clinical practice. But how can we find reliable answer in a manageable time frame? Some questions are simpler than others and some require individual practitioners' change in practice while others need complex organizational modifications. In this charter we will discuss about the emerging science of systematic reviews.

The hypothesis that organized stroke patients can benefit from being cared for in hospitals with dedicated teams and beds is not recent and we can already find the initial studies back into the mid years of the last century. However, the trials published up to 1990 comparing the outcome of patients care in stroke units vs. those cared in non-specialized units failed to demonstrate it convincingly as individual trials were too small: the largest individual trial enrolled 417 patients and the smallest 52! It was necessary to wait a systematic review that today analyzes the outcomes of

6393 patients enrolled in 31 trials and show that organized care improved beyond the play of chance patients outcomes (1).

What do we know about the risk factors of pressure soars? The assessment of individual patients' risk should be the first step. A recent systematic review has shown that despite the widespread use of risk assessment tools, there are no randomised trials that compare them with unaided clinical judgement or no risk assessment in terms of rates of pressure ulcer. Therefore, we cannot conclude whether the use of structured, systematic pressure ulcer risk assessment tools, in any health care setting, reduces the incidence of pressure ulcers (2). Even the widespread use of manual repositioning as a component of the management plan for individuals with existing pressure ulcers, is not truly based on robust evidence given that no randomised trials exist that assess the effects of repositioning patients on the healing rates of pressure ulcers (3). More reliable information exist on the risk and benefits of devices and a systematic review of 52 randomized trials showed that foam alternatives to the standard hospital foam mattress can reduce the incidence of pressure ulcers in people at risk even though the relative merits of alternating and constant low pressure devices are unclear (4).

And what about the treatment of back pain? The effectiveness of exercise as been evaluated on sixty-one randomized controlled trials with 6.390 participants. Exercise therapy appears to be slightly effective in decreasing pain and improving function in adults with chronic low-back pain, particularly in healthcare populations. In sub acute low-back pain there is some evidence that a graded activity program improves absenteeism outcomes, though evidence for other types of exercise is unclear. The conclusions of these reviews are however uncertain given the limitations of the literature, including low quality studies with heterogeneous outcome measures, inconsistent and poor reporting, and possibility of publication bias (5). The patient education approach as been studied in 24 trials testing different types of patient education for people with low-back pain measuring outcomes like pain, function and return-to-work. People with low-back pain who received an in-person patient education session lasting at least two hours in addition to their usual care had better outcomes than people who only received usual care (6).

In this chapter we will discuss what a systematic review is, how it should be carried out, and when - if the clinical question and the data from available studies allow it - a quantitative combination of data (known as meta-analysis) should be carried out.

FIRST SECTION: GENERAL PRINCIPLES

1.1 INTRODUCTION

Research synthesis has a solid tradition in science. In the field of social sciences, for example, the importance of adopting a rigorous

methodology to synthesize the results from various studies on the same topic has been accepted long ago (7). This is, unfortunately, not yet the case in medicine even though systematic reviews (SRs) and meta-analysis (MA) have gained an increasing recognition.

In the medical field the first SRs and MAs were published in important journals starting from the end of the 70's. Recent data suggest that at least 2500 new systematic reviews reported in English are indexed in MEDLINE annually (8).

An important clarification of what a SR is, and about the differences between SR and a traditional review is given in a paper published many years ago in *Annals of Internal Medicine* by Cynthia Mulrow (9). In Table 1 these differences are summarised. From here onwards the SR literature has gained recognition in epidemiological methods, thanks to the ability shown to help resolving scientific controversies and defining the priorities for the direction of future research. This growing importance is also indicated by the fact that SRs have recently obtained the status of scientific research at an academic level.

TABLE 1. Main differences between Systematic and Narrative Reviews.

NARRATIVE REVIEW	SYSTEMATIC REVIEW
<ul style="list-style-type: none">➤ Gives a “broad” view, in general dealing with the whole argument: an example is a chapter of a textbook➤ Emphasises the “background” knowledge (looks backwards): What is the cause of a disease? What are the clinical signs? What are the therapeutic options?➤ Given that most often it is commissioned to an expert alone it is susceptible to systematic errors in selecting, evaluating and combining the various studies to answer the questions.	<ul style="list-style-type: none">➤ Gives a precise “narrow” view, in general answers one clinical question or just a couple.➤ Focuses on the areas of uncertainty (looks forward)➤ Starts from a very rigorous methodology to minimise the risk of systematic errors, looking for conclusions based on solid evidence➤ Can provide a quantitative summary of the various considered studies when performing a meta-analysis is appropriate

However, even though the idea that scientific knowledge should be seen a continuous and cumulative process, “consumers” of the scientific literature in medicine tend to read and assess each study in isolation from the whole of the remaining evidence. This alarming situation was brought to the medical community’s attention by an article in 1999 where Clarke and Chalmers (10) evaluated in how many cases the results of a new study were discussed (in the discussion part of the article) in the context of already available knowledge. Only in 11% of cases the new results published were presented in the context of already available knowledge, whilst in 40% there was no mention of previous studies or competitors, and in the remaining 49% results of other similar investigated studies were quoted in an ‘unstructured fashion’. A similar study conducted by

the same Authors few years later failed to detect any significant improvement (11) and this led to the decision taken by R Horton at Lancet to require that reports of clinical trials submitted to the journal should discuss their results in the context of previous studies (12).

Also among research funding agencies there is an effort being made to face this challenge. The British Medical Research Council, for example, has adopted a policy whereby all the requests for financing phase III and IV studies must be accompanied by a Systematic Review of previous studies with the aim of evaluating the real significance and need of the new proposal (13).

BMJ publishes, in the context of the original articles, a table where Authors are requested to state: a) what is already known before the research being published had been carried out; b) what the new study being published adds to the available knowledge.

It's not by chance that the development of the modern methodological thinking, known as Evidence Based Medicine (14), sees SRs and MAS as essential instruments to define practice guidelines and as reference for continuously keeping health professionals up to date. Table 2 illustrates the two opposite (and untoward) consequences of failing to implement rigorous evaluations of the benefits and harms of medical interventions. In the upper panel of the table some intervention that had a far too rapid acceptance without empirical assessment is shown while, in the lower part of the table, some interventions that had an unacceptably long lag-time between the demonstration that they can improve patients' health and their uptake in clinical practice are also listed.

Far from being a list of exception, data presented in this table should be a strong warning signal that in modern medicine synthesising research evidence is very important and SRs are an essential means toward that goal.

TABLE 2. Some examples, past and recent, of mismatches between knowledge and utilisation in clinical practice.

PROCEDURES LACKING EFFICACY DOCUMENTATION BUT WIDELY USED
<ul style="list-style-type: none">• Bleeding and blood letting (more than 2 centuries)• Negative effect of blood letting in pneumonia (about 70 years)• Pneumothorax in TBI (Forlanini technique) (about 40 years)• Frontal lobotomy for psychiatric problems (20 years)• Albumine in critically ill patients (still today)• Hormonal replacement therapy in menopause (insufficient evidence, still today)• Routine Vitamin K supplement, calcium and magnesium in pregnancy (still today)
DELAY IN ACCEPTING PROOF OF DOCUMENTED EFFICACY
<ul style="list-style-type: none">• Vitamin C in prevention of scurvy (about 40 years)• Hydration in cholera (70-80 years)• Corticosteroids in premature babies (10-15 years)• Thrombolysis in acute myocardial infarction (about 15 years)• Therapy with beta-blockers in secondary prevention of heart-attack patients (15 years)• Administration of folates during pregnancy (still today)

1.2 WHY ARE SYSTEMATIC REVIEWS IMPORTANT?

1.2.1 Making sense of the volume of information

The quantity of information both researchers and health workers face is in continuous growth. Roughly, every year about one million articles in 30,000 scientific journals and about 17,000 books in the biomedical field are published, with a yearly growth estimated in around 7% (15).

As well as the quantity of the information, which alone represents a formidable challenge, there is also the problem of the methodological quality of it. Evaluating the validity and the reliability of the information in the scientific journals is an arduous job and many of the health professionals are called to answer without ever having been properly exposed to the appropriate preparation during their training. In this sense to be able to have access to reliable and high quality reviews is an enormous advantage in terms of time and guidance. A single study, even if properly conducted and analysed, if did not enrol a sufficient number of patients, is highly likely to fail to show a statistically significant relationship between an etiologic agent and an illness or between surgery and the improvement in the health status. In this case, if one considers only this study without looking at the totality of available evidence, will end up with a wrong conclusion that a specific agent does not have a casual role or that a specific treatment does not improve the prognosis of the illness. This is usually refereed as a 'false negative'.

A SR carried out in the 90's to evaluate the quality of trials on the treatment of lung cancer (16) found that more than 70% of the studies did not have a sufficient statistical power to demonstrate a statistically significant improvement in survival > 30%. A more recent review of the studies aimed at evaluating efficacy of different treatments for schizophrenia (17) documented that just 3% of the studies had the sample size necessary to gather clinically relevant differences.

Moreover, even where a single study showed in a significant statistical way an association, an evaluation based on the whole of the available studies is useful to rule out that it is a 'false positive'. Only evaluating it in the context of other available studies can the interpretation be appropriate.

1.2.2 Systematic Reviews and Meta-analysis

Ever since qualitative analysis works and synthesis of results of several studies carried out on the same topic have begun to be published in literature, various terms have appeared in literature in English (systematic review, overview, meta-analysis, statistical overview, best evidence synthesis) with a growing confusion on the general methodological or strictly statistical aspects of the problem. Only recently a definition has been proposed which is now accepted internationally (18).

By the term 'systematic review' we mean a review carried out according to specified methodological plan, directed at minimising possible bias and wrong conclusions due to a loss or exclusion of important studies that are available on a specific topic (18).

What, instead, is meant by the term ‘meta-analysis’ (MA) is a statistical combination of data coming from independent studies with the aim of producing an overall evaluation of the effect of an intervention/exposure (18).

In this sense a SR can but should not necessarily include a MA. It must, in other words, be clear that performing a quantitative synthesis (MA) in the context of SRs is not always appropriate. While, in fact, a systematic look and a qualitative description of the merits and pitfalls of a given body of research is always important, quantitatively combining information drawn from studies that are too different in their methodological quality, or too heterogeneous in terms of their results, may do more harm than good.

A thorough analysis of the coherence, consistence and quality of a group of studies is one of the most important features in a SR. In this sense a SR can be considered as a tool capable of providing an epidemiological assessment of the results of a group of independent studies carried out on a common argument. However one should also keep in mind that this ‘gathering of results’ (be it formal and quantitative as in a MA, or just qualitative as in a SR) suffers from the same limits that affect observational studies such as variability in the definitions of disease, in the selection of patients, as well as in the reporting and publishing rules that may have been planned differently in individual studies.

KEY MESSAGES

- 1) A single study must be evaluated in the context of similar studies available in the literature.
- 2) If not done in a systematic and consistent way, however, the review process can produce biased results and even amplify the bias carried by each individual study.
- 3) Not all SRs can and should end up combining quantitatively the results of individual studies.
- 4) When studies are homogeneous and comparable, a SR can legitimately lead to a MA. The results of the MA provide an indication of the ‘average’ likely magnitude of the effect attributable to the exposure or intervention.

SECOND SECTION: SYSTEMATIC REVIEWS

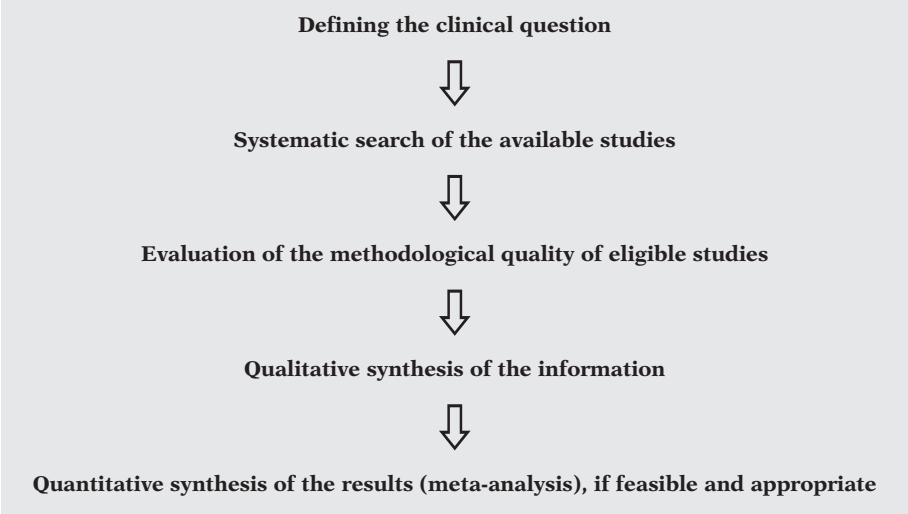
2.1 STEPS OF A SYSTEMATIC REVIEW

2.1.1 Planning a systematic review

Unlike primary studies, designed to pursue specific aims defined by the investigators, SRs are studies whose validity and informativeness are defined by the very nature of the information available for the review. This is one of the “a priori” limits of SRs as information gathering tool.

As the SRs are, by their very nature, retrospective it is important that the process be as rigorous and explicit as possible. A SR of good quality always begins with a detailed and completed research protocol. The main stages which must be dealt with in a SR, and made explicit in the protocol, are reported in Table 3.

TABLE 3. The main steps in carrying out a systematic review.



2.1.2 The definition of the question/s of a SR

The definition of the research question is the first stage and it is also where a large part of the feasibility of a SR is decided. There is no single way of defining the question of a SR, besides which the definition will largely depend on the degree of generality/ specificity which is wanted in the answer and by the availability of studies on the argument.

It is important to keep in mind that the review question influences the search of the studies, the yield of the search and the possibility of performing the analyses that the review sought to address.

Since a precise and clear formulation of these questions is not simple, a standardised and explicit procedure has been proposed which allows for the definition of a) the population, (“**P**”); b) the types of intervention (“**I**”); c) the comparator against which the new intervention will be tested (“**C**”); d) the outcomes relative to the aims of the SR (“**O**”). This is usually refereed, with an acronym, as the definition of the “**PICO**” for a systematic review.

Beside these 4 elements the study designs are also to be specified; these are the basis of the primary studies that the SR will use to answer the clinical question of interest.

- Specify the “**P**opulation” means to define which group, or groups, of subjects we are interested in vis a vis the clinical question of efficacy/

impact of a determined intervention or exposure. All factors that constitute a reason for restriction of the population are also to be specified such as sex, age, race, severity of the disease and co-morbidities.

- Define the “Interventions/exposures” means to define what the treatment alternatives we want to verify and how they are administered.
- Define the “Comparator” against which the treatment will be tested is essential as it is necessary to understand whether the efficacy or its absence can be appropriately claimed.
- Specify the “Outcomes” means to indicate the type of end points which we want to assess as a consequence of the interventions or exposures of interest.
- Finally, deciding which study designs are most appropriate to the review implies having a clear idea of the research architecture, at least in theory, most appropriate to answer the clinical question-objective of the SR.

An example of how the PICO can be defined is reported in Table 4 with reference to a SR assessing the therapeutic value of topical vs. systemic antibiotic in subjects suffering from skin ulcers in lower limbs.

TABLE 4. Example of definition of the PICO for a SR about the effects of topical o systemic antibiotics in skin ulcers of the lower limbs.

Populationin adults suffering from chronic ulcers of the lower limbs and assisted in outpatient clinic.....
Intervention/exposureadministration of topical or systemic antibiotics
Comparison while in the control group patients received a placebo.....
Outcome Can it improve the “healing rate”?
Study Design	... studies which have compared subjects suffering from skin ulcers of the lower limbs who have been treated with topic e/o systemic antibiotics in comparison to a group of subjects with the same characteristics but not receiving any antibiotics but treated with other therapeutic conditions. The assignment was purely casual (controlled and randomized study)

It is clear that the choice of the type of study design is in essence tied to the type of question to be dealt with. It is important to remember that a relationship exists between the type of questions and the type of study designs chosen; it determines the appropriateness of the research done.

For example, questions of “aetiology” need have cohort or case-control studies at the top of their preferability of study designs. Clinical questions regarding the efficacy of preventive or rehabilitation treatments need, whenever possible, randomized controlled clinical trials or alternatively useful information can also be obtained from prospective cohort studies. That said, the concept of “hierarchy of evidence” should be seen not as a rigid tool but in a dynamic way and there are not absolute and

rigid univariate correspondences between type of questions and study design; rather there are preferable and less preferable options of which everyone doing critical appraisal must be aware of.

2.1.3 Differences among studies: the advantages and dangers of heterogeneity

Most of those that are detractors of SRs and, above all MAs, base their concerns on the charge that MAs are often “fruit salads” putting together study results which are not comparable for the question being addressed, the methodological quality, the types of patients etc.

This is the problem of ‘heterogeneity between studies’ and is determined by two main components: **methodological** and **clinical**. To have methodological and clinical heterogeneity it is necessary that an interaction between the effect of the treatment/exposure and the characteristics of the study exists.

To explore clinical heterogeneity we look at the differences between studies concerning patients, treatment/exposure, settings of the study and the type of outcome. For example, patients enrolled in various studies can be different in terms of prognostic and demographic characteristics (sex, age, race, concomitant pathologies etc); the treatments/exposures can differ in terms of dosage, intensity and conditions of administration (topical antibiotic 500 mg VS oral antibiotics 300 mg); the health services where the study is being carried out can vary (teaching hospital, research centre, state hospital, private clinic); also the Authors can use different outcomes to measure the effects of the treatment/exposure).

To explore methodological heterogeneity we must look at the different ways of carrying out the studies. This refers to the type of study design (experimental or observational), the quality of the studies, type of allocation of the patients to the experimental arms or to the comparability of the groups, presence of blind criteria, number of patients lost in the follow-up etc, and to the type of analysis carried out (‘intention to treat’ or ‘by protocol’).

Since the inception of SRs, i.e. during the definition of the protocol, the researcher attempts to control the potential sources of heterogeneity. In doing so the researcher identifies the possible reasons for which the results of primary studies can vary and defines the inclusion and may come both at the protocol drafting phase as well as after the selection of the studies. During the drafting phase of the protocol, in fact, it is not always possible to understand which characteristics make the studies different. So, once the studies have been collected the question to which the researcher should try and answer is: “can the differences present amongst the studies have an importance on the results? Can these differences modify the direction of the effect of the intervention in question?” In cases where the differences were considered influential, the choice should lean towards the “non-combination of the results” and the review should limit itself to the qualitative description of the studies and the results obtained.

If the answer to the question is negative you can proceed with the quantitative combination of the data and then with the formal statistical test necessary to test the hypothesis of the homogeneity/non-homogeneity amongst the studies.

A more detailed and thorough description can be found in the Handbook of the Cochrane Collaboration as far as its reviews are concerned (15).

KEY MESSAGES

- 1) SRs must be based on a written protocol specifying: a) how the systematic search of available evidence will be carried out; b) how the evaluation of the methodological quality and of the coherence and consistency among different primary studies will be carried out.
- 2) The protocol should also specify: a) SR's primary and secondary aims; b) its target population; c) which intervention/exposures are going to be considered; c) what comparator are included; d) which outcomes are of interests; e) which studies will be eligible given their study design.
- 3) There is a relationship between the clinical question to be addressed and the most appropriate study design. When designing a SR protocol the researcher must decide whether she wants to be highly selective (i.e. including only RCTs in her review) or adopt broader inclusion criteria. The pros and cons on the choice of the type of design vary in fact depending on the aims of the review.
- 4) Clinical and/or methodological heterogeneity that may exist across component studies is the 'Achilles Heel' of a MA.

2.2 HOW DO YOU SEARCH FOR EVIDENCE?

2.2.1 The traditional model

Health professionals are used to use 5 simple strategies to find useful information:

- Seek bibliographical references from experts colleagues;
- Browse texts and volumes used during his/her training period or their more updated versions;
- Go through the journals he/she has subscriptions to, or receives for free;
- Look through the bibliography of a study or review he/she seems is well done;
- Carry out a bibliographic research, generally from the most accessible database.

The colleague can give useful advices based mainly on his/her experience and personal points of view. In this case, however, it will be based on the personal choice of that person.

Information gained from books suffers the same limitations. They are rarely completely up-to-date (at best a year has passed since an update but usually more than this) and they are also based, at best, on qualitative expert-based review of the evidence.

Quality of medical journals differs: using them is not only complex but are also very time consuming (a lot of careful reading of the indexes of each volume is needed).

This is why it is necessary to turn to biomedical literature data banks, where all the bibliographic references can be found as well as the summaries of the articles of the majority of biomedical journals from all over the world. These databases are about 100 in all: some cover all the areas in medicine, others just a specific area.

2.2.2 The databases

MEDLINE is by far the most accessible: it is produced by the National Library of Medicine, the largest biomedical library in the world with its head office in the USA. It contains bibliographical citations starting from 1969, it is updated weekly and can be searched free of charge on (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi>).

EMBASE is produced by the European publishers Elsevier and contains bibliographical citations starting from 1974. It covers all areas of medicine but is particularly specialised in the pharmacology sector and mostly reviews European journals. On the contrary to MEDLINE this is not free and can only be used through research softwares. The more than 4000 journals indexed in both data banks only overlap for about 35%. Therefore, for a good bibliographic research both these data banks should be consulted.

The main information belonging to an indexed article in a data bank go to form a bibliographic record composed of areas called 'fields'. The main fields are: Title, Authors, Authors' addresses, Bibliographic references, Summaries, Descriptors, Type of publication, and Language of publication (if different from English).

Both MEDLINE and EMBASE have dictionaries controlled by 'key words', where specific or just more general words can be inserted. A good research should be 'sensitive', that is to find all the existing articles on a determined argument, and 'specific', that is to find all the relevant articles leaving out those of no interest. A sensitive research, but not specific, would produce a 'mountain of paper', on the contrary, a research that is too specific and not sensitive will leave the 'basket empty' (15).

Both MEDLINE and EMBASE have dictionaries controlled by "key words", organized according to a structured tree-hierarchy.

Finally it's important to mention 'The Cochrane Library' (CL) (<http://www.update-software.com/cochrane/>), a database produced by the Cochrane Collaboration (20), which is in essence a "meta-data bank" containing: a) Cochrane systematic reviews (in the database CDSR, an acronym for Cochrane Database of Systematic Reviews); b) a database of abstracts of non Cochrane reviews (DARE); c) a data base of references on

controlled and/or randomized trials on the efficacy of health services (CENTRAL); d) a database of methodological review; e) a database with reference to health technology assessment reports.

In the last Issue 5, 2010 the following information are available:

- CDSR includes 6162 articles: 4.247 reviews; and 1.915 protocols with 30 new reviews, 85 updated reviews and 23 new protocols.
- DARE Database of Abstracts of Reviews of Effects (Other Reviews) includes 12,594 abstract,
- CENTRAL" contained 620.700 references of Clinical Trials, representing the most extensive data bank on controlled trials.
- Cochrane Methodology Register includes 13.164 Methods Studies,
- Health Technology Assessment Database includes 8.273 Technology Assessments reports,

Rehabilitation is by definition cross disciplinary and it is thus not easy to assess how often it is the topic of controlled trials and thus of SRs of the type that the Cochrane Collaboration does. Broadly speaking research in rehabilitation has produced several good quality studies and a search carried out in the Cochrane Library, Issue 4, 2006 using the keyword "rehabilitation" in title, abstracts or keywords found 146 SRs; 22 protocols of ongoing Cochrane SRs; 243 non Cochrane reviews in DARE; 9163 RCTs or CCTs indexed in CENTRAL. Out of the 168 Cochrane SRs the numbers could be broken down as: 57, 49, 13, 10, and 3 in orthopaedic, neurological, respiratory, mental health and cardiological rehabilitation, respectively. The other 36 were spread across many different specialties (19).

2.2.3 Publication Bias

Publication bias (PB) is one of the major threat to the validity of SRs and MAs. Not all the studies that are carried out end up in full publication in journals or books. PB is particularly serious as it does not happen randomly but rather its likelihood of occurrence is higher when the study is "negative" or presents "no difference". The probability of a scientific work is published varies on the basis of the conclusions it came to. In other words, the risk of PB is great when the possibility is high that the results of a research are published only if the results documented are in favour of the intervention (statistically significant).

Rosenthal was first, in 1979 (20), to draw people attention to this type of phenomenon (file drawer problem): "... the journals are filled with the 5% of the studies that show Type I errors, while the file drawers back at the lab are filled with the 95% of the studies that show non-significant results..."

In the start-up stage PB can be reduced but not completely avoided. Using a complementary research strategy consulting the grey literature' (thesis, monographs and documents from specialized agencies etc.) and directly asking the authors of the published studies news of other studies that have been carried out we can have an initial rough idea of the potential amount of PB in a given area.

In the interpreting the results stage it should be kept in mind that the PB tends to be more frequent in certain sectors of medicine, in the observational studies rather than the experimental ones and for the SRs based on a small number of studies with many low samples numbers.

For an in-depth look into this argument a comprehensive reference is an English monograph (21).

2.2.4 Selective reporting

Selective reporting has been defined as the selection of a subset of the original variables recorded on the basis of the results, for inclusion in the publication.

The particular concern here is that statistically non significant results may be selectively withheld from publication. The existence of selective reporting is a relatively new finding in methodological research around systematic reviews and meta-analysis. There were a few case studies done comparing protocols approved by Ethics Committees and subsequent publications of the studies.

Convincing empirical evidence comes from a series of studies conducted in Canada by Chan et al and published in 2004 (22-23) which have been subsequently confirmed by others (24).

Selective reporting of outcomes may arise in several ways, some affecting the study as a whole (selective omission of outcomes from reports) and some other related to specific outcome comparisons (such as selective choice of data for an outcome, selective reporting of analyses using the same data, selective reporting of subset of data, selective under reporting of data).

At the moment there is not an effective solution to selective reporting and some strategies to partially handle it are discussed in Chapter 8 of the Cochrane Handbook (15).

KEY MESSAGES

- 1) A valid and comprehensive search strategy should be based on electronic data bases.
- 2) "Launching" a search strategy with "free words", key words, Boolean operators and limits, is possible to reach excellent "sensitivity" and "specificity".
- 3) A thorough bibliographic search should be done with the help of skilled professionals (scientific documentalists) and not be based only from those databases that are easily accessed such as MEDLINE (PubMed).
- 4) SRs Authors should always be aware of potential for *publication bias*. If it occurs it can jeopardize the results of a SR but, even more so, in a MA (generally amplifying the effect of the association between exposure/intervention and effect, given that negative studies are those more likely to go unpublished).
- 5) *Selective reporting* of studies and outcome has been identified more recently and it is a serious threat to the validity of a MA, probably even more than full publication bias as it is more subtle and difficult to ascertain.

2.3 WHAT IT IS AND HOW WE CAN ASSESS THE QUALITY OF A SR

The extent to which a SR and a MA produce valid results depends on whether the data and results from the included studies are valid and whether all the available evidence has been found, appraised and used. We already discussed that SRs are conducted retrospectively and this makes them susceptible to potential sources of bias

The phrase “assessment of methodological quality” has been used extensively in the context of SRs methods to refer to the critical appraisal of component studies. The term suggests an investigation of the extent to which study Authors carried out their research to the best possible methodological standards. In other words, we say that a SR (and a MA, if done) is of good quality if it answers the research question “correctly”, which is in a manner that is free of bias. Said in other words, when we assess the “quality” of a SR/MA we are interested in its “internal validity”. Quite different is the concept of “quality of reporting” - which we will address later in this chapter - that refers to the extent to which SR/MA reports do provide all the information relevant to understand how they have been carried out.

Until recently there has been a strong emphasis on the term “quality assessment” both in the context of the assessment of individual studies and systematic reviews (18). However, the increasing awareness that even when we identify a bias (defined as “a systematic error”, or a “deviation from the truth”) in a study it is always difficult to predict how, and in which direction, it may affect the results, has led to the decision to focus on the concept “risk of bias” rather than “bias” in itself (15). This means to recognise that the results of a study may in fact be unbiased despite a methodological flaw which was made but has been irrelevant in terms of distorting the results. This is the reason why, for example, the Cochrane Collaboration has introduced a method to assess “risk of bias” as a more appropriate approach to the assessment of the credibility of the results of a RS/MA (15). Differences in risk of bias may explain variation in the individual studies included in a SR; similarly MA from studies of variable validity can results in false positive conclusions (erroneously concluding that an intervention is effective) if the less rigorous studies are biased toward underestimating an intervention’s effect.

The risk of bias in studies candidates to be included in a SR should be performed irrespectively of the anticipated variability in their results as, in fact, a whole set of studies carried out on a given topic can all have similar results but can be all flawed and thus the conclusions of the SR should be much more conservative.

In 1996 Moher et al identified over 25 scales and 9 checklists (25) and since then, though no specific study has updated Moher’s findings, more new tools have been proposed.

Many tools have been proposed for assessing the “quality” of studies for use in the context of a SR. Many of them are “scales” in which various components of quality are scored and combined to give a summary score, and some are checklists in which specific questions are asked (15).

Most of these tools are based on “generally accepted” criteria and there was a mix of issues having a likely influence on results and issues (such as a priori sample size calculation or specification of inclusion/exclusion criteria), which clearly have to do with precision and applicability, respectively, rather than internal validity. In general, scales appeared more likely than checklist to include a mixture of criteria (related to internal and external validity) and to produce a difficult-to-understand mix). There is a general consensus that none of this tool is the standard and that the use of scales should be discouraged in favour of checklists (26).

The Cochrane Collaboration recommended tool for assessing risk of bias is neither a scale nor a checklist but it is rather a “domain-based” evaluation in which critical appraisal is done separately for different domains. It was developed by a working group of methodologists, editors and review authors (15). This approach explicitly recognizes that it is impossible to know the extent of bias in a given study and that the most realistic assessment of risk of bias may involve some subjectivity: assessing whether the lack of blinding of patients may plausibly have affected the occurrence of an outcome in a given study.

Individual studies included in Cochrane reviews are assessed in terms of risk of bias associated with the following dimensions: a) Sequence generation (how the allocation sequence for the randomisation procedure was created); b) Allocation concealment (how it was avoided that the allocation could be guessed in advanced by study participants); c) Blinding of participants, personnel and outcome assessors (whether, given that blinding would have been appropriate to increase the validity of the study, efforts to blind had been done or not); d) Incomplete outcome data (description of the completeness of individual outcome data, attrition, losses to follow up); e) Selective outcome reporting (whether is plausible, and likely, that specific information on one or more outcome(s) has been suppressed); f) Other sources of bias.

2.3.1 Quality of reporting

A key difficulty of the assessment of risk of bias or quality is the obstacle provided by incomplete reporting. While the emphasis for this assessment should be on the risk of bias in the actual design and conduct it can be tempting to take the shortcut of assessing the quality by looking at the completeness of reporting. This has become clear since the publication of QUOROM (27) in 1999. QUOROM was a guidance aimed at helping people the completeness of reporting in articles presenting the results of SRs/MAs. While the concept was very clear in the original paper

there have been cases where authors stated that they were assessing “the quality of SRs” while in fact they were assessing the quality of reporting (28).

To respond to the advances in the science of systematic review occurred over the last two decades QUOROM has evolved into a more complete and comprehensive guidance called PRISMA - which stands for Preferred Items for Reporting of Systematic reviews and Meta-analysis (29-30).

PRISMA is a 27 items checklist developed over a three years period by an international group of experts and has already been endorsed by several peer reviewed medical journals.

The Physical Therapy Journal (PTJ) joined other journals - such as BMJ, Annals of Internal Medicine, PLoS Medicine, and Journal of Clinical Epidemiology where it was originally published - in endorsing PRISMA. In September 2009 in PTJ the publication of PRISMA was accompanied by a Christopher Maher'editorial which asked its contributing Authors to follow this statement when preparing manuscripts reporting a systematic review (31).

In Table 8 the full list of items is reported together with a short explanation of their meanings. More information is available at <http://www.prisma-statement.org/>.

KEY MESSAGES

- 1) The extent to which a SR and a MA produce valid results depends on whether the data and results from the included studies are valid, and whether all available evidence has been found, appraised and used.
- 2) Many conceptual and methodological advances have occurred in the field of the assessment of the quality of studies and systematic reviews in the last 10 years.
- 3) Many tools have been proposed for assessing the "quality" of studies in the context of a SR. Many of them are "scales", in which various components of quality are scored and combined to give a summary score, and some are checklists in which specific questions are asked.
- 4) There is a general consensus that none of this tool is the standard and that the use of scales should be discouraged in favour of checklists.
- 5) Most of these tools are based on "generally accepted" criteria and there is a mixture of issues having a likely influence on results and issues (such as a priori sample size calculation or specification of inclusion/exclusion criteria) which clearly have to do with precision and applicability, respectively, rather than internal validity.
- 6) When a bias is identified in a study it is always difficult to predict how, and in which direction, it may affect the results.
- 7) To overcome these problems the Cochrane Collaboration has decided to focus in its SRs on the concept "risk of bias" rather than "bias" in itself and has produced an instrument called "the risk of bias tool" (RoBT).
- 8) The RoBT is neither a scale nor a checklist but a "domain-based" evaluation in which critical appraisal is done separately for different domains. This approach explicitly recognizes the impossibility of assessing the extent of bias in a given study and that some subjectivity is unavoidable.
- 9) A key challenge in the assessment of risk of bias is incomplete reporting which may confound our assessment. While the emphasis should be on the risk of bias in the actual design and conduct, it can be tempting - yet inappropriate - to assess the quality by looking at completeness of reporting.
- 10) The first guidance aimed at assessing the completeness of reporting (called QUOROM) was published in 1999 and has now evolved into a more thoroughly developed instrument called PRISMA published in 2009 which is publicly available at <http://www.prisma-statement.org/>

THIRD SECTION: META-ANALYSIS

3.1 INTRODUCTORY CONCEPTS

As already mentioned in the introduction, a MA is the step in the systematic review process where we use of statistical methods to combine the results from “similar-enough” studies evaluating a “similar-enough” clinical question. When we pool the data from different studies we increase the statistical power of the analysis, which leads to more precise estimates, and look in a systematic manner at the consistency of results.

A MA can be carried out when: a) more than 1 study has assessed the treatment/therapy effect; b) the differences between studies in terms of patients, interventions and characteristics of the health care setting are minimum and/or do not allow to predict a priori an impact on the outcomes; c) the outcome of various studies has been measured in a similar manner; d) the Authors of the primary studies have published the numerical data necessary to carry out the combination.

3.1.1 The logic of meta-analysis

Before introducing statistical methodologies that are used to carry out a MA, it is important to clarify that a MA is not done by simply adding the number of patients and the events across all studies.

Let’s suppose that two studies (study 1 and study 2) are assessing the efficacy of an experimental treatment A compared to treatment B (see tables 5 and 6).

TABLE 5. Hypothetical results of 2 studies comparing the effect of treatment A and B on disease recurrence.

Studies	Treatment A		Treatment B		Relative Risk (RR)*
	N° patients with relapse	Total N° patients	N° patients with relapse	Total N° patients	
Study 1	40	100	50	200	1.6
Study 2	15	200	5	100	1.5

* The RR is calculated for each study and data are not combined

TABLE 6. Hypothetical results of 2 studies comparing the effect of treatment A and B on disease recurrence.

Studies	Treatment A		Treatment B		Relative Risk (RR)*
	N° patients with relapse	Total N° patients	N° patients with relapse	Total N° patients	
Study 1	40	100	50	200	1.6
Study 2	15	200	5	100	1.5
Total	55	300	55	300	1

* The RR is calculated for each study but, contrary to what is shown in Table 5, there is in the last column the RR calculated treating study 1 and 2 as a unique trial and adding up the number of patients: This produce a RR estimate that indicates the Simpson Paradox (ref 32).

In the first study the risk of developing the outcome in treatment A equals 0.4 (40/100), whilst in treatment B it is 0.25 (50/200).

In study 2 the risk of developing the event in treatment A is 0.075 (15/200), whilst in treatment B it is 0.05 (5/100).

The relative risk calculated in study 1 equals 1.6, whilst in study 2 it equals 1.5.

If the results of these two studies were combined adding up the cells, and the total relative risk calculated as if the data came from one large table, then the result would be the one reported in Table 6 (bottom, in bold).

Even though the RR is >1 in both studies, suggesting that the risk is higher in the group treated with intervention “A” rather than intervention “B”, when the data are combined the total RR gives an indication of no-difference between the two treatments being compared. This phenomenon is known as “Simpson’s paradox” (32) and gives a clear idea of why combining data by adding up the cells of the tables gives wrong results.

How than a proper meta-analysis should be done?

A MA is a two steps process which considers any single study as a unit of observation.

In step 1 the treatment effect of each study is calculated. If the variable is “binary” (i.e. categorical) the treatment effect is evaluated using either the odds ratio (OR), the relative risk (RR) or the risk difference (RD). In case of continuous variables results can be summarized through the mean and the treatment effect is measured calculating the mean difference (MD).

In step 2 the overall estimate of the treatment effect is calculated. This estimate is obtained calculating the weighed mean of the estimates obtained in the first step. The weights associated to each study are chosen in a way that reflects the amount of information contained in a single study. These weights are proportional to the sizes of the samples of the study and, where data are binary, and to the frequency of the observed events of the compared groups. Together with the total estimate of the treatment effect, the confidence intervals and the statistical significance are calculated.

3.1.2. Combination of the results

In Table 7 the results of 16 studies where the efficacy of antibiotic prophylaxis in reducing the risk of nosocomial infections in patients admitted to intensive care are presented (33).

One way of graphically illustrating the data shown in the table is to use the Forrest Plot (Figure 1).

The graph has, on the X axis, the summary measure related to the size of the effect (expressed as RR, OR or RD) while the Y axis indicates the line of “no difference” among between the two treatments being compared. This value equals 1 when estimates are expressed as OR or RR and 0 when the summary measure is expressed as RD.

In the graph at the right hand side of the Author’s names we see small squares - one per study and whose size is proportional to the number of patients/events generated - and the line emerging from both sides of the square indicating the “confidence intervals” (usually 95%).

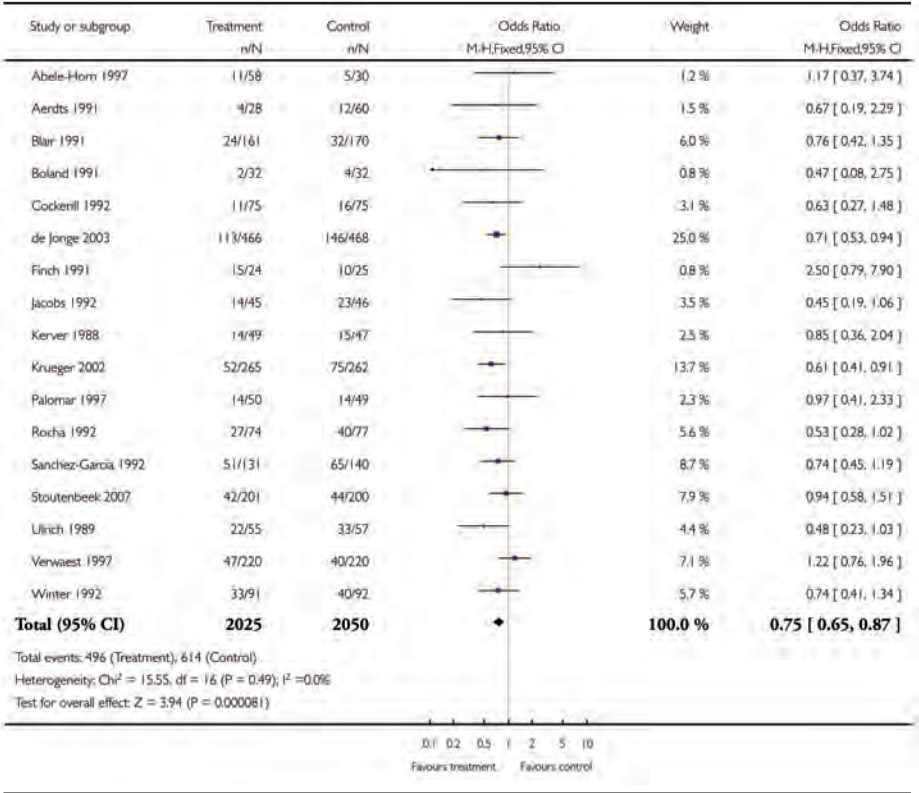
TABLE 7

Studies	Antibiotic Prophylaxis		No antibiotic Prophylaxis		Estimated treatment effect RR (95%CI)
	N° patients with infections	Total N° patients	N° patients with infections	Total N° patients	
Abel-Smith	13	58	23	30	0.29 (0.17,0.49)
Aerdt	1	28	29	60	0.07 (0.01,0.51)
Blair	12	161	38	170	0.33 (0.18,0.61)
Boland	14	32	17	32	0.82 (0.49,1.37)
Cockerill	4	75	12	75	0.33 (0.11,0.99)
Finch	4	20	7	24	0.69 (0.23,2.01)
Jacobs	0	45	4	46	0.11 (0.01,2.04)
Kerver	5	49	31	47	0.15 (0.7,0.36)
Palomar	10	50	25	49	0.39 (0.21,0.73)
Rocha	7	47	25	54	0.32 (0.15,0.68)
Sanchez	32	131	60	140	0.57 (0.40,0.81)
Stoutenbeek	61	202	99	200	0.61 (0.47,0.78)
Ulrich	7	55	26	57	0.28 (0.13,0.59)
Verwaest1	22	193	40	185	0.53 (0.33,0.85)
Verwaest2	31	200	40	185	0.72 (0.47,1.09)
Winter	3	91	17	92	0.18 (0.5,0.59)

In a MA each study provides a contribution to the overall estimate (which is called “weight”). Large studies (and with many events in case of binary variables) have more weight. For example as can be seen in the Figure, Stoutenbeek’s study - with many patients (n=402) and many events (N=162) -carries a weight of 11, 1% while Jacobs’ trial - fewer patients (N=91) and only very few events (N=4) - has a weigh that is only 1.0%.

Results of the studies whose confidence intervals cross the no difference-line are considered non statistically significant while those whose confidence intervals do not cross the no-difference line are considered statistically significant, regardless of whether they are on the side of the beneficial or harmful effect. In the antibiotic prophylaxis example (a total of 16 studies) 13 yield statistically significant results and 3 did not.

The overall estimate is represented by the diamond at the bottom of the graph and the lateral angles of the diamond represent the confidence intervals. The overall estimate is, in the example, 0.28 while the 95% confidence interval goes from 0.20 to 0.38. Thus the combined estimate of this MA suggests that antibiotic prophylaxis can reduces the risk of nosocomial infections by 72%.



Analysis 1.4. Comparison 1 Topical plus systemic versus no prophylaxis, Outcome 4 RTIs.

Review: Liberati A, D'Amico R, Pifferi S, Torri V, Brazzi L, Parmelli E. Antibiotic prophylaxis to reduce respiratory tract infections and mortality in adults receiving intensive care. *Cochrane Database of Systematic Reviews* 2009, Issue 4.

FIGURE 1. Forrest plot of the 16 studies assessing the effectiveness of antibiotic prophylaxis in reducing the risk of nosocomial infections in patients admitted to intensive care

3.1.3 Statistical heterogeneity

In assessing the results of a MA a critical issue is to understand whether studies that have been combined statistically were “similar-enough” thus making the overall estimate the best summary obtainable from the data. This is why - in doing a meta-analysis properly - we need to assess the heterogeneity. The estimate of effect obtained in anyone study may differ from that obtained in another study for the play of chance alone. This may happen even if the two studies used the same protocol (same patients, same treatment, same setting) This variation depends essentially on the number of patient (events): large scale studies tend to produce very similar estimates of effects while small studies can produce even substantially different results. This is called statistical heterogeneity. The presence/absence of statistical heterogeneity can be assessed via formal statistical tests or informal methods.

3.1.4 Informal methods

Two approaches can be used to assess, informally, the presence of statistical heterogeneity.

- Visual inspection of the Forrest plot
- Using the Q value (weighted average of the sum of squares of individual studies estimates and their mean values) and its respective confidence intervals

The Forrest Plot approach is based on the idea that if different studies estimate the same treatment effect then their confidence intervals would tend to overlap. A way to thus assess the presence of heterogeneity would be to explore how much these intervals overlap (see Figure 1).

The second approach is based on the fact that the number of degrees of freedom (d.o.f.) of the test for heterogeneity corresponds to the value that the test would assume if the results would vary among themselves only by the play of chance. A value of the statistical test greater than the number of d.o.f. is an indication that heterogeneity exists while a lower value suggests absence of statistical heterogeneity.

3.1.5 Formal methods: the heterogeneity test

The statistical test used to test the homogeneity hypothesis is called “heterogeneity test”. In the lower part of Figure 1 we see the results of the c.

Chi-square test, often called “Q statistic”, together with its confidence intervals and the p-value.

The results of the Q test are considered statistically significant when the p-value is lower than a predefined threshold defined α , often < 0.05 . When the test is statistically significant we reject the null hypothesis of homogeneity and accept that results do not differ only for the play of chance alone.

In Figure 1 we see a value of the Q equals to 33.77 ($p=0.004$), an highly significant result which would suggest that studies differ more than we would have expected by chance alone, and that therefore the overall estimate may be inappropriate because the heterogeneity is too high.

In such case there are different actions that can be taken.

The first is to perform a subgroup analysis. Authors can group studies according to some meaningful characteristic that they suspect could be the source of heterogeneity (i.e. studies that have focussed on good prognosis vs. those that enrolled mostly poor prognosis patients). We then re-calculate summary estimates of effects within subgroups and see whether results are internally more homogeneous. It would be a marker of a more reliable MA if these types of “heterogeneity exploring analyses” were pre-specified in the SR protocol.

A second approach is to use statistical techniques that incorporate the heterogeneity in the estimates of the pooled effect. The most common technique of this type is called the “random effect model” where both “within-study” and “between-studies” variability is incorporated into the

pooled estimates. For more details about the random effect model readers can go to the Cochrane Handbook (15).

It is important to remember that heterogeneity tests have a low statistical power: when applied to situation where the number of studies (not of patients) is small the test may yield non statistically significant results even if there is heterogeneity. By the same token, when the number of studies is large, even small variability among studies will yield statistically significant results, despite the non-clinically relevant nature of the heterogeneity.

3.1.6 An index to measure heterogeneity (I²)

The I² index measures how much of the observed variability can be explained by inter study differences. It is calculated as $[1-(df/Q)]*100\%$ where d.o.f. are the degrees of freedom and Q the value of the heterogeneity test.

In the example about antibiotic prophylaxis reported in Figure 1 the value of the I² is 56% indicating that half of the observed variability is not due to the play of chance but could rather be attributable to differences of methodological and/or clinical origin.

3.1.7 An example of statistical heterogeneity of questionable clinical relevance

The visual inspection of the Forrest plot (Figure 1) indicates that all point estimates lie at the left hand side of the graph, thus suggesting a protective effect of antibiotic prophylaxis. The result of the Q statistic, however, indicates statistically significant heterogeneity: $Q=33.7$ with 15 d.o.f. ($p=0.004$). Under the homogeneity hypothesis this result would very low plausibility and the variability among studies could hardly be attributed to the play of chance. In this case the value of I² (56%) supports the view that a high proportion of the variability is not due to chance. How did the Authors of the review use these data? They did not take the statistical heterogeneity on the ground of clinical considerations. They based their decision on the following considerations: a) all studies pointed to a beneficial effect of the treatment; b) patients-mix in different studies was composite but was, at the same time, broadly comparable considering the proportional representation of main patients groups (surgical, medical, trauma and others); c) the drugs used in different studies could be considered of comparable effectiveness and in all studies a combination of topical and systemic antibiotics was used. In short, therefore, Authors provided an example of a judgment interpretation of heterogeneity and refused a mechanistic interpretation of the statistical test in favour of a clinical/epidemiological reasons. On the other hand, the review provides an explicit account of the heterogeneity from the statistical point of view and if readers disagree with Authors' interpretation they may still consider inappropriate the pooled estimate and interpret the message of the review differently.

3.1.8 Meta-analysis of observational studies

It is widely accepted that the RCT is the gold standard for assessing the comparative effects of different interventions while it is not so for etiologic, prognostic as well as the assessment of the performance of a diagnostic test.

When we assess the effect of an intervention using RCTs we accept the two following assumptions:

- a) each study offers an unbiased estimate of the treatment
- b) the between studies variability is due to the play of chance.

These two assumptions allow us to say that the combination of unbiased estimates will yield an unbiased overall estimate with a precision that is greater than the one that can be drawn from each component study.

Can we hold the same assumptions for observational studies?

As the estimates are not unbiased (given that the populations were not randomised and selection bias cannot be ruled out) the estimate from a MA are at higher risk of being wrong.

Those notwithstanding there are several examples of MA of observational studies. In those cases results should be interpreted much more conservatively than it is usually done.

Statistical combination of observational studies that aims at assessing the effects of a intervention are not recommended, while SRs of these studies should be strongly recommended.

That said the empirical research that is available on the comparison between MAs of randomised and observational studies had not yield clear results (34-38).

KEY MESSAGES

- 1) The term “Meta-analysis” (MA) refers to the quantitative combination of data from different studies using appropriate statistical techniques in order to have more statistical power to answer a given research question.
- 2) MA is not carried out by summing the number of patients and number of events across all studies but is done calculating a weighted average of the estimate of effect obtained in each study; the weights are determined by the size (number of events) in a study.
- 3) MA's results can be presented both numerically and graphically (Forrest plot).
- 3) Results of individual studies included in a MA may vary for the play of chance or for true differences in terms of study design, type of patients, interventions, settings. This is called “heterogeneity”.
- 4) Heterogeneity is not a drawback in itself but its determinants (statistical, clinical, and methodological) must be understood before we can accept pooled results.
- 5) The interpretation of a MA of observational studies assessing the effect of a interventions is more problematic and, in principle, it is substantially less reliable than a MA based on RCTs.

TABLE 8. PRISMA Checklist: Items to include when reporting a systematic review (with or without meta-analysis)

Section/topic	#	Checklist item
TITLE		
Title	1	Identify the report as a systematic review, meta-analysis, or both.
ABSTRACT		
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants and interventions; study appraisal and synthesis methods; results; limitations; conclusions; and implications of key findings; funding for the systematic review; systematic review registration number.
INTRODUCTION		
Rationale	3	Describe the rationale for the review in the context of what is already known.
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes and study design (PICOS).
METHODS		
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g. web address) and, if available, provide registration information including the registration number.
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility giving rationale.
Information sources	7	Describe all information sources in the search (e.g., databases with dates of coverage, contact with study authors to identify additional studies) and date last searched.
Search	8	Present full electronic search strategy for at least one major database, including any limits used, such that it could be repeated.
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in the systematic review and, if applicable, included in the meta-analysis).
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level, or both), and how this information is to be used in any data synthesis.
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).

continue

continue **TABLE 8**

Section/topic	#	Checklist item
Planned methods of analysis	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).
Additional analyses	16	Describe methods of additional analyses (e.g. sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.
RESULTS		
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citation.
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).
Results of individual studies	20	For all outcomes considered (benefits and harms) present, for each study: (a) simple summary data for each intervention group, (b) effect estimates and confidence intervals, ideally with a forest plot.
Synthesis of results	21	Present the main results of the review. If meta-analysis are done, include, for each, confidence intervals and measures of consistency.
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see item 15).
Additional analyses	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression).
DISCUSSION		
Summary of evidence	24	Summarise the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).
Limitations	25	Discuss limitations at study and outcome-level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).
Conclusions	26	Provide a general interpretation of the results in the context other of evidence, and implications for future research.
FUNDING		
Funding	27	Describe sources of funding and other support (e.g., supply of data) for the systematic review; role of funders for the systematic review.

REFERENCES

1. Stroke Unit Trialists' Collaboration. Organised inpatient (stroke unit) care for stroke. *Cochrane Database of Systematic Reviews* Issue 1 2009.
2. Moore ZEH, Cowman S. Risk assessment tools for the prevention of pressure ulcers. *Cochrane Database of Systematic Reviews* 2008, Issue 3.
3. Moore ZEH, Cowman S. Repositioning for treating pressure ulcers. *Cochrane Database of Systematic Reviews* 2009, Issue 2.
4. McInnes E, Cullum NA, Bell-Syer SEM, Dumville JC, Jammali-Blasi A. Support surfaces for pressure ulcer prevention. *Cochrane Database of Systematic Reviews* 2008, Issue 4.
5. Hayden J, van Tulder MW, Malmivaara A, Koes BW. Exercise therapy for treatment of non-specific low back pain. *Cochrane Database of Systematic Reviews* 2005, Issue 3.
6. Engers A, Jellema P, Wensing M, van der Windt DAWM, Grol R, van Tulder MW. Individual patient education for low back pain. *Cochrane Database of Systematic Reviews* 2008, Issue 1.
7. Light A and Pillemer D. Summing up. The science of reviewing research. Harvard University Press, 1984 Boston.
8. Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG (2007) Epidemiology and Reporting Characteristics of Systematic Reviews. *PLoS Med* 4: e 78.
9. Mulrow CD. The medical review article: state of the science. *Ann Int Med* 1987; 106: 485-488.
10. Clarke M, Chalmers I. Discussion sections in reports of controlled trials published in general medical journals. Islands in search of continents? *JAMA* 1998; 280(3): 280-2.
11. Clarke M, Alderson P, et al. Discussion sections in reports of controlled trials published in general medical journals. *JAMA* 2002 Jun 5; 287(21): 2799-801.
12. Young C, Horton R. Putting clinical trials into context. *Lancet* 2005 Jul 9-15; 366(9480): 107-8.
13. Canadian Institutes of health research (2006) Randomized Controlled Trials Registration/ Application Checklist (12/2006). Available at: http://www.cihr-irsc.gc.ca/e/documents/rct_reg_e.pdf
14. Straus S, Richardson S et al. Evidence based medicine: how to practice and teach EBM. Third Edition, Churchill Livingstone, London, UK 2005.
15. Higgins PT, Green S (eds). *Cochrane handbook for systematic reviews of interventions*. J Wiley Blackwell, West Sussex, England 2008 also available online at <http://www.cochrane.org/training/cochrane-handbook>
16. Nicolucci A, Alexanian A, Apolone A et al. Quality, evolution, and clinical implications of randomized, controlled trials on the treatment of lung cancer: A lost opportunity for meta-analysis. *JAMA* 1989; 262: 2101-7.
17. Thornley B, Adams C. Content and quality of 2000 controlled trials in schizophrenia over 50 years. *BMJ* 1998; 317: 1181-84.
18. Egger M, Smith GD, Altman DG. *Systematic Reviews in Health Care. Meta-analysis in Context*. BMJ Books, London UK 2001.
19. Negrini S, Minozzi S, Taricco M, Ziliani V, Zaina F, Eura A systematic review of physical and rehabilitation medicine topics as developed by the Cochrane Collaboration *Eura Medicophys* 2007; 43: 381-390.
20. Rosenthal R. The "file drawer problem" and tolerance for null results. *Psychol Bull* 1979; 86(3): 638-41.
21. Song F, Eastwood AJ, Gilbody S, Duley L, Sutton AJ. Publication and related biases. *Health Technol Assess* 2000; 4 (10).
22. Chan AW et al. Empirical evidence for selective reporting of outcomes in randomised trials; comparison of protocols to published studies *JAMA* 2004; 291: 2457-65.
23. Chan AW et al. Outcome reporting bias in randomised trials funded by Canadian Institutes of health research *Can Med Ass J* 2004; 171: 735-40.
24. Ghera D. et al. Selective reporting of the primary outcomes in clinical trials: a follow up study Abstract at the XIV Cochrane Colloquium, Dublin 2006.
25. Moher D, Jadad AR, Tugwell P. Assessing the quality of randomized controlled trials. Current issues and future directions. *Int J Technol Assess Health Care*. 1996; 12: 195-208.

26. Juni P. The hazard of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999; 282: 1054-1060.
27. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, et al. Improving the quality of reports of meta-analysis of randomised controlled trials: the QUOROM statement. *Quality of Reporting of Meta-analysis*. *Lancet* 1999; 354: 1896-1900.
28. Vigna-Taglianti F, Vineis P, Liberati A, Faggiano F. Quality of systematic reviews used in guidelines for oncology practice. *Ann Oncol* 2006; 17(4): 691-701.
29. Moher D, Liberati A, Tetzlaff J, Altman DA. Preferred reporting items for systematic reviews and meta-analysis: the PRISMA Statement. *Ann Intern Med* 2009; 151: 264-269.
30. Liberati A, Altman DA, Tetzlaff J, et al. The PRISMA Statement for reporting systematic reviews and metanalyses of studies that evaluate health care interventions: explanation and elaboration. *Annals of Internal Medicine* 2009; 151: w1-w30.
31. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-analysis: The PRISMA Statement. *Phys Ther*. 2009; 89: 873-880.
32. Blyth C: On Simpson's paradox and the sure-thing principle. *J Am Statistical Assn* 1972, 67: 364-366.
33. Liberati A, D'Amico R, Pifferi S, Torri V, Brazzi L, Parmelli E. Antibiotic prophylaxis to reduce respiratory tract infections and mortality in adults receiving intensive care. *Cochrane Database of Systematic Reviews* 2009, Issue 4.
34. Egger M. Spurious precision? Meta-analysis of observational studies. *BMJ* 1998; 316: 140-144.
35. Britton A et al. Choosing between randomised and non-randomised studies. A systematic review. *Health Technology assessment*. 1998; Vol. 2, N° 13 (<http://www.ncchta.org>).
36. Kunz R. The unpredictability paradox: review of empirical comparison of randomised and non-randomised clinical trials. *BMJ* 1998, Vol. 317: 1185-1190.
37. McKee M. Interpreting the evidence: choosing between randomised and non-randomised studies. *BMJ* 1999; 319: 1312-1315.
38. Deeks J et al. Evaluating non randomised intervention studies. *Health Technology assessment* 2003; Vol. 7, N° 27 (<http://www.ncchta.org>).

THORSTEN MEYER¹, MONIKA SCHWARZE²
KERSTIN BIELEIT³, CHRISTOPH GUTENBRUNNER³

QUALITATIVE RESEARCH

¹ Swiss Paraplegic Research, Nottwil,
Switzerland and University of
Luebeck, Institute of Social
Medicine, Luebeck, Germany

² Hannover Medical School, Department
of Rehabilitation Medicine,
Coordination Centre for Applied
Rehabilitation Research,
Hannover, Germany

³ Hannover Medical School, Department
of Rehabilitation Medicine,
Hannover, Germany

WHAT IS QUALITATIVE RESEARCH?

Before going into detail we would like to make an important preliminary remark. Dealing with and doing qualitative research could threaten scientific habits and assumptions that you might have gotten used to, or with which you might have even gotten painfully accustomed during the course of your scientific development. Thus, reading this chapter might come with some “adverse effects”. On the contrary qualitative research offers studies going beyond numbers (1), studies that provide techniques to derive meaning from what we observe. These benefits may outweigh the “risks”.

Characteristics of qualitative research

Qualitative research has been around for a long time. However, it has only recently started to enter the field of health research on a larger scale. E.g. the Journal of Rehabilitation Research explicitly appreciates qualitative studies and has included them as a topic in the information and instructions to authors. Qualitative research has different traditions. One research tradition stems from *ethnography*, i.e. researchers became part of a social group, e.g. a tribe, by taking part in the daily lives of the tribe members, talking with them and observing their behaviors

they developed an insight into the norms, rules, habits, attitudes etc. of the group.

This idea of participatory observation has been taken up by *sociologists* who have entered social groups or institutions (e.g. a psychiatric hospital) and who have tried to understand the norms and rules like a stranger in a tribe. Sociologists have developed a plethora of other qualitative approaches and techniques thus contributing most to furthering methodological developments in qualitative research.

In the field of *psychology* an independent tradition of qualitative research has evolved, partly stemming from the technique of introspection. While these fields of science have been dominated by a so-called quantitative research tradition that is rooted in natural sciences, qualitative methods have found their respective place within these fields. Especially in the applied human sciences qualitative research methods are valued and respected, e.g. educational research, action research, evaluation research, as well as medical research in terms of clinical or health care research (2). The more important the role of people as social beings is for answering a question, the more qualitative methods can contribute.

What are the characteristics of qualitative research? First of all there is no shared definition of qualitative research. Qualitative research is an umbrella term encompassing a wide spectrum of different approaches. One popular approach to defining qualitative research is to contrast it with what is called "quantitative research". We believe that this kind of definition contributes primarily to myths about qualitative research, some of which are discussed below. Quantitative and qualitative research do not represent two distinct research paradigms (3). There are different research paradigms within the field of qualitative research. Also, some qualitative and quantitative approaches might be thought of within one research paradigm. Still, there are some characteristics specific to qualitative research.

The *most fundamental characteristic* is the focus of qualitative research on the *interpretation of meaning*. You could argue that this is what every kind of research should be about. We would not disagree. However, in quantitative studies this is usually not the focus of enquiry. There, interpretation of meaning (of numbers) is restricted to the *discussion* section. Other characteristics specific to qualitative research are:

- *Adaption of research question and methods during the course of the study:* Quantitative research should be characterized by a distinct research question that leads to the selection of an appropriate study design und research methods. In qualitative research, the research question itself might change due to results that emerged during the study. Perhaps the research question ignores other even more interesting phenomena that evolved during the study, or the research question might have been developed on assumptions that can no longer be sustained. For example, the question "*which factors con-*

tribute to good compliance in patients with chronic back pain" might have to be changed, because the concept of compliance does not match the situation the patients have to deal with after the rehabilitation intervention has been terminated. Consequently, the new research question may be: "*which factors contribute to a feeling of self-efficacy and to mastery of daily life by patients with chronic back pain*". In addition, if the research methods are found to be unsuitable for the research question, it is absolutely acceptable and even expected that the research methods are adapted accordingly.

- *Open-format data*: The raw data obtained in qualitative research has an open format. It still has to be translated to codes or categories. Normally data is provided as written texts. While pictures and experiences from participatory observations can also be viewed as raw data, this kind of data should also be translated into written text in order to make them accessible for analysis, e.g. a description of a picture or a documentation of an experience.
- *Analysis as a re-iterative process*: Analysis of quantitative studies can be very straight-forward and can usually be determined prior to the start of data retrieval in a study protocol. On the contrary, qualitative analysis is characterized by repetitively working through the material. For example, a text is read and categories or codes are developed. Then further text is read, categories and codes are added or refined, hypotheses are developed and tested at other parts of the text, are discarded or modified and so on. The results of the analysis can prompt the researcher to gather new data that has to be subsequently analyzed, and so forth.
- *Integration of the principle of induction in the research process*: Induction, i.e. to infer a general statement from a specific statement, is believed to be unscientific in the quantitative research tradition. The basic problem is very simple: the fact that every swan we have seen so far is white does not validate the general statement "all swans are white". You only need one swan of a different color to disprove this sentence. However, qualitative researchers use the principle of induction to build up general statements that are again and again checked against the data in an iterative process. Thus the principle of induction has become a dominant principle in qualitative research. Of course, this does not rule out the danger of false generalizations. Strictly speaking, quantitative researches have to regard the *development of theories* as unscientific. The *prove of theories* can be made scientifically sound. However, it would be senseless to keep the development of theories to an unscientific exercise. Therefore, here qualitative has an important tradition to offer.
- *Exploration and comprehensive theory development*: A major strength of qualitative studies lies in the thorough exploration of a phenomenon. Quantitative researchers appear to value this role of qualitative research the most. Yet, qualitative research has much more to offer. Qualitative research is an excellent approach for de-

veloping theories. Furthermore, these theories are more capable of representing complex interactions between different factors than statistical modeling that is common in the quantitative research tradition.

- *Attitudes, values, and role of the researcher are made explicit and are integrated in the research process:* In quantitative research tradition researchers are eager to keep their own personal characteristics out of the research process in order to get unbiased results. In contrast, qualitative researchers assume that they influence the results anyway. Therefore, the researcher is obliged to explicitly indicate his or her motives and interests for doing the research, as well as experiences with the topic, expected results and possible ways how his or her personal characteristics could influence the research process. By being explicit the researcher can take certain steps to deal with these influences, e.g. by letting a person with a different background conduct the interviews or by analyzing the data in a heterogeneous team. In every phase of the research process the researcher is requested to critically reflect his or her views and values so that their influence on the results and their interpretation can be discerned. This enables the researcher to be more sensitive to results that are contrary to his personal opinions.
- *Search for contradictions:* In quantitative research deviances from average trends are usually thought of as statistical error. Qualitative research is more sensitive towards these deviances. They call for explanations that in turn can enhance the development of a more substantive theory. This is why qualitative researchers are encouraged to actively look for data that is contrary to their own assumptions or hypotheses.

It is tempting to construct a distinction of qualitative versus quantitative studies. This has been done in different textbooks but runs the severe danger of producing myths. E.g. qualitative studies are claimed to be “naturalistic”, quantitative “experimental”. While there are a lot of good examples for quantitative naturalistic studies, e.g. in epidemiology, it is not quite convincing to call semi-structured interviews or focus-groups, which all follow their specific rules of conduct, naturalistic. The following table 1 summarizes the main characteristics of qualitative studies as introduced above. It is set up to convey the notion that the terms qualitative and quantitative research are used as endpoints of what might actually be regarded as a continuum. Also, this list differs from other introductions for the reasons stated above.

Basic assumptions in qualitative research

One central assumption of qualitative research is that realities are socially constructed. This is characteristic of but not specific to qualitative research. The assumption means we can only have a certain perspective on a phenomenon. It depends on our personal characteristics,

TABLE 1. Special characteristics of qualitative studies and their quantitative counterparts

Qualitative research	Quantitative research
Research question	
Might be adapted within the course of the study	Is fixed and defined prior to the study
Raw data format	
Open	Closed
Analysis	
As an iterative process	Should be according to preset hypothesis; or data mining
Principle of research process	
Induction; also induction - deduction cycles	Deduction (in theory)
Important strengths	
Exploration and comprehensive theory development	Numerical representation, generalizability
Attitudes, values, and role of the researcher	
Are made explicit; are integrated into the research process	Are implicit; their role on the research process is not documented
Aberrant cases	
Are treated as substantive data; active search for contradictions	Are usually treated as statistical error

the place and era we live in, or in other words: our socio-cultural context. Each person perceives a phenomenon differently. This is the main reason why qualitative researchers do not search for the “one truth” or are not enthusiastic about “finding laws” that are independent of time and place. Research results are always dependent on a perspective or specific perspectives that should be made explicit to be trustworthy. It also explains why qualitative research should always be a multi-personal endeavor, i.e. the data has to be examined by different people so that a shared understanding of the phenomenon can be discussed from the different perspectives.

There is no straightforward epistemological reason that only qualitative researchers can cling to the idea of multiple (social) realities. In statistical modeling there is also sensitivity for the aspect of constructing realities that should not be mixed up with the truth. You do not have to be a qualitative researcher to adhere to such a constructivist epistemology.

Another myth about qualitative research that still prevails in the literature is that it is *naturalistic*. The idea still exists that qualitative re-

search is conducted in natural settings and the researcher comes into contact with the research subject in their natural environment. In fact, most qualitative research studies in health care are conducted by means of open or guided interviews or focus groups. Both situations are first of all artificial. They are set up to provide a special environment for eliciting talk. It is rather unnatural to talk with a group of strangers on intimate topics following a set of rules for showing mutual respect, listening to others, waiting until their lines of thought have come to an end etc. The methods this myth refers to is participatory observation, or even document analysis (see below). Note that both methods are uncommon in qualitative health care research. On the other hand, there are quantitative studies that manage to be very naturalistic, e.g. pragmatic trials or cohort studies.

Research questions

Qualitative research in health care is especially used to reproduce and analyze subjective perceptions of different stakeholders, i.e. patients, physicians, other health professionals, relatives or caregivers etc. Patients' beliefs about their illness or disability can be explored (causes, duration and time courses, interventions deemed necessary, social consequences). The decision making process of health professionals in routine practice can also be explored and analyzed.

A second field of enquiry relates to the description and analysis of social interactions in health care, e.g. interactional styles and role behavior of physiotherapist and patients in rehabilitation centers, characteristics of team interaction in relation to staff satisfaction, analysis of organizational processes within institutions or the interplay of different institutions that deal with rehabilitation clients within the system of care.

Qualitative research methods are especially suited for specific research question. The least disputed questions are explorative questions that help to identify essential aspects of a problem to address. Qualitative research methods are also more appropriate in complex research questions involving various facets or levels that are strongly interdependent. In these cases quantitative studies are hardly capable to provide adequate representations.

Implementation studies are another important field of application of qualitative research methodology. Here, the implementation of a new health technology in health care and the analysis of varying adoptions of the technology and different conditions of their adoptions are of great value.

In general, qualitative research methods are well suited to integrate the perspective of subjects in studies, e.g. experiences of care-givers for mentally ill persons. They also provide multi-perspectivity, i.e. the integration of different perspectives of all relevant stakeholders. In addition, qualitative studies allow us to study populations of people or individuals who are unable to be included in quantitative studies (so called

“hard-to-reach” persons), e.g. the exploration of need for medical care or rehabilitation interventions in illegal immigrant populations, or exploration of rehabilitative needs of drug abusers who are unable to be admitted to a regular rehabilitation unit. In qualitative research there is a tradition to call attention to deprived social groups, and to give these people a public voice that they would lack otherwise (“giving-voice”), e.g. dementia patients without relatives or social support, or persons suffering from sexual abuse.

DATA COLLECTION METHODS

There are four main types of data collection methods in qualitative health research. They include individual interviews, group interviews (especially focus groups), observation and document analysis.

Individual interviews

Interviews are characterized by a direct interaction between an interviewer and an interviewee. Individual interview allow for in-depth exploration of personal topics. They can provide very rich material. However, they are laborious and time-consuming. Therefore they can only be used in a restricted number of persons. The interview can be semi-structured, i.e. the interviewer has certain questions prepared and will work through a couple of questions or topics during the interview. Depending on the degree of the preset structure of questions or topics, the interviewer does not have to follow a specific order, and can be flexible with the phrasing of questions.

Interviews can also be almost completely open: here the interviewer provides a stimuli to talk and tries to be as passive-receptive as possible during the interview. The interviewer has the task to keep the flow of words going. This approach is often used in biographical enquiries and narrative interviews, e.g. by letting interviewees talk what they thought and experienced when they were confronted with a diagnosis of paraplegia. In these open interviews the selection of topics by the interviewee can be an additional object of analysis.

In health care, subjects are not only patients but often experts (“expert interviews”). Here, the expertise of a health professional representing an expert in a specific field of knowledge is used to answer research questions. Also, patients with chronic disorders or disabilities can be approached as experts of their own conditions.

Group interviews

Group interviews represent a more efficient data collection method than individual interviews. They are suitable if there is a need for a broader anchoring of a topic that individual interview would allow. However, the individual perspective may be substantially reduced and there is very limited opportunity to go into detail in single participants.

A clear-cut advantage of group interview is to use the group dynamics that are always involved in group discussions. Especially questions relating to attitudes or judgments might be more pronounced in group discussions where pros and cons of a topic can emerge more easily than in individual interviews. For example, an appraisal of the introduction of a new health technology could make use of group dynamics.

Focus groups are a special type of group discussions. They originated in marketing research but have been established in the social sciences, too. These groups usually focus on a specific topic, invite participants who do not know each other before hand, have usually around 6-8 participants who agree on certain rules of conduct, last from 1 - 2 hours, and are usually tape-recorded or additionally video-recorded. It is helpful to have a group moderator and a second person being attentive to missing answers or participants not getting the opportunity to share their views adequately. A practical introduction into focus group is given by Krueger & Casey (4).

Observation

Observations are a corner-stone of qualitative enquiry. However, their importance has declined compared to interview approaches in practical qualitative health research. Scientific observations are different from observations in everyday life. They serve a specified research aim, are planned and documented systematically, and are tested for validity, reliability and scrutiny. There are different types of scientific observations:

Open vs. covered observations - When the person to be observed knows that he is observed we refer to this as an open observation. When the person to be observed is unaware of being observed this is called a covered observation.

Participatory vs. non-participatory observation - Ethnology introduced the concept of participatory observation. The researcher is or becomes part of the social group he wants to research. This approach is especially suitable for enquiries into sub-cultures or for the analysis of complex social interactions and processes. In non-participatory observation the observant is detached from the situation, e.g. by means of a video-documentation of patient-physician encounters.

Structured vs. open observations: Observations can differ markedly by the degree of how fine-grained an observation scheme is elaborated. On the one hand there might be a pre-defined classification scheme focusing on different aspects of social interactions. On the other hand there might be a researcher taking notes after the situation he has observed has ended.

Field- vs. laboratory conditions - Observations can be made within in a laboratory setting. There it is easier to control for possible influences that might otherwise disturb the observation. However, observation in the "field", i.e. under natural occurring conditions, has the great advantage of being close to everyday life.

Observation by others vs. self-observation - Self-observation (introspection) has become a rare phenomenon in social sciences. In psychology it had been used as an important research method for some time but has lost its importance there, too. Therefore, most observational studies are third-party observations.

Document analysis

Document analysis is part of so-called *non-reactive research methods* or *unobstrusive measures*. This term is used for methods in which the person analyzed does not know that his or her behavior will become part of a scientific analysis. This means non-reactive research methods should be free of social desirability biases - to a certain degree. The way a health professional fills in a patient chart is determined by specific social codes. The analysis of minutes of meetings are based on statements of persons that are not just deliberately say what comes to their mind but what might be appropriate by the social role taken up in the meeting. This is why it is very important in document analysis to reflect on the socio-historical context of the development of the document. The big challenge of document analysis and of unobstrusive measures per se is to give a sound reasoning on the relationship between the indicators of behavior present in the documents and the research question to be answered. Document analysis has become more popular in recent years due to the increasing digitalization and accessibility of documents in the internet along with the increasing importance of the internet in the daily lives of people.

DATA ANALYSIS METHODS

As mentioned above, data in qualitative research is usually transformed to texts. This transformation is called transcription. The transcript is then the written basis of analysis.

Transcription

Most often data in qualitative health care data represents oral language. There are different levels of detail a transcript of oral language can adhere to. It has to be decided at the outset whether para- or non-verbal aspects of speech should be transcribed. Also, there should be clarity how to deal with dialect words. In the simplest cases, the words said are transcribed into standard language without consideration of dialect idioms or para- or non-verbal expressions. This might be sufficient in expert interviews or in situations where the researcher is primarily concerned with what the interviewees have said without going too deeply into interpretative work. If para- or even non-verbal aspects are deemed important for analysis then there should be a clear guidance how they are represented in the transcript. There are different ways to do that. The following list of transcription symbols in table 2 is taken from the textbook of Denzin & Lincoln (5).

TABLE 2. Transcription symbols in conversation analysis (taken from (6), page 882)

(Starting point of overlapping speech)	End point of overlapping speech
(2.4)	Silence measured in seconds	(.)	Pause of less than 0.2 seconds
↑	Upward shift in pitch	↓	Downward shift in pitch
word	Emphasis	wo:rd	Prolongation of sound
°word°	Section of talk produced in lower volume than the surrounding talk	WORD	Section of talk produced in higher volume than the surrounding talk
w#ord#	Creaky voice	£word£	Smile voice
wo(h)rd	Laugh particle inserted within a word		
wo-	Cut off in the middle of a word	word<	Abruptly completed word
>word<	Section of talk uttered in a quicker pace than the surrounding talk	<word>	Section of talk uttered in a slower pace than the surrounding talk
(word)	Section of talk that is difficult to hear but is likely as transcribed	()	Inaudible word
.hhh	Inhalation	hhh	Exhalation
.	Falling intonation at the end of an utterance	?	Rising intonation at the end of an utterance
,	Flat intonation at the end of an utterance		
word.=word	"Rush through" without the normal gap into a new utterance	((word))	Transcriber's comments

Qualitative content analyses

Content analysis is a technique that has its origin in communication sciences. It has also a quantitative tradition, e.g. by counting special phrases in newspaper articles or politicians' talks. Qualitative content analysis can aim at the content of the text, formal characteristics of the text or - which is the most challenging and interesting - latent, i.e. unobserved meanings that have to be distilled during analysis. Therefore, qualitative content analysis is a truly interpretative technique. It follows some basic principles:

- The analysis has to take into account that the text is not a product of itself but has been generated in a social communication context. Content analysis should take this context into account: Who is the author? What are the characteristics of the material? Who is the addressee of the talk? Which interests might have played a role?
- Scientific analysis is systematic analysis. Before doing a qualitative analysis there should be a good rationale which approach or theory

you want to apply, which rules are used to structure and analyze the text and which steps the analysis will follow.

- Different authors have differing views about the application of quality criteria in content analysis. If you develop a category system that you use to code further text material it can be useful to do an inter-rater-reliability analysis, i.e. to show how good independent coders agree in their coding. This approach can also be used to identify parts of a category system that are not defined precisely resulting in disagreement between coders.
- Qualitative content analysis might result into quantifying statements. However, there are only a few special cases where such a quantification proves to be meaningful. The mere frequency of a phrase does not have to directly relate to its relevance. In addition, only very few qualitative studies are set up in a way to allow generalization in terms of statistical representativeness.

There are four types of content analyses that differ in their aims and procedures (7): content analysis for summarizing; inductive formation of categories; content analysis for explication; and content analysis for structuring.

Content analysis for summarizing is applied when comprehensive text should be reduced in a way that the main content is preserved. It is useful when the focus lays on the manifest content of a text. It can also contribute to a subsequent qualitative analysis e.g. by providing summary abstracts from each case or each interview conducted. These vignettes can help to gain a good overview of the texts.

Inductive formation of categories is a central technique in qualitative content analysis. Having your research question in mind you have to go through the text and start to identify everything that might be related to the question. Every piece of relevant text is described in your words. By reading more and more texts or different interviews there usually appear similar themes and topics. You have to develop codes or categories that are able to capture all relevant aspects that have emerged. In the end you might have a multitude of different aspects to your question. But you can be sure they have emerged from the text and are not - at least to a certain degree - the product of your expectations or theoretical knowledge on the issue. These codes or categories can then be further analyzed with regard to their commonalities or discrepancies, with the aim to develop underlying explanatory categories that could represent the main result of your study.

Content analysis for explication is applied when specific parts of the text have to be explained thoroughly. These might be unusual word choices or astonishing reactions of the interviewee. The task is to use other context data to make the part under investigation understandable. This context data can be the transcript itself (narrow analysis) or data that goes beyond the transcript is included in the analysis (wide analysis), e.g. further documents, background material on the interviewee or information on the socio-cultural context of the interview situation. The

aim of this analysis is truly interpretative. It can represent a major part of a qualitative analysis, e.g. by identifying statements or words that cannot be accounted for by present theoretical backgrounds and lead to a wider or new theoretical perspective.

The types of content analyses introduced so far have something important in common. They use the text as their starting point of analysis. On the contrary, the **content analysis for structuring** uses prior defined schemes or frameworks that are used to analyze the data. There is a very special perspective that is used to structure the data accordingly. This perspective can look at formal aspects of the text, at special content, or even rating scales can be applied to appraise certain characteristics of the text. It becomes clear that this type of analysis should only be applied in much focused research questions.

Grounded theory

Grounded theory is more than a data analysis method. It is a comprehensive and consistent framework to do qualitative research. It would be impudent to give a proper introduction into grounded theory within the space available. The interested reader is referred to the textbooks of Glaser & Strauss (8) or to a short introduction of *what grounded theory is not* (9). One main point of grounded theory is its focus on developing theory, i.e. qualitative enquiry should have the aim to build theory. A thick description of a phenomenon might be a prerequisite of such an analysis, but we should pursue to learn something from these phenomena by developing theory. The theory itself should be developed iteratively, step by step, with a truly inductive approach to begin with, i.e. grounded in the data. An extensive example of a grounded theory project of interest to rehabilitation has been published by Glaser & Strauss where they analyzed illness trajectories of chronic ill persons (10).

Computer-assisted analysis

There are a number of different software tools to assist qualitative analysis. However, they do not take over the most central work of a qualitative researcher: doing interpretation. There is no ghost in the machine that comes up with surprising results. But this software can help to organize texts, codes, memos and the like in a very efficient way. For example they offer graphical displays that help to visualize associations that the analyst has defined between codes or categories. They also offer the opportunity to do frequency analysis with coded categories. Examples for wide-spread software programs are Maxqda, NVivo or Atlas/ti (11).

QUALITATIVE RESEARCH IN CONTEXT

How scientific is qualitative research?

Qualitative research has always been accused for being too subjective or even unscientific. The motivation in quantitative research is

to exclude as much of subjectivity of the researcher as possible. Subjects are seen as a source of nuisance or bias in the quest for objectivity. Qualitative research to the contrary assumes that subjectivity is inevitable. Therefore, it tries to integrate subjectivity into research, by appealing to the researcher to unfold his or her background, interests and motives in doing the research; by integrating multiple views on a research topic (multi-perspectivity); by actively seeking for evidence which is contrary to one's own expectation (and by that adhering to Popper's idea of falsification that can make a hypothesis even stronger).

There is a long discussion on validity of qualitative research. Criteria for good qualitative health research have been suggested by: Greenhalgh & Taylor (1), Grenn & Britten (12) and Mays & Pope (13) in the British Medical Journal; Kitto et al. (14) in the Medical Journal of Australia; Giacomini et al. (15, 16) in the Journal of the American Medical Association; Malterud (17, 18) in the Lancet; Morse et al. (19) and Sandelowski & Barroso (20) from the perspective of nursing sciences; Murphy et al. (21) from the perspective of health technology assessment; Tong et al. (22) from a Quality in Health Care perspective; and Cohen & Crabtree (23) from the perspective of family medicine. The latter publication is of interest because it aims at summarizing and integrating published quality criteria of qualitative research. The result is displayed in table 3.

TABLE 3. Quality criteria of qualitative research (according to Cohen & Crabtree 2008 (23))

Qualitative research should ...
1. be ethical (i.e. respectful, humane, honest, embodying values of empathy, collaboration, and service);
2. be important (i.e. pragmatically and theoretically useful, advancing current knowledge base);
3. be clearly and coherently articulated (reports should be concise and provide a clear and adequate description of the research question, background and contextual material, study design, rationale of methodological choices, relationship between data and interpretation);
4. use appropriate and rigorous methods;
5. use reflexive processing by the researcher; or consider researcher bias;
6. be valid or credible, adequate, trustworthy, accurate, plausible;
7. be verifiable or reliable.

A thorough look through these criteria reveals that they are not specific to qualitative research but are more appropriate called criteria of empirical research per se - being it qualitative or quantitative. This is one expression of the fact that qualitative and quantitative research cannot be simply discerned into two research paradigms. They are both

overlapping and there are also different paradigms apparent within these research traditions.

An introduction of what should be reported in a qualitative research publication has been published by Elliot et al. (24).

Qualitative research in an evidence-based world

Evidence-based medicine (EBM) means the integration of the best available research evidence with clinical expertise and patient values into clinical decision making (25). It originated in the field of clinical epidemiology. Therefore, it is not surprising that qualitative studies have not been valued or integrated in the idea of evidence-based medicine. E.g. they cannot be found in the "hierarchy of evidence" table of the Oxford Centre of Evidence-based Medicine (see www.cebm.net). However, there are different ways how qualitative studies fit into an evidence-based world. EBM starts with question of practical relevance. There are questions that are more appropriately answered by qualitative research, e.g. what are the motives, interests or preferences of the patients? Why do patients not take their prescribed medicine? Besides being a source of evidence in their own right qualitative studies can be used to test the feasibility of concepts of conduct of quantitative studies or to deepen the insight into results of quantitative studies, especially when the results are contrary to expectations. They can also be used to get a good picture of patient preferences to prepare for clinical or system level decisions on care. They are also useful to analyze decision-making and health care processes, including the implementation of EBM into health care, e.g. "why do physicians abstain from using clinical guidelines in decision making?", "Evidence does not make decisions, people do." (26, p. 1350). Qualitative research is well suited to analyze these decision making processes.

Learning qualitative research

There are different challenges in doing and learning qualitative research. First of all, it is easy to get lost when there is a lot of text as the typical basis of analysis. The challenge is to organize the data and develop an understanding of what the unit of analysis should be and distinguish the important and irrelevant aspects of the data. Second, the task of interpretation is a process that needs both rigor and creativity. In this vein qualitative analysis is quite more challenging than most applications of statistical methods. Third, a lot of decisions have to be taken during analysis and there is also a great need for multi-perspectivity. It appears to be reasonable to conduct qualitative research in a research group lead by an experienced qualitative researcher. I.e. qualitative research should be taught by researchers themselves and not solely by book. Also, the conduct of a qualitative study should be integrated into a qualitative research groups. Unfortunately, there are a lot of bachelor, master or doctoral theses conducted by a single person with dearth supervision. They are often characterized by insufficient training and missing integration into a respective research group.

Qualitative research in the rehabilitation sciences

Specific aspects

Physical and Rehabilitation Medicine aims at improving functioning in patients with disability (or experiencing disability) within context of health conditions (27). This includes the diagnosis of the underlying pathology and the assessment of functioning according to the definition given by within the ICF-model (28). Due to this complex model a large spectrum of interventions are needed ranging from medication and physical modalities to psychosocial interventions and patient education (29, 30). This requires interdisciplinary team processes that have to integrate the patient's perspective and his or her individual goals and concepts. Last but not least contextual factors such as the physical and social environment and personal factors have also to be taken into consideration within the rehabilitation process.

Due to this conventional study designs that mainly have been developed for the evaluation of drug treatments are only partly applicable for the evaluation of interventions, programmes or services in Physical and Rehabilitation Medicine. This is the case especially for comprehensive and multiprofessional rehabilitation programs. Reasons for this are e.g.

- complex interactions of the dimensions of functioning (body structures and functions, activities and participation) and with the environmental factors
- multidimensional interventions
- interventions that cannot be blinded
- individual goal setting taking into consideration the person's own concept
- continuous adaptation of the applied program during the rehabilitation process

The mostly used study designs as randomized controlled studies (including studies using naturalistic designs) are appropriate to decide if an intervention is effective as compared to negative controls (e.g. placebo, sham treatment) or as effective as positive controls (standard treatment) (6). Here the methods tested have to be elaborated in advance. The development process of such interventions strategies needs other methods ("to find out how to intervene"). On the other hand probability statistics are used to evaluate the validity of diagnostic tools such as assessment questionnaires can be investigated (7). Such studies are not suitable to develop new intervention strategies or to find out what dimensions may be measured in patient-oriented outcome assessment instruments. Here too the selection of relevant dimensions to be measured needs other scientific strategies ("to find out what to measure").

For these reasons qualitative research methods are of major importance in Physical and Rehabilitation Medicine especially to solve the following problems (8):

- to evaluate the patients' perspectives in rehabilitation including evaluating the experience of disability and the individual rehabilitation goals

- to figure out relevant dimension for the design of assessments in rehabilitation (this includes the relevance of different aspects of a patient's perceived quality of life)
- to develop new rehabilitation strategies especially if complex structures of the rehabilitation systems are involved
- to analyze and to improve communication within rehabilitation teams and people to be rehabilitated or treated

Qualitative research in the rehabilitation literature

How are qualitative studies represented in the rehabilitation sciences literature? An informal survey of publications¹ within the last three years of five prominent international journals in field of rehabilitation medicine and research done by the authors offers valuable insights. Only a few studies used qualitative approaches, between 0 and 10% of all original articles. There were two main focuses of research areas and questions: 1. description and analysis of individual perspectives, experiences or attitudes of patients and carers on health or health-care related issues; 2. development of new interventions or innovative approaches to care delivery. The majority of papers dealt with neurological disorders, especially stroke and spinal cord injuries, followed by pain, mental disorders and other neurological aspects. Participation problems were elaborated in wheelchair users and in work related issues.

From a methodological point of view it should be noted that (semi-structured) individual interviews and focus groups were the most prominent approach to data acquisition. Observations and concept mapping was used only in single publications. One study used the technique of meta-synthesis to provide a systematic overview on different qualitative studies, concerning the experience of living with stroke. Content analysis was the predominant method of data analysis, as well as approaches in reference to grounded theory. Some of the studies used a qualitative approach; some used a combination of quantitative and qualitative methods in the study design.

It is obvious that qualitative research has only started to be represented in journals of rehabilitation medicine. The focus seems to be on patient groups with severe and long-term disorders or disabilities. As expected, there was hardly any study applying methods of participatory observation or even document analysis. Also, the analysis of interaction of different stakeholders within the rehabilitation process is not represented. There are much more facets of qualitative research that those represented in the present literature of rehabilitation medicine. We would like to encourage researchers to utilize this potential!

¹ Journal of Rehabilitation Medicine, European Journal of Physical and Rehabilitation Medicine, American Journal of Physical Medicine and Rehabilitation, PM & R, International Journal of Rehabilitation Research (June 2007 to June 2010).

EXAMPLE OF QUALITATIVE STUDIES

Medina-Mirapeix et al. Personal characteristics influencing patients' adherence to home exercise during chronic pain: a qualitative study. *J Rehabil Med* 2009 (31)

Medina-Mirapeix et al. identified the beliefs and perceptions of patients with chronic neck and low back pain that influence adherence to home exercise during exacerbation and/or remission of pain. As a data collection method they used the focus group technique with patients participating in a home exercise programme. Seven groups of patients were interviewed about how they perceived their adherence to a home exercise programme during chronic pain. Data were analyzed using descriptive and categorizing methods. As results several themes about patients' beliefs and perceptions were identified as factors related to adherence. These factors changed when pain or disabilities was present or had disappeared for a longer time.

As an example three citations about patient's beliefs about illness in the in the categories: "when pain or disabilities appears", "when pain disappears for long time" and "when pain or disabilities decrease" are listed:

IP 16: "I have my problem since so many years and nobody could help me. Because that often I don't do advice of the brochure that physical therapist gave me" (Female, 63 years).

IP 3: "I exercise because I am afraid that I will have the pain again (Male, 53 years)".

IP 20: "When I feel better, I forget the exercises and do other things; besides that, I don't have pain again" (Male, 35 years).

In clinical practice as PMR specialists we treat a lot of patients with chronic pain. Adherence to home exercise is an essential prerequisite for successful treatment. However, in clinical practice it is often difficult to judge the level of adherence of a patient and to focus our attention on possible factors that might contribute to adherence. This study, exemplified by the citations of three different patients, highlight the important role that patients beliefs about their illness and its course are a major determinant in adherence to home exercise. The results of this study made us thought about exploring the beliefs about what causes the pain, how stable the symptoms are perceived and how they can be affected by exercise or other interventions in a systematic way. By this way we might be not only concerned with encouraging patient to do their exercise but with talking about their beliefs openly and trying to change them in a favorable manner.

CONCLUSIONS

All in all qualitative methods are of major importance in Physical and Rehabilitation Medicine to respond to the need of patient-centered intervention approaches, to involve the individual's perspective, to optimize rehabilitation programs and to design innovative intervention approaches taking the complexity of functioning and disability and respective interventions into account.

REFERENCES

1. Greenhalgh T, Taylor R. Papers that go beyond numbers (qualitative research). *BMJ*. 1997; 315: 740-743.
2. Pope C, Mays N (eds.) *Qualitative research in health care*, 3rd ed. Oxford: BMJ books; 2006.
3. Rost J. Drei Thesen zum Konzept qualitativer Forschungsmethoden. *Zeitschrift für Didaktik der Naturwissenschaften*. 1998; 4: 35-42.
4. Krueger RA, Casey MA. *Focus groups: a practical guide for applied research*, 3rd ed. Thousand Oaks: Sage; 2000.
5. Denzin NK & Lincoln YS (eds). *The Sage handbook of qualitative research*, 3rd ed. Thousand Oaks: Sage; 2005.
6. Peräkylä, A. Analyzing talk and text. In: Denzin NK & Lincoln YS (eds). *The Sage handbook of qualitative research*. Thousand Oaks: Sage; 2005, p. 869-886.
7. Mayring, P. *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. 8. Auflage, Weinheim: Beltz; 2003.
8. Strauss AL, Corbin J. *Basics of qualitative research*, 2nd ed. Thousand Oaks: Sage; 1998.
9. Suddaby R. From the editors: what grounded theory is not. *Acad Manage J* 2006; 49: 633-642.
10. Corbin J, Strauss AL. *Unending work and care: managing chronic illness at home*. San Francisco: Jossey-Bass; 1988.
11. Lewins A, Silver C. *Using software in qualitative research*. London: Sage; 2007.
12. Grenn J, Britten N. Qualitative research and evidence based medicine. *BMJ* 1998; 316: 1230-1232.
13. Mays N, Pope C. Qualitative research in health care: assessing quality in qualitative research. *BMJ*. 2000; 320: 50-52.
14. Kitto SC, Janice Chesters J, Grbich C. Quality in qualitative research. *Med J Aust*. 2008; 188: 243-246.
15. Giacomini MK, Cook DJ, for the Evidence-Based Medicine Working Group. Users' Guides to the Medical Literature. XXIII. Qualitative Research in Health Care. A. Are the Results of the Study Valid? *JAMA* 2000; 284: 357-362.
16. Giacomini MK, Cook DJ, for the Evidence-Based Medicine Working Group. Users' Guides to the Medical Literature XXIII. Qualitative research in health care. B. What are the results and how do they help me care for my patients? *JAMA* 2000; 284: 478-482.
17. Malterud K. The art and science of clinical knowledge: evidence beyond measures and numbers. *Lancet*. 2001; 358: 397-400.
18. Malterud K. Qualitative research: standards, challenges, and guidelines. *Lancet* 2001; 358: 483-488.
19. Morse M, Swanson J, Kuzel AJ. *The nature of qualitative evidence*. Thousand Oaks: Sage; 2001.
20. Sandelowski M, Barroso J. Reading qualitative studies. *Int J Qual Methods* 2002; 1: 74-108.
21. Murphy E, Dingwall R, Greatbatch D, Parker S, Watson P. Qualitative research methods in health technology assessment: a review of the literature. *Health Technol Assess*. 1998; 2: iii-ix, 1-274.
22. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care*. 2007; 19: 349-357.
23. Cohen DJ, Crabtree BF. Evaluative criteria for qualitative research in health care: controversies and recommendations. *Ann Fam Med*. 2008; 6: 331-339.
24. Elliott R, Fischer CT, Rennie DL. Evolving guidelines for publication of qualitative research studies in psychology and related fields. *Br J Clin Psychol*. 1999; 38: 215-229.
25. Sackett DL. *Evidence-based medicine: how to practice and teach EBM*, 2nd ed. Edinburgh: Churchill Livingstone; 2000.
26. Haynes RB, Devereaux PJ, Guyatt GH. Physicians' and patients' choices in evidence based practice. Evidence does not make decisions, people do. *BMJ*. 2002; 324: 1350.

27. Stucki G, Melvin J. The International Classification of Functioning, Disability and Health: a unifying model for the conceptual description of Physical and Rehabilitation Medicine. *J Rehabil Med.* 2007; 39: 286-29.
28. World Health Organisation. International Classification of Functioning, Disability and Health: ICF. Geneva: WHO; 2001.
29. Gutenbrunner C, Ward AB, Chamberlain MA (eds.). White Book on Physical and Rehabilitation Medicine in Europe. *Eura Medicophys* 2006; 42: 287-332; *J Rehabil Med* 2007; 45: 1-48.
30. Gutenbrunner C, Lemoine F, Yelnik A, Joseph PA, de Korvin G, Neumann V, et al. The field of competence of the specialist in Physical and Rehabilitation Medicine (PRM). *Ann Phys Rehabil Med.* 2010; in press.
31. Medina-Mirapeix F, Escolar-Reina P, Gascon-Canovas J J, Montilla-Herrador J, Collins SM. Personal characteristics influencing patients' adherence to home exercise during chronic pain: a qualitative study. *J Rehabil Med.* 2009; 41:347-352.

ELENA GIGLIA¹
NIKOLAOS BAROTSIS²
XANTHI MICHAEL³

INTERNET RESOURCES FOR PHYSICAL AND REHABILITATION MEDICINE: DATABASES, SEARCH ENGINES, WEBSITES, E-BOOKS, IMAGES AND THE TOOLS TO ORGANIZE THEM

¹ University Library System,
University of Turin, Italy

² Physiatrist, PRM Clinic,
Naxos, Greece

³ Professor of Rehabilitation Medicine
in Physiotherapy Department,
Technological University,
Athens, Greece

Aim of this contribution is to present some useful Internet resources in order to run effective searches in PRM. We shall consider both some searching strategies within well-known tools such as PubMed or PEDro and some interesting features of other databases or websites. You will find the references within the text, with the corresponding URL. Some few notes are added in order to give in-depth suggestions. As the choice of the most useful database depends on each single information need or question, it's impossible to define a list of the "better" tools. We tried to organize the material according to topics, in order to orient the user.

1. PUBMED: HOW TO PERFORM AN EFFECTIVE SEARCH, SAVE IT AND AUTOMATICALLY RECEIVE UPDATES

PubMed (<http://www.pubmed.gov>) is the worldwide best known and accessed service for biomedical researchers. It is the free interface to MEDLINE, the bibliographical database edited by the U.S.

National Library of Medicine (<http://www.nlm.nih.gov/>), with its over 20 million citations since 1947 from about 5400 journals in 38 languages. MEDLINE became freely accessible via Web in 1997, and it is the core of PubMed; yet, in the latest years PubMed has been integrating free databases, free full texts, and value-added services we are going to explore. PubMed has been developed by NCBI (National Centre for Biotechnology Information, <http://www.ncbi.nlm.nih.gov/>) and also gives access from a pull-down menu in homepage to all the free NCBI databases dealing with genes, proteins, molecules, substances (fig. 1).

Like every easy, user-friendly application, PubMed is often only partially used, not exploiting all its potentiality: the user types a word in the Search Box, and clicks the “Search” button. Often thousands of items - not all relevant - are displayed, and the reader immediately gets discouraged. Actually, there are some useful tools to run an effective search we are going to explore. Moreover, the traditional PubMed interface we all were used to search is no longer available since October 27th, 2009, so we shall see the developments as to date, because they are yet in progress.

1.1 PubMed homepage

Essential, streamlined, the new PubMed homepage (fig. 1) is structured in:

- a blue NCBI header, with a pull-down **Resources** menu for all the Entrez (<http://www.ncbi.nlm.nih.gov/sites/gquery>) databases (see below, 2.5), clustered by topic; a new **How to** menu, with these databases instructions; and the traditional **My NCBI Sign in**, always in the upper right end (see below, 1.6);
- the Search Bar, with the **Search Box** with a **Search button** instead of the Go button; the pull down selection menu for the active searched database; the links to **Limits**, **Advanced search** and **Help** and, after a run search, the **RSS** feed and **Save Search** links;
- a three-column point of access to the former PubMed blue sidebar tools, labelled respectively **Using PubMed** (for Quick Start, Animated Tutorials with audio, Tutorials in pdf, News...), **PubMed tools** (Single citation Matcher, Clinical Queries, Topic-Specific queries, i.e. former Special queries); **More resources** (MeSH database, Journals database, Clinical queries, E-utilities);
- a footer, which includes links to other NCBI resources (e.g. PubMed Central, Bookshelf, Human Genome) help, tools, contacts, not shown in the figure.

1.2 Basic search: how the system works

To run a search in PubMed, just type your keywords in the Search Box and press “Search”.

A list of results is then displayed, indicating the authors' names, the title of the article, the title of the journal which contains the article, with

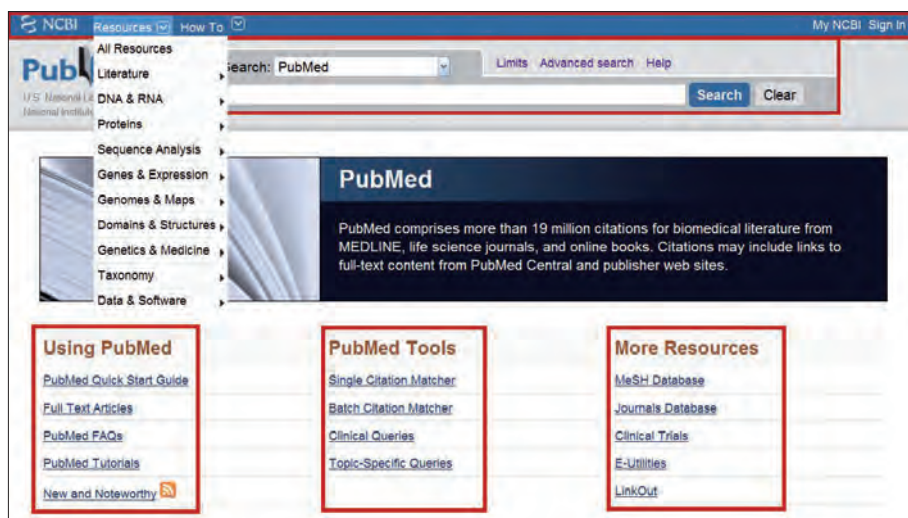


FIGURE 1. PubMed New Homepage with NCBI header, search bar, and tools columns.

all its bibliographic references (year, volume, issue, pages), the *status* indications that contains basic information such as [In Process] or [As supplied by publisher], which means that the recently added article is being indexed so it has not yet a matching MeSH term as the [Indexed for MEDLINE] ones (see below, 1.3).

Display settings can be modified by the corresponding link which appears on the top of the results list: you can choose how many items per page, their format, and the criteria to sort them (Recently added, Publication date, Authors' name...).

From the results list, you can scroll on the right side of the screen (fig. 2):

- **Filter your results**, provides a Manage filter to quickly vary selections. By default are set now: All, Free full texts and Reviews. They can be set in MyNCBI (see below, 1.5);
- **Titles with your search terms**, i.e. what are supposed to be the most pertinent items;
- **Try also**, pre-set suggestions to refine the search;
- **Free full text in PubMedCentral**: it highlights the Open Access (see below, 2.3) papers freely accessible in PubMedCentral;
- **Find related data**: a related search within all the Entrez databases (see below, 2.5);
- **Search details**: available right here or accessible from the Advanced search link near its Search Box. Remind that the Details feature is very suitable because it lets you know how the system translated your query. The Search Details box being as usual rewritable, it allows you to refine your query just by adding or changing e.g. a Boolean operator or by deleting or adding a term;
- **Recent activity**, which tracks your searches and your downloads. Each one of these sections is collapsible, if not applied.

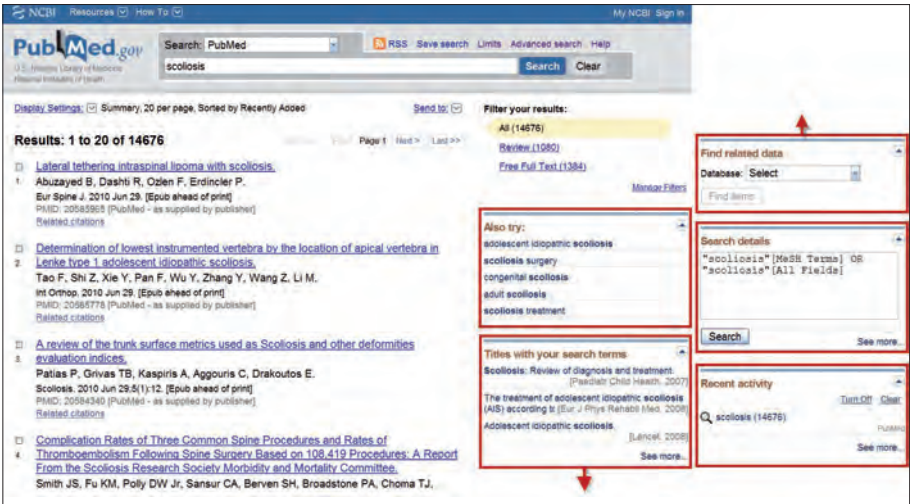


FIGURE 2. PubMed results list with the right side options (composed).

A useful **Auto suggest** feature appears while typing in the Search Box.

Being active the “**Sensors**” introduced in 2008, if the typed keyword refers to a gene or a protein a yellow box is shown as a gate to other NCBI or external databases supplying genes or protein additional information or sequences or tools.

In the same way, being active another “Sensor” for journal titles, if you type “Negrini scoliosis” the system automatically runs a query for “Negrini” as Author AND “scoliosis” as MeSH term or journal title, clustering the results for the journal in a yellow box.

The system works in order to obtain the widest list of results. It automatically matches your entered keyword into the corresponding MeSH term (see below, 1.3) to cover all the articles indexed with that subject, and therefore return a very specific list of results, but at the same time it runs a query with your keywords in each field (title, abstract, author...), in order to retrieve also the most recent contribution not yet indexed.

1.3 MeSH (Medical Subject Headings): the bearing structure of MEDLINE

MeSH is the U.S. National Library of Medicine’s controlled vocabulary of biomedical terms that is used to describe the subject of each journal article in MEDLINE. “Controlled vocabulary” means that in MeSH a “term” is an univocal entry which includes all the synonyms and variants, the most useful example being “Neoplasm” which stands for “cancer”, “cancers”, “tumor”, “tumour” etc. It simplifies the search because without MeSH you should have to run a query with each term. One or more MeSH terms are attached to each article, in order to better organize them by topic and to ensure an effective retrieval.

MeSH is a database itself (<http://www.ncbi.nlm.nih.gov/mesh>), with its own interface and logic, connected to MEDLINE to give it a structure. It contains over 25.000 terms, and it is constantly updated. If you need to search a MeSH term, you have to access the MeSH database from the homepage, by the corresponding link in the “More Resources” section, or by the pull-down menu above the Search Box, which indicates the active database.

MeSH has a hierarchical tree structure: each term like a branch has narrower branches below and broader above. It is important to notice that PubMed search engine automatically “explodes” the term, i.e. it catches the term - and the articles this term was assigned to - you enter and all its narrower terms, like in the example for “Spinal diseases” (fig. 3). The figure also shows the “Entry terms”, meaning the unified variants or synonyms:

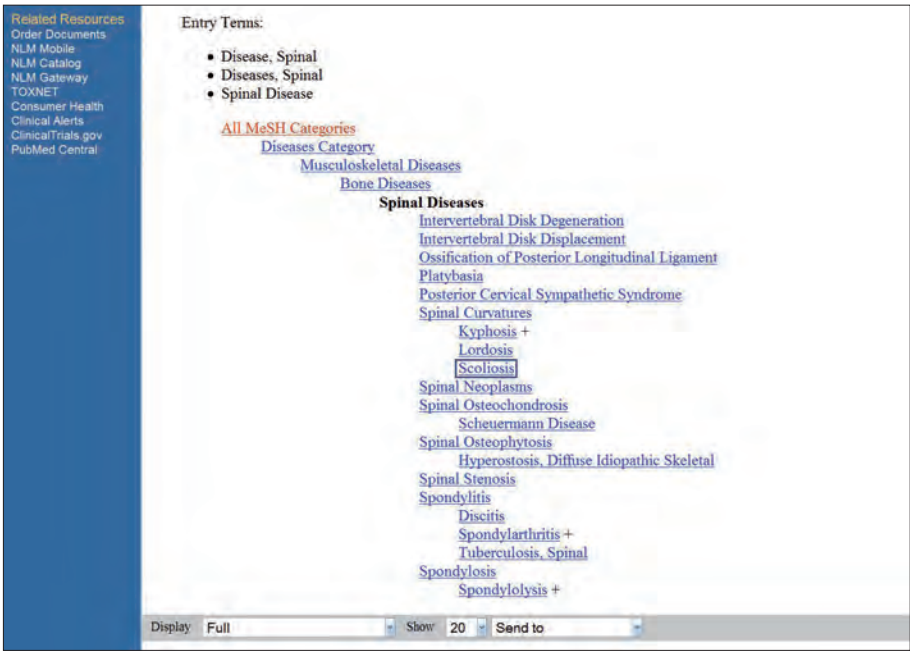


FIGURE 3. MeSH tree structure.

The system helps you showing the whole tree, so that you can navigate among the terms and find out the one that is more fitting to your search. When you type your keyword in the MeSH database, the system maps it to the existent MeSH term, and then shows a list of MeSH terms. They can contain your keyword or they can stand for the concept. Each term has a brief definition and the date of its addition to the MeSH vocabulary: it's obvious that if you decide to use a MeSH term introduced in 2003, you won't retrieve all the previous records, you need to use the related terms. Once you have selected your term, the system shows you

also its specific **Subheadings** (fig. 4). Using the Subheadings by flagging them is a first, effective way to limit your search: e.g. if you are interested in “Spinal diseases” but only in the side of “surgery” AND “rehabilitation”, you flag those Subheadings and you sensibly reduce your results.

In the same page you find other two check boxes:

- **Restrict Search to Major Topic headings only:** it means that the system retrieves only the items in which your term is the first for relevance
- **Do not Explode this term** means that the system will not include terms found below in the MeSH tree. It could be dangerous in some cases, because some MeSH term could include synonyms that otherwise you’ll never find. Anyway: not exploding our term you will restrict again your search.

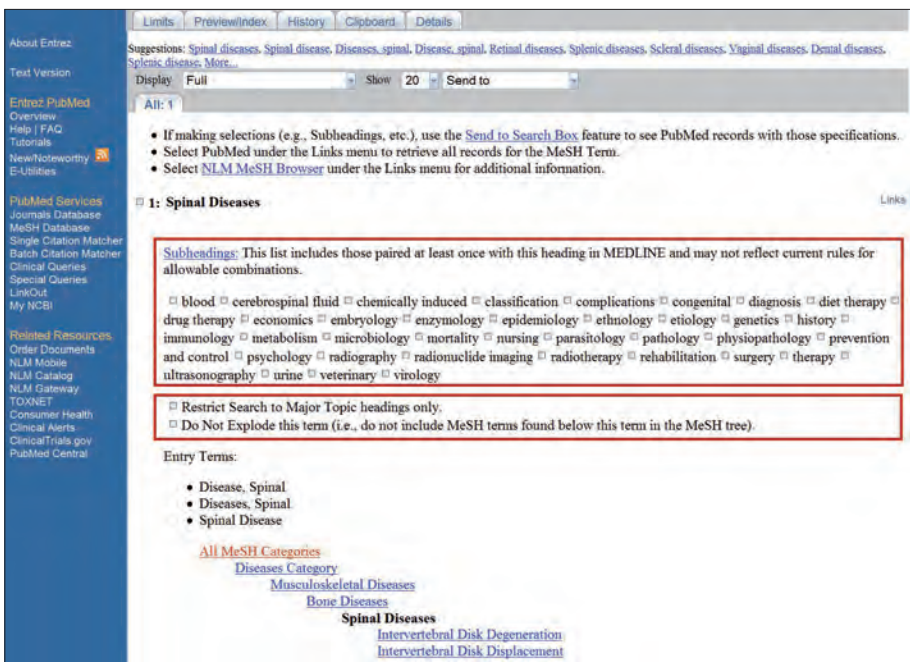


FIGURE 4. MeSH database: SubHeadings and Major Topic/Don't explode options.

A technical information about MeSH terms: the librarians working at the Nation Library of Medicine always choose, in indexing an article, the most specific term (an article dealing with “Muscle stretching exercise” will be indexed with this MeSH term, not with the broader “Exercise therapy”). Take it into account in deciding to explode or not your term.

To run your search after having set all your preferences in this page, you need to go at the very end of the page, open the “Send to” window, and select “Search Box with [the appropriate Boolean operator]”. Then click on the “Search PubMed” button just below the new window that the system shows to you, containing the translation of your query in the engine language.

There are differences between a search run only by MeSH terms and a search run by a free keyword (technically, [Text Word]). As said, the system assists you: if you simply type a keyword in the Search Box, it runs a search both for MeSH terms and for Text Word. There are advantages in both choices: it depends on the topic or the aim of your search, there is no recipe. Running a search with a MeSH term ensures a more specific list of results, because finds articles which have been uniformly indexed by subject, whatever the author's words, or whatever synonyms exist.

The search by Text Word gives you back a broader set of items, not so specific, because the system looks for your typed keyword even if it appears few times. If you want to search only by keyword and exclude the MeSH term search you have to type after your keyword [TIAB] if you want to look in the Title or Abstract or [TW] if you want a search by Text Word in more fields. Remember to run as many searches as the possible synonyms or variants of the keyword.

Just a complementary note about the Boolean operators AND, OR, NOT: be careful that AND narrows down your search (it looks for a term AND another all together, in the same record), whereas OR expands it by finding all the records which contain the first term, all the records which contain the second one, and all the records that contain the two together. Remember also that the system reads your query from left to right. For a complex strategy, you need to use brackets to «nest» groups of terms that must be processed first: if you are searching the implications of both ligament injuries and reconstruction with osteoarthritis, the right strategy is “osteoarthritis AND (ligament injuries OR reconstruction)”.

1.4 Tools to run an effective search

To run an effective search, one of the most useful tools is the **Limits** page: they are parameters to refine your search. “Limits” displays as a link just above the Search Box on all PubMed pages.

You flag one or more of these parameters to refine your search according to:

- Date of publication
- Type of article (Clinical trial, Practice guideline, Randomized controlled trial...)
- Text Options (Full text, Free full text, Abstract)
- Languages
- Subsets in PubMed (Core clinical, Dental, Nursing journals respectively) and Topics (Cancer, AIDS, Systematic reviews...)
- Ages (with ranges)
- Gender

Please remind that subsequent searches will include “Limits” until they are removed. You can clear them within the “Limits” page clicking the “Reset” button, or the “Remove” link from the “Limits activated” message bar. The yellow “Limits activated” message bar displays above the results

summary line on search results page; when active, it displays also both under the Search Box on the homepage and the Advanced search page.

PubMed **Advanced search** has been streamlined.

The new "Search Builder" is a very powerful tool which allows users to build their own search, even complex, by using the cited Boolean logic operators AND (conjunction), OR (disjunction), NOT (complement or negation) and combine more than 30 fields to run your search into (author, affiliation, journal, publication date...).

The "Search History" tracks the history of your previous searches (to a maximum of 100, expires after 8 hours or if you close the work session).

From the Advanced Search page you can directly access also other resources among which "Clinical queries": they are specialized PubMed searches for clinicians, with built-in search filters, to find out:

- **Search by Clinical Study Category:** enter your keyword in the Search Box; you are then requested to select a category between Therapy, Diagnosis, Etiology, Prognosis, Clinical Prediction Guides and a scope between Narrow or Broad;
- **Find Systematic reviews:** entering a keyword in this Search Box you will be provided by a specific list of results covering a broad set of articles - systematic reviews, meta-analyses, reviews of clinical trials, evidence-based medicine, consensus development conferences, guidelines, citations from journals specialized in clinical review studies;
- **Medical genetic searches:** enter a term in the Search Box, then select a category between Diagnosis, Differential Diagnosis, Clinical description, Management, Genetic Counselling, Molecular genetics, Genetic testing.

1.5 How to save items of interest

Once you have obtained a list of results, you can select your items of interest by flagging them and then operate with the **Send To** pull-down menu and its five options:

- **Order**, a Document Delivery service (by charge);
- **File**, for reference management tools;
- **E-mail**, to send the selected items;
- **Clipboard**, to **temporarily** save items in the "Clipboard" tool for a maximum of 500 items. When the "Clipboard" is active - it expires after 8 hours of inactivity - the system generates:
 - a link in homepage, with the numbers of added items;
 - a green message bar on the results page;
 - a green link below each single item once added;
- **Collection**, to **definitely** save items in a "Collection" in MyNCBI. If you are not logged in yet, the system prompts a pop up window to let you sign in (or freely register in MyNCBI the first time). Then you can choose either to create a new collection or add items to an existing one. These items will be available once and for all. Once saved, you can then edit your Collections (delete, merge, add...).

1.6 How to save a search strategy in MyNCBI and automatically receive updates

MyNCBI is a feature tailored to customize PubMed services for each user. It is a personal space you can access anytime from anywhere once logged in. It displays on the upper right side of each PubMed page. The first time you have to register, choosing a username and a password. It is completely free and takes only few minutes. Then, when you connect again to PubMed, you only need to “Sign in” to MyNCBI: you will see all your Saved Searches, Collections, and Bibliography. MyNCBI also allows you to change some preferences to customize your PubMed layout on your screen (format, filtering options...).

To save your whole Search strategy, once run the search simply click above the Search Box on “Save search” from the results list. A pop-up appears (your browser needs to allow pop-ups at least for this time), in which you can give your Search a name to store it. Then you can select if you want automatic e-mail updates and how often (every day, week, month): the system automatically re-runs the search and sends the new results to your e-mail address. Now the search strategy is saved in MyNCBI. If you didn’t choose the automatic e-mail update, signing in to MyNCBI you can select the desired search and click the button “What’s New for Selected”. The search will be re-run. You can always edit your Searches (delete, change update preferences...).

Another innovative way to make the system run a determinate search strategy is to create a RSS feed. RSS (Really Simple Syndication) is an XML-based format used to send new items or information to recipients who use RSS feed readers. Once run a search, click on the “RSS” link above the Search Box. Following the instructions, you have only to cut and paste the URL that the system gives to you in your RSS reader. You will receive updated results in your RSS reader.

1.7 Googling PubMed: new interfaces to ask a myth

No doubt that Google has deeply changed the users’ expectations in terms of easiness, speedy and efficiency of the query. Assuming that, and harnessing the potentiality of the Web 2.0 and the Semantic Web, some new projects offer a Google-way to search PubMed.

GoPubMed (<http://www.gopubmed.org/>) is far the most interesting experiment in providing a new, inviting interface to navigate PubMed and its lists of results. GoPubMed is a knowledge-based search engine for bio-medical texts.

GoPubMed retrieves PubMed abstracts for your search query in the same way that PubMed does, but then sorts relevant information to the 4 top level categories, i.e. “What”, “Who”, “When”, “Where”.

Navigating the tree of the extracted concepts you can get an answer to your question, e.g. “Which are treatment outcomes for the spine traumatic injuries?”. PubMed retrieves more than 27.000 items for “spine trau-

matic injuries”, GoPubMed then allows you to go further and intuitively refine your search just by clicking on the corresponding label in the navigation tree on the left. Please notice that the system shows directly in the first places pertinent items which in PubMed run lower in the results list.

Hubmed (<http://www.hubmed.org>) is an alternative way to ask PubMed. It has a simple Google-like search box, but offers the possibility of clustering the results. It can also associate to each article a tag, different from the MeSH, assigned by the users, in the logic of the Web 2.0.

Finally, **eTBLAST** (<http://etest.vbi.vt.edu/etblast3/>) a unique search engine that doesn't work with keyword but with phrases: it lets you input a whole paragraph and returns abstracts in PubMed that are similar to it. It is useful to run searches in plain text. A curiosity: this search engine has been used to rate the similarity between scientific articles, with the surprising outcome that in too many cases they look like too equal, or better... plagiarized?

2. BIOMEDICAL SEARCH ENGINES AND META-SEARCHERS

Please, raise one's hand who doesn't begin any search on the Web “googling” a keyword. “To google” stood out as a neologism in English, due to the popularity the search engine reached. “Googling” has deeply influenced the users' expectations about the answers to their questions and the way of searching the Web. Needless to say that we can't ignore these premises, so we shall start from Google itself and its Advanced Options. But Google is not alone, and there are other search engines which could be useful for scholarly researches, both because they are dedicated to explore only biomedical resources, reducing the noise, or because they offer features - like clustering the results - that make easier the life. Don't forget that there is no recipe at all: each time, it depends on the question, the aim and the perspective of the query.

2.1 Google behind the curtain

Google provides an Advanced Search mask that allows you to refine your query. First of all you can associate words with the Boolean operator, i.e. you can tell Google to find ALL the words you type (meaning AND), e.g. stroke AND rehabilitation; AT LEAST ONE of the words (meaning OR), e.g. stroke OR rehabilitation; WITHOUT a word (meaning NOT). If you want you can also search for the EXACT PHRASE you have to include your phrase in brackets, e.g. “stroke rehabilitation”.

You can also limit your search by format (.pdf for academic articles, .ppt for meeting presentations), domain (.edu will consider only academic URLs), occurrences, date, usage rights.

If you need a definition, Google can provide for it, if you type the operator “define:” with no spaces between it and the term, it will show you a list of definitions gathered from various online sources: e.g. “define:scoliosis” get a list of more than 20 different entries from medical glossaries.

In the Web 2.0 momentum, Google is also useful to find out blogs or newsgroups - which are not to be underestimated - about subject of interest. The right URL to start from are, respectively, <http://blogsearch.google.com/> and <http://groups.google.com/>.

2.2 Google Scholar: standing on the shoulders of giants, but...

Google Scholar (<http://scholar.google.com/>) is a service provided by Google aimed at finding peer-reviewed papers, theses, books, abstracts and articles from academic publishers, professional societies, universities. The interface is the usual easy Search box of Google, the ranking of the result is quite the same of the general search engine: weighing the full text of each article, the author, the publication in which the article appears, and how often the piece has been cited in other scholarly literature. But... Does it works applied to scholarly literature? Perhaps if I'm looking for the most cited article, it does, but if I'm looking for the most recent one, the criterion of the most cited one doesn't work at all, because it would retrieve only old stuff. The Advanced Search page allows to set limits by date, asking for articles published between a range of years; from the results page you find a pull-down menu with the option "Any date" or single year.

This emphasizes another evident lack of Google Scholar: it never indicates its coverage, nor the period of embargo that the publishers establish, i.e. you never know what you are searching, where, since when and till when. So, you stand on the shoulders of the giants, as the logo states, but you don't know neither who the giants are nor their age or weight...

What Google Scholar provides in an excellent way is the **Citation Tracking** feature: for each item it shows the link "Cited by" that allows tracking back the history and the relevance of a work according to the received citation.

A last positive annotation about the coverage of Google Scholar: being a free-of charge service, it harvests the Open Access repositories, so it lists with a good relevance ranking scholarly material self-archived in full-text by the authors (i.e. the pre-print of the final versions submitted to the traditional journals).

2.3 Scientific Commons: an Open Access search engine

Scientific Commons (<http://www.scientificcommons.org/>) is a project of the University of St. Gallen (CH). It is a search engine dedicated to Open Access resources, and indexes about 35 million records from about 1200 repositories all over the world.

According to the principles of the Open Access movement¹, all of the contributions have to be freely available in full text, via two channels:

¹ Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, 2003, available at <http://oa.mpg.de/openaccess-berlin/berlindeclaration.html>; see also Open Access Directory, available at http://oad.simmons.edu/oadwiki/Main_Page.

- “Green road”: the author self-archives in Open Archives (<http://www.opendoar.org/>) the preprint or post print of the paper, following the copyright permission of the different publishers (listed in Project ROMEO - <http://www.sherpa.ac.uk/romeo/>). This is the option chosen by more than 220 institutions and funders all over the world, according to the principle that the results of a publicly funded research must be publicly available (<http://www.eprints.org/openaccess/policysignup/>). The US National Institutes of Health adopted such a mandatory policy in 2008. PubMedCentral (<http://www.ncbi.nlm.nih.gov/pmc/>) is the repository which contains these Open Access contributions: they appears in PubMed as “free full text” articles.
- “Gold road”: the author publishes in an Open Access journal (<http://www.doaj.org/>). Open Access journal are peer-reviewed but have no subscription fees. Some of them (23%) request an Article Processing Charge to cover publication costs (<http://www.biomedcentral.com/info/authors/apccomparison/>).

Open Access is an option to be considered by researchers, because of its advantages in term of visibility, dissemination and citations².

2.4 Biomedical dedicated search engines

TRiP, Turning Research into Practice (<http://www.tripdatabase.com/index.html>) is a search engine specialized in Evidence-based resources. It clusters the results in categories like Systematic reviews or Guidelines. It also offers images and a box with background information on the requested topic, and a link to BMJ Doc2Doc service (<http://doc2doc.bmj.com/>) which connects physicians into debates and discussions.

MedHunt (<http://www.hon.ch/MedHunt/>), powered by the Health On the Net Foundation, is a search engine dedicated to biomedical resources. Many of these resources have a HON code, a seal of quality (see below, 4.2).

Scirus (<http://www.scirus.com/srsapp/>), searching more than 370 million science-related pages, focuses only on Web pages containing scientific content and goes deeper than the common search engines, looking for reports, peer-reviewed articles, patents, pre prints and journals. It allows also selecting a subject area, to narrow by author, journal, date, and to customize and save searches.

The Diseases database (<http://www.diseasesdatabase.com/begin.asp>) is a cross-referenced index of human disease, medications, symptoms, signs. It gives definitions, links to special Web sites, or sends your search to other resources (PubMed, Wikipedia, Scirus...).

² Swan, A. (2010) The Open Access citation advantage: Studies and results to date. Technical Report, available at <http://eprints.ecs.soton.ac.uk/18516/>.

2.5 Metasearch engines: simultaneous queries for time-effective searches

NLM Gateway (<http://gateway.nlm.nih.gov/gw/Cmd>) is a Web-based system that lets users search simultaneously among the National Library of Medicine free databases. NLM Gateway integrates the search in PubMed with some of the toxicological databases and some of the genetic databases illustrated below, and, above all, it gives the unique access to the Meeting Abstract database, the National Library of Medicine's online collection of abstracts coming from medical Meetings. As usual within a meta-search application, the user enters one query that is sent automatically to multiple retrieval systems having different characteristics but potentially useful results. Results from the target systems are presented in broad categories (bibliographic resources, consumer health resources, other information resources) rather than by database (fig. 5).

The screenshot shows the NLM Gateway interface. At the top, it says "NLM Gateway" and "Your Entrance to Resources from the National Library of Medicine". Below this is a navigation bar with links: Home, Term Finder, Limits/Settings, Search Details, History, My Locker, About, Help, and FAQ. A search bar contains the term "scoliosis" and buttons for "Search" and "Clear". Below the search bar, it says "Results Summary: 16475 records found" and "[Bookmark this Search]". The results are organized into three columns: Bibliographic Resources, Consumer Health Resources, and Other Information Resources. Each column lists various databases and the number of records found for each.

Bibliographic Resources	Consumer Health Resources	Other Information Resources
14615 MEDLINE/PubMed - journal citations, abstracts	6 MedlinePlus - Health Topics	1 Images from the History of Medicine
315 NLM Catalog - books, AVs, serials	1 MedlinePlus - Drug Information	0 HSRProj - Health Services Research Projects
665 Bookshelf - full text biomedical books	43 MedlinePlus - Medical Encyclopedia	476 OMIM - Online Mendelian Inheritance in Man
155 TOXLINE Subset - toxicology citations	0 MedlinePlus - Current Health News	20 HSDB - Hazardous Substances Data Bank
46 DART - Developmental and Reproductive Toxicology	5 MedlinePlus - Other Resources	2 IRIS - Integrated Risk Information System
10 Meeting Abstracts	56 ClinicalTrials.gov	0 ITER - International Toxicity Estimates for Risk
	8 DIRLINE - Directory of Health Organizations	0 GENE-TOX - Genetic Toxicology (Mutagenicity)
	51 Genetics Home Reference	0 CCRIS - Chemical Carcinogenesis Research Information System
	0 Household Products Database	0 Profiles in Science

FIGURE 5. NLM Gateway metasearch: results list.

Entrez (<http://www.ncbi.nlm.nih.gov/sites/gquery>) is the integrated, text-based search and retrieval system used at National Centre for Biotechnology Information (NCBI) for its major free databases, including PubMed, Nucleotide and Protein Sequences, Protein Structures, Complete Genomes, Taxonomy, and PubChem databases.

By using the Entrez "Global query", a search across all Entrez databases is performed by entering a simple search term or phrase in the "Search across databases" query box. Select the Go button to execute the search. The results found in each database are displayed on the Global Query page, as shown in fig. 6.

If you are looking for information about a chemical substance, e.g. Naproxen, you can access **TOXNET** - TOXicology Data NETwork (<http://toxnet.nlm.nih.gov/>), in which you can run a cross-search among databases covering toxicology, hazardous chemicals, environmental health and related areas. HSDB (<http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?HSDB>) is the core of TOXNET and it focused on the toxicology

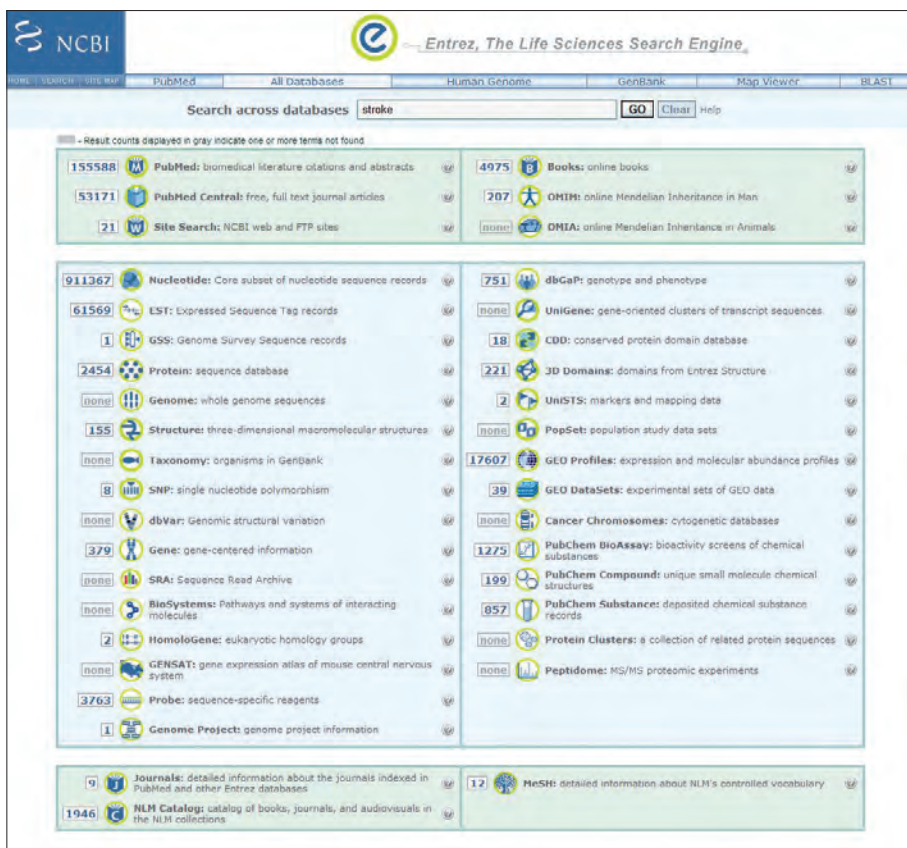


FIGURE 6. Entrez cross-search engine.

of potentially hazardous chemicals. It is organized into over 5000 individual chemical records, and is enhanced with information on human exposure, industrial hygiene, emergency handling procedures, environmental fate, regulatory requirements, and related areas. The data are fully referenced and peer-reviewed by a scientific review panel.

3. FREE BIOMEDICAL DATABASES ON THE WEB

As we have seen, NCBI and the US National Library of Medicine give free access to all their databases. Other useful databases are available.

3.1 BioMedCentral Databases: Open Access resources

A useful list of other free biomedical databases is provided by the Open Access publisher BioMedCentral: <http://databases.biomedcentral.com/search>. It is searchable by subject area or by content.

3.2 Genetic implications for diseases, conditions, drugs

If your research is dealing with any genetic implication, you have several possibilities to perform a query, going from quick information to ultra-specialist genetic databases.

The referring database is **OMIM** - Online Mendelian Inheritance in Man, (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=OMIM>). It is a catalogue of human genes and genetic disorders, with links to literature references, sequence records, maps, and related databases.

Gene (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene>) is the Entrez database for gene-centred information. If you run for instance a search for “Human muscular dystrophy”, the system retrieves all the genes involved, and for each of them the nomenclature, the map location, gene products and their attributes, markers, phenotypes, and links to citations, sequences, variation details, maps...

A useful tool for testing laboratories, more European-oriented, is **Orphanet** (<http://www.orpha.net/consor/cgi-bin/home.php?lng=EN>), which provides the testing laboratories for a rare disease or syndrome, beyond other information and news³.

4. BIOMEDICAL WEBSITES, SUBJECT GATEWAYS, BLOGS...

No doubt that the Web, with its about 206 millions sites (according to Netcraft as of May, 2010, http://news.netcraft.com/archives/web_server_survey.html), in the last ten years has represented a real Copernican revolution in the information search and retrieve: the user is in the centre now, he/she can access information - no matter where it is based in the world - just typing few words on the keyboard, and can build his/her own virtual library each time collecting documents answering to a specific question. After the first justified enthusiasm, now we have to cope with the so called “information overload”: a matter of quantity, indeed, but mainly of quality. Reliability, updating, authoritativeness ought to be a requirement, all the more reason for the biomedical field. Here are presented Subject gateways and projects aimed at supplying trustworthy health information in Web sites - we are not talking about medical literature or article in journals, indexed in MEDLINE, but about Internet resources in general.

4.1 Medline Plus

Medline Plus (<http://medlineplus.gov/>) is the National Library of Medicine’s Web site tailored for consumer health information. It is patient-

³ For in-depth genetic searches, see Giglia, E Beyond PubMed. Other free-access biomedical databases, *European Journal of Physical and Rehabilitation Medicine* (Europa Medicophysica) 2007 December;43(4):563-9, available at <http://www.minervamedica.it/en/journals/europa-medicophysica/article.php?cod=R33Y2007N04A0563>

oriented, but it provides quick references, definitions, practical overview and useful links also for the physician. It lays out many sections:

- **Health Topics:** covering 800 topics on conditions, diseases and wellness.
- **Drugs and Supplements:** searchable by generic or brand name, it gives information on side effects and interactions.
- **Medical Encyclopaedia:** includes over 4000 articles about diseases, tests, symptoms, injuries, surgeries and an extensive library of medical illustrations.
- **Dictionary:** spelling and definitions of medical words.
- **News.**
- **Interactive tutorials:** 165 tutorials on diseases and conditions (e.g. cerebral palsy), test and diagnostic procedures (e.g. knee arthroscopy), surgery and treatment procedures (e.g. hip replacement - physical therapy), and prevention and wellness (e.g. back exercises).

Let's explore the **Health Topics** section, which is so rich of useful data. You can browse by Body Location/System, Disorders and conditions, Diagnosis and therapy, Demographic groups, Health and wellness. For each topic, the system opens a real "hub", like the one you find in fig. 7, with an overview, the latest news, specific conditions and related issues, financial issues, information on clinical trials, selected journal articles... For the diseases, on the right end you find also a link to the US National Institute of compe-

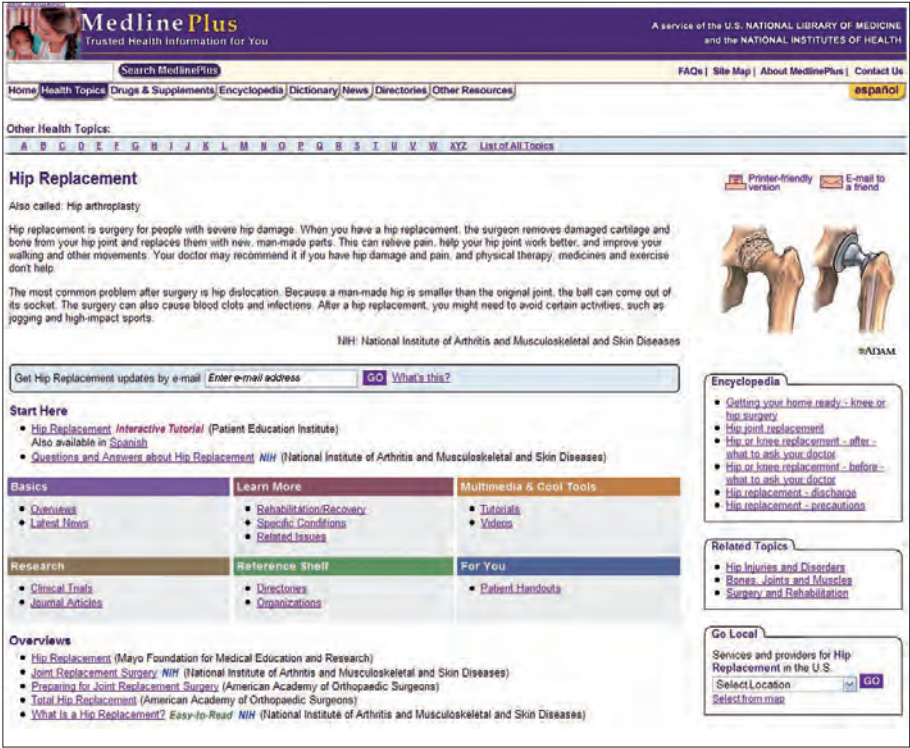


FIGURE 7. Medline Plus resources for Hip Replacement.

tence, for instance the National Institute of Arthritis and Musculoskeletal and Skin Diseases (<http://www.niams.nih.gov/>) or the National Institute of Neurological Disorders and Stroke (<http://www.ninds.nih.gov/>), whose Web site are other precious mines of information.

4.2 HON Code: medical information you can trust

Health On the Net (HON) Foundation is a Swiss non-profit, non-governmental organization created in 1995 and accredited to the Economic and Social Council of the United Nations (<http://www.healthonnet.org/>). HON is the leading organization promoting and guiding the deployment of useful and reliable online medical and health information. To address this issue HON has elaborated a *Code of Conduct* (HON Code) to help standardise the reliability of medical and health information available on the Web. HON code states 8 principles (<http://www.hon.ch/HONcode/Pro/Conduct.html>):

1. Authoritative (indicate the qualifications of the authors)
2. Complementarity (information should support, not replace, the doctor-patient relationship)
3. Privacy (respect the privacy and confidentiality of personal data)
4. Attribution (cite the sources)
5. Justifiability (site must back up claims relating to benefits and performance)
6. Transparency (accessible presentation, accurate email contact)
7. Financial disclosure (identify funding sources)
8. Advertising policy (clearly distinguish advertising from editorial content)

Medical Web sites that apply for HON code are submitted to an accreditation process in order to evaluate the compliance with the eight principles. Then they receive the HON logo that they have to put on their Web page. On the HON's homepage (<http://www.hon.ch/home1.html>) you are provided with HON code, HONsearch, HONtools and HONtopics, all useful. Above the Search box you have also a string with three profile tabs: if you make the system recognize you as a "Medical professional" you will be shown other tailored tools more suitable with your role.

4.3 INTUTE: access to the very best Internet resources

INTUTE is a long-term collaborative project between 7 Universities in the UK, providing access to quality resources through a process of evaluation (<http://www.intute.ac.uk>). Materials you can find are journal articles, patient leaflets, practice guidelines, associations websites and so on. INTUTE also provides a "Virtual Training Suite" section, a set of free Internet tutorials to help you develop Internet research skills (<http://www.vts.intute.ac.uk/>).

The “Medicine” section (<http://www.intute.ac.uk/medicine/>) is both searchable by keyword and browsable by MeSH term. It also provides a list of many active blogs in Medicine. The same appears in the “Nursery, midwifery and Allied Health” section (<http://www.intute.ac.uk/nmah/>), which includes Rehabilitation Therapy. Signing in into the MyINTUTE service, you can easily save your searches, receive e-mail alerts on subject of interest, and export records.

4.4 Orthogate: the gateway to the orthopaedic Internet

Orthogate is a project developed between Canada and the USA aimed at improving orthopaedic care and musculoskeletal health, education and research using Internet technologies (<http://www.orthogate.org/>). It is compliant with HONCode. Orthogate provides for a sort of virtual community of specialists with a forum, a blog, mailing lists, and discussions online. Among the “Resources” listed on the navigation toolbar on the left, there is a very interesting «**Orthopaedist’s guide to the Internet**», designed to provide researchers with self-teaching tools to improve the use of the Internet in order to obtain the best and most useful orthopaedic information (<http://www.orthogate.org/guide/chapter1.html#guide>). Another innovative tool is Orthopaedia (<http://www.orthopaedia.com/display/Main/Home>), a collaborative orthopaedic knowledge base written by users (like Wikipedia).

The most useful instrument in Orthogate is a virtual reference desk, linked from the homepage as “Web links”, (direct URL: <http://owl.orthogate.org/>), a directory of more than 10.000 Web sites clustered by topics.

4.5 CISMef: Internet speaks French

CISMef - Catalogue et Index des Sites Médicaux Francophones (<http://www.cismef.org/>) is a project of the University of Rouen cataloguing Internet resources in French. You can search by keyword, or browse by MeSH term, or search by resource type.

4.6 WEMOVE: Worldwide education & awareness on movement disorders

WEMOVE (<http://www.wemove.org/>) is a comprehensive resource for movement disorder information and a hub of movement disorder activities on the Web. The website presents articles and multimedia material focused on major movement disorders.

4.7 The Internet Stroke Center

The **Internet Stroke Center** (<http://www.strokecenter.org/>) is an educational service provided by Stroke Center at Barnes-Jewish Hospital and Washington University School of Medicine. Apart from a plethora of

scientific resources about stroke, a registry of ongoing clinical trials is hosted on this website. Resources include clinical information for patients, students, health care professionals, a neurology image library and a collection of stroke scales and clinical assessment tools.

4.8 NIDRR: The National Institute on Disability and Rehabilitation Research

NIDRR conducts comprehensive and coordinated programs of research and related activities to maximize the full inclusion, social integration, employment and independent living of individuals of all ages with disabilities. NIDRR's focus includes research in areas such as employment, health and function, technology for access and function, independent living and community integration and other associated disability research areas. NIDRR operates as a component of the Office of Special Education and Rehabilitative Services (OSERS) at the U.S. Department of Education and provides the following Web services:

4.8.1 NARIC: National Rehabilitation Information Center

NARIC (<http://www.naric.com>) offers access to databases, resources and research tools dedicated to rehabilitation. A well known database is REHABDATA, which contains a big number of book abstracts, reports, articles, and audiovisual materials relating to disability and rehabilitation research.

4.8.2 AbleData

AbleData (<http://www.abledata.com/abledata.cfm>) is an online database of assistive technology products and rehabilitation equipment. In the literature section of this website, searchable listing of articles, books, paper and electronic publications about assistive technology are available.

4.9 BMJ: Research Methods & Reporting

The **British Medical Journal** (BMJ) has published a series of articles on research methodology (http://www.bmj.com/cgi/search?tocsectionid=Research%20Methods*). The aim is to present information and guidelines on issues such as data collection, scientific trials, analysis methods and reporting of scientific data. Free access is available for some articles.

4.10 EQUATOR: enhancing the quality and transparency of health research

The **EQUATOR Network** is an international initiative that seeks to improve reliability and value of medical research literature by promoting transparent and accurate reporting of research studies. The EQUATOR Network Resource Centre (<http://www.equator-network.org>) provides up-to-date resources related to health research reporting mainly for au-

thors of research articles, journal editors, peer reviewers and reporting guideline developers. EQUATOR is developing a digital library that will provide publications related to writing research papers, use of reporting guidelines in editorial and peer review processes, development of reporting guidelines, educational materials and tools for editors, peer reviewers and researchers.

4.11 ProQolid: Patient-Reported Outcome & Quality of Life Instruments Database

The **Quality of Life Instruments Database** project was initiated by Mapi Research Institute to provide all those involved in health care evaluation with a comprehensive source of information on patient reported outcome and quality of life measures available through the Internet. The website addresses practical issues, such as the availability of different translations and copyrights. Musculoskeletal section of ProQolid (http://www.proqolid.org/proqolid/search__1/pathology_disease?pty=1925) is dedicated to instruments related to rehabilitation.

4.12 Orthopaedic Scores

Orthopaedic Scores is a list of assessment scales relevant to orthopaedic disorders (<http://www.orthopaedicscores.com/>). Questionnaires are sorted according to region of the body and grouped in those completed by patient or clinician. They are automatically scored and can be saved or printed.

4.13 IN-CAM: Outcomes Database

IN-CAM (<http://www.outcomesdatabase.org/>) presents a database of outcome measures of particular importance to Complementary and Alternative Medicine research.

4.14 WHO & Disability

World Health Organization's (WHO) portal on disability (<http://www.who.int/disabilities/cbr/en/>) includes articles about its policy, projects, resolutions, activities and other resources related to rehabilitation and disability. An important project for Physical and Rehabilitation Medicine is the International Classification of Functioning, Disability and Health (ICF). The online version of ICF and other useful tools can be accessed at <http://www.who.int/classifications/icf/en/>.

4.15 UEMS PRM Section & Board website

Physical and Rehabilitation Medicine Section & Board of European Union of Medical Specialists (www.euro-prm.org/) has created an interactive

distance learning platform offering selected free scientific resources focused on education and research, e-books, self assessment-activities and a forum⁴.

5. EVIDENCE BASED MEDICINE RESOURCES: PEDro AND OTHER TOOLS

We have already seen (2.4) TRIP - Turning Research into Practice, a biomedical search engine for EBM resources. Now we are going to explore PEDro and other useful gateways to practise EBM; yet, there are a lot of EBM databases, most of them being bibliographical and, unfortunately, for charge, such as the Cochrane Library.

5.1 PEDro, the pre-eminently Physiotherapy Evidence Database

PEDro stands for Physiotherapy Evidence Database, and surely it is well known in the world of rehabilitation medicine. We can try to explore it, to find out how to make our searches more effective.

PEDro (<http://www.pedro.org.au/>) is a free database of evidence relevant to physiotherapy. It is maintained by the Centre for Evidence-Based Physiotherapy, located at the Musculoskeletal Division, The George Institute for International Health (University of Sydney).

The database is available in many languages (English, Portuguese, German, and Japanese). PEDro contains - as of May, 2010 - citations of over 16.000 randomised controlled trials, systematic reviews and evidence-based clinical practice guidelines relevant to physiotherapy. Where possible, abstracts and links to full-text versions of the documents are also provided. It contains no items of other scientific literature, i.e., not scientific journal articles. It is updated once per month, on the first Monday. It includes any Clinical trial, Systematic review or Practice guideline that satisfies the rigorous criteria of inclusion as stated on the website (<http://www.pedro.org.au/english/downloads/criteria/>), regardless of the time of publication: the oldest record, yet suitable, was published in 1929. To give just an idea of the solidity of such criteria: Systematic Reviews are not included unless they contain a "Method" section; Guidelines are not included unless they are produced under the auspices of a health profession association, or professional societies, or public/private organizations, or government agencies: they also must be publicly available and they must contain a "Recommendation" section useful for physiotherapy practice; Clinical trials are eligible only if they involve comparison of at least two interventions - which have to include treatments, prevention strategies, diagnostic tests... -, one of which has to be part of physiotherapy practice. Clinical trials are also rated with a checklist known as "PEDro

⁴ e-learning website - PRM Section and Board of UEMS. Available at: <http://www.euro-prm.org/elearning/>.

scale”, whose number appears in the results list. The PEDro scale deals only with internal validity of the trials and whether they contain sufficient statistical information to make them interpretable. It does not rate the external validity (i.e. the meaningfulness), or the size of the treatment effect. The 10 single scale items are reported for each item in the complete record visualization - obtained by clicking on the title in the results list, as shown in fig. 8:



FIGURE 8. “PEDro scale” score in detail for a clinical trial.

The PEDro score determines the ranking of the trials in the results list. Systematic reviews and Practice guidelines, which get a score of N/A that stands for Not Applicable, are presented sorted by year, starting from the more recent one. The system presents - please keep it in mind when facing a long list of results - first the Guidelines, followed by the Systematic reviews, and then by the Clinical trials with their own score.

5.2 How to search PEDro and get effective results

Starting from the PEDro homepage, a Simple search and an Advanced search are provided. In the “Simple search” query box you can

just type your query term, and the system retrieves all the documents containing your words, no matter if in the title or the abstract or anywhere else in the record. It could be a little bit dispersive, e.g. for “stroke” it gives back 1025 results, going from rehabilitation to prevention to generic cardiovascular risks.

The “Advanced search” interface has been appositely created to bypass this lack of effectiveness.

You can type your query term in the search box and then you can refine or limit your search by further steps like Therapy, Problem, Body part, Sub-discipline. Please note that you do not need to enter search term in each of the pull-down menus: in many cases it could be absolutely counterproductive and bring to a zero result list. Imagine that you are interested in finding a Practice guideline dealing with the post stroke rehabilitation for patients with problems in hand coordination. Let’s see step by step - but in the practice you can obviously make all these choices at once. By typing “stroke” we get 1025 results. Refining only by Therapy, choosing “behaviour modification”, we reduce to 44 records. Refining again by Problem, with “motor incoordination”, we get 17 items. Limiting by Body Part, “hand or wrist”, we have the 8 results shown in Fig. 9. Limiting by Method, “Practice guidelines”, we get only one record, precisely fitting with our information need.

Other useful limits are: author, title, source, date of publication, date of inclusion in the database, and even PEDro score. To select a range of years, you have to use three periods (...) between the years, e.g. 2000...2008. Use the same operator to specify records published during a single year: e.g. 2009...2009.

Using wildcards for word variations is another trick to run an effective search. If you are looking for a singular/plural term, just type the root and the system, by default, searches for any record containing these letters at the beginning of the word: e.g. typing “fractur” you will retrieve items containing “fracture” and “fractures”. The same for adjectives and related nouns: typing “spin” you will get all the “spine” and “spinal” records. Making the system consider a word not like a root is simple, just by typing = before your term: e.g. “=brain” will retrieve only “brain” itself and not “braincase, brainstem, brainwave” and so on.

Instead, if you are looking for a variation that includes a letter or a group of letter at the beginning of the word, you need to put an asterisk before your term: e.g. “*edema” will retrieve “edema”, “oedema”, “lymphedema” and “lymphoedema”. Please notice that in this case PEDro will not perform at the same time the final variations.

For a single letter truncation, PEDro requires the @ symbol: e.g. if you don’t remember the exact spelling of an author, you can type “R@uch” and have results for “Rauch” or “Rouch”.

To find an exact phrase, let’s enclose the words in inverted commas: “idiopathic scoliosis” will not retrieve “scoliosis” alone.

Once you got the result list, you can flag the most interesting items and then, clicking on the “Display selected records” link on the toolbar in the upper side of the results page, you can visualize the complete record

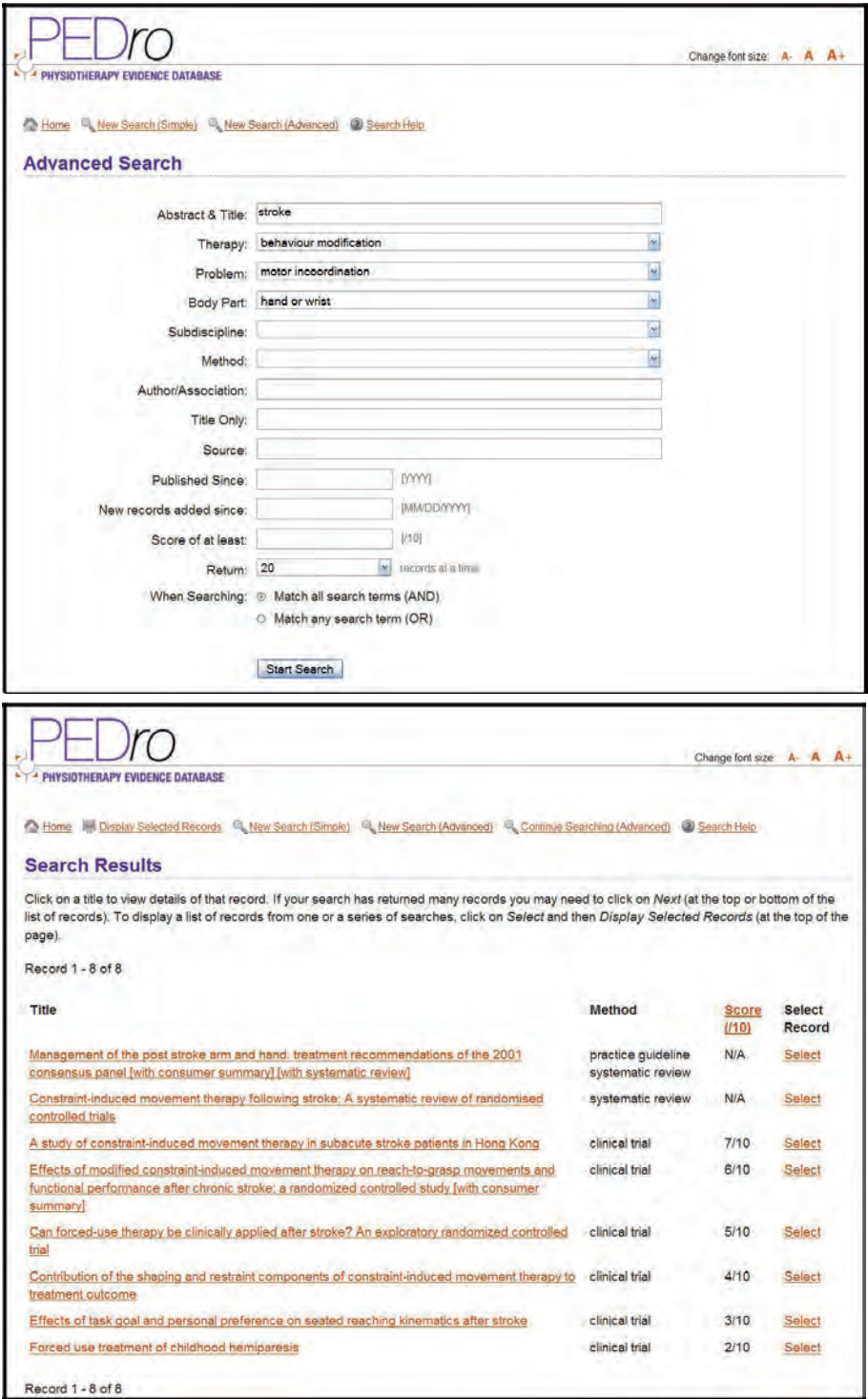


FIGURE 9. PEDro advanced search interface and refined results

with the abstract, the complete bibliographical references, the PEDro score in case of clinical trials, and the link to the full text when available. You can print the list of selected just by activating the “Print” option of your Web browser, or save it just by using the “Save as” option of your Web browser. If you want to e-mail the list, you can simply do it by clicking the “E-mail results” link on the navigation toolbar in the upper side of the “Display selected records” page.

If you want to export your saved items of interest in EndNote or RefWorks (they are reference manager tools, see below, 7.2), PEDro requires you to download an EndNote filter (available from the link in the homepage, section Download on the left side of the screen, direct URL <http://www.pedro.org.au/english/downloads/endnote-filter/>) or a RefWorks filter (<http://www.pedro.org.au/english/downloads/refworks-filter/>).

5.3 EBM gateways

Physiotherapy Choices (<http://www.physiotherapychoices.org.au/>) is a database designed for use by consumers of physiotherapy services, including patients, health service managers, and insurers. It is an initiative by Centre for Evidence-Based Physiotherapy (CEBP) at the University of Sydney. The database provides a catalogue of the best research evidence of the effectiveness of physiotherapy interventions - the same Clinical trials, Systematic reviews, and Practice guidelines, appearing in the PEDro database - associated with a summary in plain English.

There are several EBM Centres in the world, providing tools and services on the Web:

- **CEBM at Toronto University** (<http://ktclearinghouse.ca/cebm/>) provides tools in order to practise EBM and teach EBM. It offers a rich EBM Glossary (<http://ktclearinghouse.ca/cebm/glossary>) and all sort of EBM calculation tools (<http://ktclearinghouse.ca/cebm/toolbox>)
- **CEBM Net at Oxford University** (<http://www.cebm.net/>) offers CAT-Maker (<http://www.cebm.net/index.aspx?o=1216>), a software to create Critically Appraised Topics, and a list of EBM search tools (<http://www.cebm.net/index.aspx?o=1900>);
- **CRD (Center for Reviews and Dissemination)** at the University of York (<http://www.york.ac.uk/inst/crd/index.htm>) deals with evaluating the research evidence on health and public health questions. It gives access by a meta-search engine to:
 - **DARE** - Database of Abstracts of Reviews of Evidence, contains 15,000 abstracts of systematic reviews including over 6,000 quality assessed reviews and details of all Cochrane reviews and protocols;
 - **NHS EED** - National Health Service NHS Economic Evaluation Database, contains 24,000 abstracts of health economics papers including over 7,000 quality assessed economic evaluations.
 - **HTA** - Health Technology Assessment brings together details of over 8,000 completed and ongoing health technology assessments from around the world.

5.4 National Guideline Clearinghouse

National Guideline Clearinghouse publishes syntheses of selected guidelines that cover rehabilitation topic areas and expert commentary on issues of interest and importance to the clinical guideline community. The website of National Guideline Clearinghouse (<http://www.guidelines.gov/>) represents a resource for evidence-based clinical practice guidelines, published by the Agency for Healthcare Research and Quality (U.S. Department of Health and Human Services).

5.5 BestBETs: Best Evidence Topics

Physiotherapy section of **BestBETs** database (<http://www.bestbets.org/database/browse-by-topic.php?CategoryID=354>) presents evidence based answers on specific clinical questions. Each topic answers a carefully worded 3-part question, using a structured approach to finding and reviewing the literature. The BET method allows the use of lower quality research, and lists the shortcomings of the evidence used. Being brief and well-structured, BETs can be reviewed at regular intervals, to ensure the evidence remains the best available.

6. E-BOOKS, IMAGES, DICTIONARIES, ENCYCLOPAEDIAS...

When you need any sort of reference material (encyclopaedias, dictionaries, e-books, images) you can access a lot of free and qualified material on the Web.

Let's spend a word about the complex copyright question. Finding an image or a text on the Web doesn't mean automatically that it is free to use. Most of the images on the Web are under copyright, unless otherwise stated, e.g. by the Creative Commons licenses (<http://creativecommons.org/>), which allow the reuse saving the attribution. But it's the author that states them. Please be aware that «linking» to images or texts is always permitted; what is strictly forbidden is «copying and pasting» an image or a text into your own work without the explicit permission of the rightful owner.

6.1 Medical encyclopaedias

A **Medical Encyclopaedia** (<http://www.nlm.nih.gov/medlineplus/encyclopedia.html>) is hosted by Medline Plus, the US National Library of Medicine gateway to trusted health information we talked about above (4.1). It includes over 4,000 articles about diseases, tests, symptoms, injuries, and surgeries. It is available also in Spanish.

The **Medical Encyclopedia Index** (http://www.umm.edu/ency/index/eng_index.htm), edited by the US University of Maryland, is searchable by disease, symptom, injury, surgery, special topics and is available also in Spanish. For some term it offers also images free to get and reuse.

Please keep in mind that just for a definition you can also use the PubMed MeSH, Medical Subject Headings (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=mesh>), which provides 25,588 descriptors in 2010 MeSH. And don't forget **Wikipedia** (http://en.wikipedia.org/wiki/Main_Page) whose medical voices are very accurate and always up-to-date and, in the spirit of social networking, verified by thousand of readers in the world.

The International Encyclopedia of Rehabilitation (<http://cirrie.buffalo.edu/encyclopedia/index.php?language=en>) is a collaborative effort from the Center for International Rehabilitation Research Information and Exchange (CIRRIE), at the University at Buffalo, SUNY, and the Laboratoire d'informatique et de terminologie de la réadaptation et de l'intégration sociale (LITRIS), from the Institut de réadaptation en déficience physique de Québec (IRDPO). It is meant to provide researchers, professionals, students, persons with disabilities, and the general public with an easily accessible resource via the Internet. Its articles include the principal disabilities, as well as rehabilitation evaluation, interventions and associated topics related to physical, psychological, social and environmental aspects of rehabilitation.

On its completion, the encyclopaedia will include hundreds of articles on rehabilitation and disability topics identified through terms found in the CIRRIE and REHABDATA thesauri, the World Health Organization's International Classification of Functioning, Disability and Health (ICF) and the International Index and Dictionary of Rehabilitation and Social Integration (IIDRIS). The encyclopaedia is available in English, French and Spanish.

6.2 Medical dictionaries

The **Merriam-Webster Medical Dictionary** is accessible from its own general homepage (<http://www.merriam-webster.com/>) by flagging "medical" in the option window above the Search box before searching or, with a simpler and more direct interface, from the cited Medline Plus (<http://www.nlm.nih.gov/medlineplus/mplusdictionary.html>). You can type the word you have to find, or just begin it, followed by an asterisk (to truncate and get all the terms with this root).

MedTerm (<http://www.medterms.com/script/main/hp.asp>) is a medical glossary of over 16,000 terms hosted by MedicineNet (<http://www.medicinenet.com/script/main/hp.asp>) a medical gateway appointed by the cited HON code.

The **Multilingual Glossary of technical and popular medical terms** (<http://users.ugent.be/~rvdstich/eugloss/welcome.html>) is a project commissioned by the European Commission and developed by the University of Ghent (Belgium). You can find 1830 technical and popular medical terms in eight of the nine official European languages: English, Dutch, French, German, Italian, Spanish, Portuguese and Danish.

For general purposes, **Wordreference.com** (<http://wordreference.com/>) is a service with free online dictionaries (English, French, Italian, Spanish, and Portuguese) for translations and cross-search. For each voice a "listening button" is provided with English or American correct pronunciation.

The University of San Diego (California, USA) provides a list of **Dictionaries and thesauri** (<http://tinyurl.com/3qdbkf>) for general and specialist purposes, including acronyms and symbols.

6.3 Atlases, e-books, open courses

The **Visible Human Project** (http://www.nlm.nih.gov/research/visible/visible_human.html) is the project of the US National Institutes of Health aimed at the creation of complete, anatomically detailed, three-dimensional representations of the normal male and female human bodies. From the homepage lot of material is available for free (the files of the whole images are for charge).

Musculoskeletal Radiology (<http://www.rad.washington.edu/academics/academic-sections/msk>) is the subject of a medical gateway edited by the University of Washington (USA) School of Medicine. It gives access to a muscle atlas and to a lot of teaching material for diagnostic radiology.

MRI and radiology teaching files (<http://www.mritutor.org/mriteach/index.html>) presents a lot of images divided by body location (spine, musculoskeletal, brain...), searchable by diagnosis or unknown cases. All the images are under copyright.

6.4 E-books free on Web

There are lots of sites providing free access to e-books:

- **Bookshelf** - US National Library of Medicine is accessible from the PubMed homepage by the black navigation toolbar on the upper side of the screen - Books-, or directly at <http://www.ncbi.nlm.nih.gov/sites/entrez?db=books&itool=toolbar>. The interface is the same and well known of the old PubMed. You can type a word in the search box and the system gives back all the references it found in all the chapters of the available books (about one hundred). For some books, Bookshelf also provides links to related topics in other databases and animated tutorials that printed books cannot provide. From PubMed itself, in many abstracts, there are link to "Books" when available.
- **Virtual books** - Golden Hour (http://www.goldenhour.co.il/virtual_books.html): a list of about 40 texts, including the Merck manuals and a useful Textbook of Physical Medicine and Rehabilitation (<http://www.emedicine.com/pmr/index.shtml>).
- **e-Medicine** (<http://www.emedicine.com/>) is a comprehensive site that offers e-books divided into specialities and also a useful link to free accredited CME courses.
- **Free books 4 Doctors** (<http://www.freebooks4doctors.com/index.htm>) provides access to about 360 e-books, accessible by title or by speciality.
- **Books online** - National Academies Press (<http://www.nap.edu/topics.php?browse=1>) gives access to more than 3700 free books in lots of disciplines, among which Health and Medicine or Biology and Life Sciences.

The last initiative to report is the Boston Massachusetts Institute of Technology (MIT) **OpenCourseWare** (<http://ocw.mit.edu/OcwWeb/web/courses/courses/index.htm>): it is a Web-based publication of virtually all MIT course content. Sections to be seen: Biology, Biological Engineering, Brain and cognitive sciences, Health and technologies.

6.5 Medical images

If you need a medical image - beware of the copyright law, not all of the sites listed below are free - you can have a look at:

- **Yale Image finder** (<http://krauthammerlab.med.yale.edu/imagefinder/>) is a database of over 544,000 (and growing) images coming from the Open Access repository PubMedCentral (<http://www.ncbi.nlm.nih.gov/pmc/>). It means that each image is freely reusable, according to the Open Access principles, as we have seen above, 2.3.
- **HONMedia** (<http://www.hon.ch/HONmedia/>) is a repository of over 6.800 medical images and videos, pertaining to 1.700 topics and themes, edited by the Health on the Net Foundation. Search by keyword or by steps (category, subcategory, voice).
- **Medical Images - TRIP** Turning Research into Practice, the EBM search engine (<http://www.tripdatabase.com/>) displays on the right side of the results list a box with the corresponding images for the searched keyword.
- **Martindale's Health Science Guide** (<http://www.martindalecenter.com/HSGuide.html>) offers more than 4.800 images and about 1.000 videos in its disciplinary sections. In some cases are listed also atlases and textbooks.
- **Google images** (<http://images.google.it/>) offers a good response for medical images.

7. HOW TO ORGANIZE CITATIONS AND WEBSITES: REFERENCE MANAGER TOOLS AND SOCIAL BOOKMARKING

Researchers in the age of Internet cope with the so called “information overload”. There are useful tools to simplify both the process of bookmarking, tagging and retrieving Web sources of interest, and the process of recording, managing and importing bibliographic citations into texts.

7.1 Social bookmarking: share your personal bookmarks within the research community

Imagine extending the service of your browser's bookmark to the whole research community: this is the concept of “social bookmarking”, where your list of websites or articles of interest are made available to anyone. Unlike in file sharing, the *resources* themselves aren't shared,

merely bookmarks that *reference* them. Content is organized by “tags”, which are a sort of label or keywords any user can add to a resource to classify and remember it, according to his/her personal vocabulary. So, other users may understand the content of the resource without first needing to download it for themselves. The advantage of this system is that content is accessed, rated and tagged by human beings who understand the meaning; on the contrary, search engines only works on algorithms to calculate the most linked website. Moreover, a social bookmarking system can rank a resource based on how many times it has been bookmarked by users, which may add value to the resource itself: collaborative tagging is a form of self-organizing system.

Wikipedia provides a list of the most used social bookmarking Web services (http://en.wikipedia.org/wiki/List_of_social_bookmarking_websites#Social_bookmarking).

We shall focus on **Connotea** (<http://www.connotea.org/>), which is the most research-oriented one.

Connotea is a Web-based service, so you don't have to download any program to your computer. Being a Web-based service means also that, once logged in, you can access your list of bookmark from any computer with an Internet connection. When you first register in Connotea, you can add a button to the bookmark toolbar of your browser; it lets you easily add references.

When you are reading a Web page that you want to save a reference for, you just have to click the ‘Add to Connotea’ browser button: a pop-up form appears allowing you to immediately save the page, without needing to break off from your current work nor to open a separate program. Connotea will automatically fill in the title and URL for the page.

You just have to add your tags to describe or comment the item you're saving. Each reference can have as many tags as you like, so there is no need to decide between different categories. At the end, click the “Add to my library” button to save the reference. By default, the links a user saves are public for anyone to see. There is an option to make them private, or shared with just a group of other Connotea users.

Each time you sign in, Connotea opens your library. We registered a user on Connotea as “PRM1”. In this library we put all the websites we have talked about in this chapter, so you can easily access them from <http://www.connotea.org/user/PRM1>. They are clustered by tags.

On the top, the system displays a search box which allows a search within your own library or everyone else's. Please notice that this feature only searches the information held in the Connotea database (titles, bibliographic information, descriptions, tags and comments), not the full text of the saved article or Web page. On the left column, you can see all your used tags; clicking on anyone of them will display the corresponding saved items. The central column shows your items in order of addition, the most recent on top. On the right side you can find the “Related tags” and “Related users” lists, which are powerful tools to navigate and discover new content.

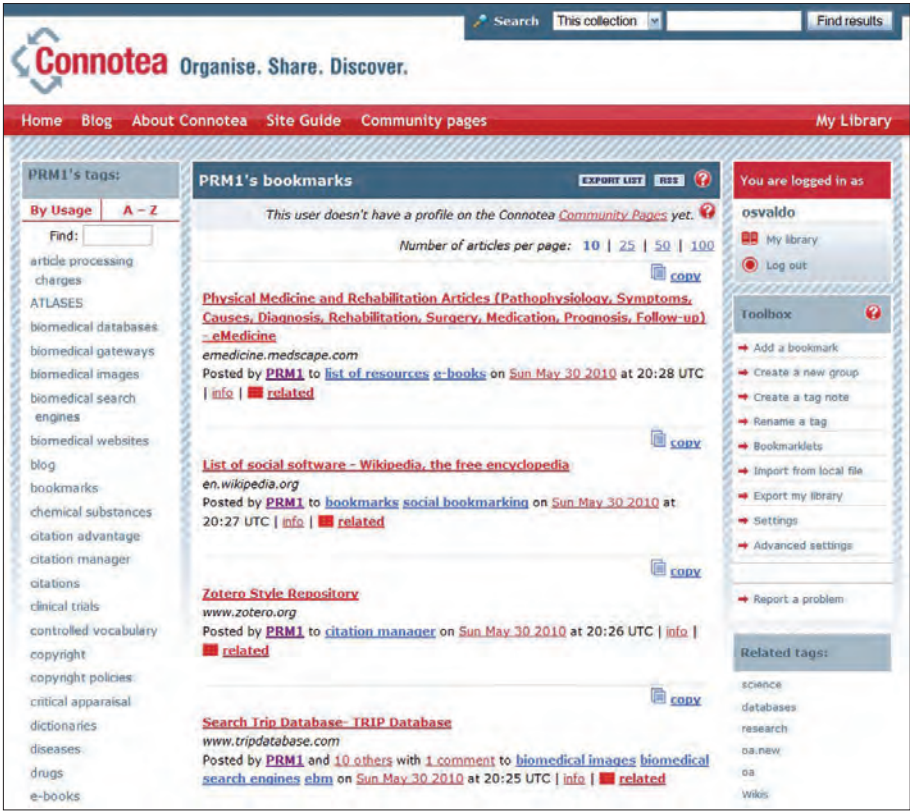


FIGURE 10. Connotea’s PRM1 library with all the websites cited in this contribution.

7.2 Reference manager tools: once stored, always available

Reference manager software packages consist of a database in which full bibliographic references can be entered, plus a system for generating selective lists according to the user’s criterion. These are very useful tools because once stored a bibliographic record, it will be available time and again for multiple bibliographies or article references. As a sample, we shall focus on **Zotero**, which is free, performing and easy-to-use. Zotero is a Mozilla Firefox extension, which means it works only within the Firefox browser. You have to download the program from the homepage (<http://www.zotero.org/>).

Zotero is, at the most basic level, a citation manager. It is aimed at storing, organising, and citing bibliographic references, such as books and journal articles. In Zotero, each of these references is an “item”. In the Zotero pane (which you can easily open from the browser Tools pull down menu) items are displayed in the central column, with metadata in the right column. “Metadata” means the titles, creators, publishers, dates, and any other data needed to cite the item. They are automatically collected by Zotero: the system “senses” when you are looking at an item (or items) on a Web page, and displays a book icon in the browser navigation toolbar, at the right side.

Simply by clicking on the book icon you will save all of the citation information into your library, in the active collection. Zotero allows you to organize your records into “collections”. Collections and subcollections are shown in the left column of the Zotero pane, under the label “My Library”. The central column shows the contents of the selected collection. Zotero allows a single item to be in multiple collections at the same time. You can tag your items to categorize them. Items can have notes, files, and links attached to them.

Zotero uses Citation Style Language (CSL) to properly format citations in many different bibliographic styles. Zotero supports all the major styles; if you need a different one, you can choose and download it from the Zotero Style Repository (<http://www.zotero.org/styles>).

Microsoft Word and OpenOffice plugins allow users to insert citations directly from their word processing software: in-text citations, footnotes and endnotes are all supported. Icons on the Zotero toolbar allow you to insert citation, to create bibliographies, to select a citation style. It is also possible to switch citation styles for the entire document at once or automatically generate a bibliography from the items cited.

If you want to create a bibliography or a custom C.V., highlight one or more references in your Collections and then right-click (or control-click on Macs) to select “Create Bibliography from Selected Item(s).” Then select a citation style for your bibliography format and choose if you want to Save as RTF, to Save as HTML, to Save to Clipboard (to paste into any text), to Print. Zotero will automatically create a numbered and alphabetized bibliography.

If you want to access your Zotero library online, you have to register to a Zotero account. Then fill in the “Sync Server section”. By default, Zotero will sync your data with the server whenever any changes are made.

For those who don’t use Firefox, a valid alternative could be **Bibus** (Bibus (http://bibus-biblio.sourceforge.net/wiki/index.php/Main_Page)). It has to be downloaded to your computer. It features a hierarchical organization of the references with user defined tags; it allows a multiuser environment. As Zotero does, it supports the insertion of references and formatting of bibliographies in Open Office and Microsoft Word, with its own Style Repository (http://bibus-biblio.sourceforge.net/wiki/index.php/Styles_repository). It performs “Live Queries”, which updates as the database changes. It runs online queries in PubMed and e-TBLAST, with whom it interfaces seamlessly.

HERVÉ COLLADO, LAURENT BENSOUSSAN
GUILLAUME LOTITO, JEAN-MICHEL VITON
ALAIN DELARQUE

CLINICAL CASE REPORTS IN PHYSICAL AND REHABILITATION MEDICINE

Department of Physical and Rehabilitation
Medicine, Faculty of Medicine,
University of the Mediterranean,
University Hospital la Timone,
Marseille, France

DEFINITION

Case studies are a common research method used in social and medical science. They consist in an in-depth investigation of a single individual, group, or event to explore causation in the hope of discovering the underlying principles (1-3). Rather than using a selected study population and following a rigid protocol to examine a limited number of variables, the case study method involves an in-depth, longitudinal (over a long period of time) examination of a single instance or event, i.e. the **case**. It offers a systematic mode for looking at events, collecting data, analyzing information, and reporting the results. Through it, the researcher seeks to gain an enhanced insight into why the event occurred as it did, and identify what elements would be good to look at more extensively in future research. Case studies are a vehicle for both generating and testing new hypotheses (4).

Case studies should not be confused with qualitative research - they can be based on any mix of quantitative and qualitative evidence and benefit from the prior development of theoretical propositions. Single-subject research provides the statistical framework for making inferences from quantitative case-study data (2, 5).

HISTORY

The idea of using case studies for the creation of new theory in social sciences was formally introduced by the sociologists Barney Glaser and Anselm Strauss who presented their research method - grounded theory - in 1967. However, the popularity of case studies for testing hypotheses became established only in later decades. One area in which case studies have gained great popularity is education and in particular educational evaluation (6).

The casebook method, also known as the case method, is the primary method for teaching law in law schools in the United States. It was pioneered at the Harvard Law School by Christopher Columbus Langdell. It is based on the principle that rather than studying highly abstract summaries of legal rules the best way to learn American law is to read the actual judicial opinions which become law under the rule of *stare decisis* (due to its Anglo-American common law origin). To this end, American law professors traditionally collect the most illustrative cases concerning a particular area of law in special textbooks called casebooks.

CLINICAL CASE STUDIES IN MEDICINE

Traditional medical research designs in which a group receiving treatment is statistically compared to a similar group not receiving treatment may not be feasible in many clinical settings. Such investigations, typically referred to as randomized clinical trials, are a powerful method for demonstrating causal relationships between variables under controlled conditions.

Randomized clinical trials are associated with efficacy research. In a randomized clinical trial, the investigator attempts to identify homogeneous groups of patients (subjects) with similar disorders, similar levels of severity, and similar demographic characteristics. The patients are randomly assigned to a treatment or control arm. The treatment is generally governed by a highly specific protocol and every attempt is made to keep the level and type of intervention constant across the patients receiving it. The outcome measures are determined in advance, and it is important that all patients (subjects) - both those receiving the treatment and those not receiving the treatment - be measured in the same manner by the same tests. The limitations of this form of experimentation have been discussed by Kramer and Shapiro (7), who note that "*despite the obvious advantages and impressive track record of randomized clinical trials, clinical investigators have become increasingly aware of certain difficulties in their interpretation, feasibility and ethics.*" These difficulties include: the need to calculate and obtain an appropriate sample size, the ability to randomly assign participants to groups or conditions, the difficulty inherent in providing the same type, level, or intensity of treatment to all persons in the treatment groups, the problem of using the same standardized measure

for assessing outcome in all subjects, the feasibility of masking the treatment and outcome measures, and, finally, ethical objections associated with denying treatment to participants in control or placebo conditions.

Single-system design is another research and clinical evaluation methodology that has been advocated for use in rehabilitation settings (8,9). The single-system model of accountability and evaluation provides a mode for incorporating research strategies into clinical practice and building the foundation for evidence-based practice at the level of the individual patient. Group comparison designs, on the other hand, are often aimed at determining differences in the average change between two or more groups. Clinicians, however, are usually not concerned with the average change in a group but rather with change in an *individual* patient who possesses a unique set of characteristics and particular conditions. The single-system methodology thus permits the clinician to achieve two major goals: the monitoring of patient progress and the individual determination of intervention effectiveness.

Some famous case reports have provided the basis on which scientific theory was built. One of the first indications of brain function lateralization came from the research of French physician Pierre Paul Broca, back in 1861. His research involved the male patient nicknamed “Tan”, who suffered a speech deficit (aphasia); “tan” was one of the few words he could articulate, hence his nickname. In Tan’s autopsy, Broca discovered he had a syphilitic lesion in the left cerebral hemisphere. This left frontal lobe brain area (which became known as Broca’s Area) is an important speech production region. The motor aspects of speech production deficits caused by damage to Broca’s Area are known as Broca’s aphasia. In clinical assessment of this aphasia, it is noted that the patient cannot clearly articulate the language being employed.

William MacIntyre is another example highlighting the scientific potential of a clinical case. He was a British physician known for publishing the first case report of multiple myeloma (Kahler’s disease) in 1850.

CASE REPORT DESIGN AND PUBLICATION

A clinical case publication should be structured in the same way as a traditional medical research publication, with an Introduction, Material and methods, Results and Discussion.

By way of example, in their clinical case Bensoussan et al. report on the gait improvement obtained in a stroke patient after applying three treatments for spastic equinus varus foot: botulinum toxin injection (BTI), tibial nerve neurotomy (TNN), and orthopaedic surgery (triple arthrodesis), during a 7-year longitudinal follow-up period (10). They used a classical design to talk about spasticity management through a single patient. The method used was a quantified analysis of a stroke patient’s gait with a walkway system before and after applying BTI at 3 years after the stroke, TNN at 4 years and orthopaedic surgery at 7 years after the stroke.

In the results the authors describe the improvement in the spasticity after the three treatments: *"The spasticity disappeared, the range of ankle motion improved and voluntary command of the tibialis anterior became possible. Upon comparing the gait parameters before treatment and after the three treatments, it was observed that the comfortable gait velocity had increased (from 0.42 m/s to 0.70 m/s), the step length had become more symmetrical (from left 19 cm, right 57 cm to left 50 cm, right 51 cm), the step width had decreased (from 23 cm to 12 cm). In terms of participation, walking barefoot had become possible without a cane, as well as going up and down stairs and walking outdoors on uneven ground without any help."*

In the discussion, the authors evaluate and compare the different treatments used for this patient: *"BTI and TNN were found to be efficient but only for a short time. Orthopaedic surgery gave the most long-lasting results..."*. Obviously, in this kind of publication, the conclusion must open the door to future, larger studies: *"Further studies should be performed on a larger number of patients to determine the most suitable options for treating stroke patients with an equinus varus foot."*

Another example of this kind of publication is that of Luc et al. (11). Through a clinical case the authors illustrate their experience in terms of medical assessment and therapeutic strategy including fitting. The clinical assessment is based on the International Classification of Functioning Disability and Health (ICF) (12). The aim of this team was to develop a more systematic method for describing and assessing the therapeutic solutions used for patients with neurological disease. This is the first paper describing a fitting of this kind applied to a patient with spinal muscular atrophy (SMA). The walking analysis performed showed that fitting the patient with orthopaedic shoes and a patellar tendon-bearing orthosis was a highly effective therapeutic strategy. The main reason why no similar cases treated with fittings have been reported in the literature is that SMA is a very rare and variable disease.

In the same way, Mirlicourtois et al. (13) used a clinical case report method to determine whether an orthotic fitting with a combination of orthopaedic shoes, plantar orthoses and a custom-made patellar tendon bearing orthosis could improve the gait of an adult patient with generalized secondary dystonia. Complete clinical assessments were performed before fitting and one year afterwards. The patient's complaints, impairments (neurological and musculoskeletal), activities (gait) and participation were assessed. The Functional Independence Measure (FIMTM) and the Functional Ambulatory Classification (FAC) scores were determined. An instrumental gait assessment was performed with the GAITRite® (CIR Systems, Inc. Havertown, PA, USA) system (14), which can be used to measure the temporal and spatial parameters of gait. The fitting resulted in a significant improvement in gait, reduced pain and ankle instability, decreased cadence, increased step length and single foot support time, and reduced asymmetry of the temporo-spatial patterns and body weight distribution. In the case of this patient with a disabling pathology where other treatments had failed, an orthotic fitting applied to the lower limbs

efficiently restored upright stance and gait. The authors concluded that an appropriate orthosis can enable a severely impaired dystonic patient with severe deformities to continue to walk.

Many other clinical case publications could be cited in the field of Physical and Rehabilitation Medicine (15-21). All of them contribute to improving medical knowledge and build a framework for further randomized research.

While many international journals publish case reports, there are few that are devoted to publishing case reports alone. *Case Reports in Medicine*, *Journal of Medical Case Reports*, and *Cases Journal* are three such journals, publishing Open Access peer reviewed case reports in all areas of medicine. *BMJ Case Reports* is an online, peer-reviewed journal publishing cases in all disciplines. *The Journal of Radiology Case Reports* is an open-access peer-reviewed journal focusing on medical imaging. *Journal Of Surgical Case Reports* is an open access peer reviewed journal that considers case reports in the field of surgery.

CONCLUSION

Not all rehabilitation clinicians are able to carry out traditional group comparison research designed to develop or refine theory. However, all clinicians, in their daily clinical practice, have the responsibility to document the effectiveness of the services provided to individual patients. Single-system designs can be used by clinical practitioners to fulfill this purpose and also contribute to the advancement of rehabilitation science.

REFERENCES

1. Shepard, JM, Greene RW. Sociology and You. Ohio: Glencoe / McGraw-Hill. 2003, pp. A-22.
2. Yin RK. Case Study Research: Design and Methods. 4th Edition. SAGE Publications. California, 2009.
3. Case study. Available from: http://en.wikipedia.org/wiki/Case_study
4. Flyvbjerg B. Five misunderstandings about case study research. *Qualitative Inquiry*. 2006; 12: 219-245.
5. Lamnek S. Qualitative Sozialforschung. 4th Edition. Beltz Verlag: Weinhhein, Basel, 2005.
6. Stake RE. The art of case study research. Thousand Oaks: Sage, 1995.
7. Kramer MS, Shapiro SH. Scientific challenges in the application of randomized trials. *JAMA* 1984; 252: 2739-45.
8. Backman CL, Harris SR, Chisholm JM, Monette AD. Single-subject research in rehabilitation: a review of studies using AB, withdrawal, multiple baseline, and alternating treatment designs. *Arch Phys Med Rehabil* 1997; 78: 1145-53.
9. Ottenbacher KJ. Clinically relevant designs for rehabilitation research: the idiographic model. *Am J Phys Med Rehabil* 1990; 69: 286-92.
10. Bensoussan L, Mathelin A, Viton JM, Collado H, Delarque A. Improvement of gait in a stroke patient. A 7-year longitudinal study. *Disabil Rehabil*. 2010 Feb 23. [Epub ahead of print].

11. Luc M, Bensoussan L, Viton JM, de Bovis VM, Collado H, Pouget J, Delarque A. Gait recovery in a distal spinal muscular atrophy patient wearing a patellar tendon-bearing orthosis and orthopaedic shoes. *J Rehabil Med.* 2007; 39: 181-4.
12. Bensoussan L, Viton JM, Barotsis N, Delarque A. Evaluation of patients with gait abnormalities in physical and rehabilitation medicine settings. *J Rehabil Med* 2008; 40: 497-507.
13. Mirlicourtois S, Bensoussan L, Viton JM, Collado H, Witjas T, Delarque A. Orthotic fitting improves gait in a patient with generalized secondary dystonia. *J Rehabil Med.* 2009; 41: 492-4.
14. Bilney B, Morris M, Webster K. Concurrent related validity of the GAITRite walkway system for quantification of the spatial and temporal parameters of gait. *Gait Posture* 2003; 17: 68-74.
15. Collado H, Nazarian F, Le Corroller T, Bensoussan L, Viton JM, Delarque A. Osteolysis of the medial malleolus due to a sports injury - an unusual cause of ankle pain. *Joint Bone Spine.* 2008; 75: 741-3.
16. Goulipian C, Bensoussan L, Viton JM, Milhe-De Bovis V, Ramon J, Delarque A. Orthopedic shoes improve gait in Friedreich's ataxia: a clinical and quantified case study. *Eur J Phys Rehabil Med.* 2008; 44: 93-8.
17. Avril Y, Collado H, Coudreuse JM, Viton JM, Delarque A, Chagnaud C, Leluc O. Acute leg pain caused by plantaris muscle injury in a professional soccer player. *Joint Bone Spine.* 2007; 74: 660-2.
18. Guzian MC, Bensoussan L, Viton JM, Mihle de Bovis V, Ramon J, Azulay JP, Delarque A. Orthopaedic shoes improve gait in a Charcot-Marie-Tooth patient: a combined clinical and quantified case study. *Prosthet Orthot Int.* 2006; 30: 87-96.
19. Pellas F, Olivares JP, Zandotti C, Delarque A. Neuralgic amyotrophy after parvovirus B19 infection. *Lancet.* 1993; 342(8869): 503-4.
20. Grangeon M, Guillot M, Sancho PO, Picot M, Revol P, Rode G, Collet C. Rehabilitation of the elbow extension with motor imagery in a patient with quadriplegia after tendon transfer. *Arch Phys Med Rehabil.* 2010; 91: 1143-6.
21. Chaler J, Müller B, Maiques A, Pujol E. Suspected feigned knee extensor weakness: Usefulness of 3D gait analysis. Case report. *E. Gait Posture.* 2010 Jul 13 [Epub ahead of print].

NICOLAS CHRISTODOULOU¹
ROLF FRISCHKNECHT²
KATERINA CHRISTODOULOU³

HOW TO PREPARE AND PRESENT A LECTURE

¹ School of Sciences of European
University Cyprus

² Unit of Neurorehabilitation and Physical
Medicine, Department of Clinical
Neuroscience, Lausanne
University Hospital, Switzerland

³ Limassol Centre of Physical and
Rehabilitation Medicine

1. INTRODUCTION

“Lecturing is not simply a matter of standing in front of a number of people and reciting what you know. A lecture is a special form of communication in which voice, gesture, movement, facial expression, and eye contact can either complement or detract from the content” (1).



To prepare a lecture may seem to be a simple task, but the truth is that it requires a long process of study and comprehension of the subject as well as a careful and well-organised preparation. Otherwise stated, a lecture preparation requires the ability of the speaker to define the learning objectives of the lecture, gather the relevant specific scientific material, assess the scientific value of the collected documents, work out the key points of the lecture, put them into a concise text and represent them clearly and accurately, using the available aids in front of an audience of a specialised background or not within the allotted period of time. It also requires a combination of a thorough knowledge and an extensive experience on the subject to be presented as well as a strict application of basic steps that will lead to the successful delivery of the lecture.

Making a first-rate presentation not only opens doors to many career opportunities and boosts the self-esteem (2), but also can influence the beliefs, thoughts and ideas as well as broaden the knowledge of the attendees in the domain of their specialization. Given that presenting well, especially at scientific meetings is important in developing one's career (3), it becomes clear the delivering of a scientific lecture is not only a demanding but also a challenging task. Reviewing recent publications on the subject to be lectured might also be an opportunity for the author to broaden his own knowledge on the relevant topic and improve regular lectures which are part of his academic tasks.

The course leading to the presentation of a lecture consists of many steps: 1) the receipt of the invitation, 2) the acceptance of the invitation, 3) the preparation of the lecture and 4) the actual delivery of the lecture. The person who is directly involved in all the above-mentioned phases is the lecturer. Before accepting a lecture, the lecturer should necessarily take time to read carefully the invitation to deliver a lecture and take into consideration all the relevant factors such as the subject of the lecture, the relation to his field of competence, the audience to be addressed, the context of the invitation, date and place of the lecture. He should take into account the sponsors involved and think about potential conflicts of interest. If the lecture will take place in another town or country arrangements for transportation, accommodation and meals should also be considered.

The present chapter will provide a detailed analysis on how a lecture should be properly prepared and delivered. The two basic phases of a lecture (the *before-lecture phase* and the *during-lecture phase*) will be clearly explained, by mentioning what should be taken into consideration and avoided both before and during a lecture. All the afore-mentioned will be supported by a series of examples thereupon provided.

2. PREPARING A LECTURE

The preparation of a lecture can be divided into two main phases, namely the *before-lecture phase* and the *during-lecture phase*. Both phases are interdependent and a great emphasis should therefore be thereon laid.

2.1. The *before-lecture* Phase

As the Greek proverb goes, “the beginning is the half of all things”, the *before-lecture phase* is indeed the very beginning of a lecture preparation on which the success of the lecture to be presented is based to a great extent. It is well known that about 80% of the effort comes before the presentation. The actual talk comprises only 20% (4). Only if the lecture is meticulously prepared, an outstanding presentation can be achieved. The before-lecture phase comprises everything that has to be taken into consideration before the delivery of the lecture. There is a number of key factors, namely the *where*, *how*, *who* and *what* factors, that are of vital importance for the speaker to know.

2.1.1 The “*what*” Factor

Being aware of the type of lecture to be given is of vital importance for the lecture preparation. The speaker must always be informed from the very beginning whether it will be a university course lecture, a talk at a review seminar, a state of the art lecture at a congress or a presentation of his own work at an international meeting, considering that the type of the lecture dictates the way one assembles it (5). What is more, the speaker should identify and perfectly understand the scope and purpose of the lecture to be delivered, in order to clarify the material to include and the points to stress during the lecture. Beyond the specific scientific content, the intent of a lecture can vary: spread information or knowledge, convince the audience about a fact, a practice, a theory, etc. or stimulate a debate. The context of the lecture must also be known. What is the theme of the university course the lecture is integrated into? What is the subject of the congress or the theme of the session the lecture takes place? It could be underlined that usually, a university course lecture or a seminar talk requires few of the lecturer’s own data, whereas a presentation at a meeting with specialists often comprises the lecturer’s results, set in the perspective of a research sector (5).

2.1.2 The “*who*” Factor

The speaker should know its target audience and consider its needs both before and during a lecture. Otherwise stated, the audience’s interests, background and expertise will shape an effective presentation (6). In general, the amount of the technical details, the types of examples and the major goals of a lecture vary according to each audience (6). The broader the audience is, the more background information must be included and the more results must be explained on the way (5). Moreover, a lecture addressing experts of the speaker’s field may include far more results and deductions and promote further research on the subject (5).

On the first hand, using specific medical terminology in front of an audience that comprises doctors and specialists of the speaker’s domain is likely not only to draw their attention but also to enhance the intended message. On the other hand, using the same specialized terminology in

front of listeners with little background is likely to provoke confusions, misunderstandings and their indifference about the subject. It is generally known that listeners are happier when the presented material is so simply and clearly delivered that they can follow all parts of a lecture (5). Lastly, what the lecturers should always remember is that in lecturing it is almost impossible to underestimate the audience. As a rule, the audience should always be overestimated (5).

2.1.3 The “where” Factor

Lecturing is a performing art and good staging helps the audience to retain the content of the lecture. Therefore knowing a maximum about the place where the lecture will be delivered and the technical equipment of the lecture hall is important for a successful preparation and staging of the lecture. The speaker should get this information well in advance in order to arrange accommodation and transportation so as to arrive at the lecture hall relaxed and unaffected by stress and time lag. Whenever possible, the speaker should familiarize himself with the location and its surroundings before the lecture by visiting the place in person. If the speaker lives close to the location of the lecture he should visit the place well in advance, if he is traveling from far, he should check out the lecture hall a few hours before lecturing. Such a visit allows him to know the characteristics of the room, the position of the lecture desk and the screen as well as the type of microphone available (fix or mobile). He will also be able to estimate the number of people who are likely to attend the lecture and get informed about the projection equipment available, the lightning and the podium height (6). The speaker has also the right to order any additional equipment he may need for the lecture purposes, check whether the available ones work properly, or learn how to operate the lightning and the sound system of the room. Many lectures are hampered by incompatibilities between the programs of the computer at disposal in the lecture hall and those used for preparing the lecture. Therefore it is advisable to check this out beforehand. It might be cautious to have a pdf file ready if there is any doubt about the compatibility of computer programs. During a visit on site other practical information such as connection with public transport, parking space etc. might be gathered more easily. Nowadays many universities and convention centers display the plans of their lecture rooms and information about the equipment which is available on their website. It can be worthwhile to check these websites if a visit on site is not possible.

2.1.4 The “how” Factor

Once the objective, the type, the audience and the place of the lecture are defined, it is mandatory to identify the way in which the lecture should be presented in terms of the language used according to the audience background, the material to be collected according to the subject and the “take home messages” according to the scope. The lecturer should also identify the visual aids to be used and create a written version of the talk in order to practice before the actual delivery.

2.1.4.1 Language Used

A language is a system of signs (indices, icons, symbols) for encoding and decoding information and the ability to use speech originated in remote prehistoric times (7). It is clear thus, that being able to express in words everything that exists in mind as well as to convey successfully the intended message to a number of people sitting in front of you indicates the great gift of mastering a language. The speakers of international lectures could be characterised as cultural mediators, since through their way of talking they constitute a bridge that connects people of different scientific backgrounds, cultures and most likely of different languages. Given that most international medical or scientific conferences are held in English (8), all lecturers are required to have a proficient knowledge of the English language, be able to state and explain clearly the basic points of the lecture and to conduct easily a discussion on subjects of a specialised domain, such as medicine, technology, computer science etc. The challenge is to keep brief written materials in order to expand verbally beyond what is written, to change the word order, to include more interesting points, examples or clarifications (8). The verbal expression will help the non-native English speakers to understand the meaning both in writing and orally and the native English speakers to remain engaged by the additional oral information (8). The loud speaking should be avoided since it does not increase the non-native English speakers' comprehension. On the contrary, loud speaking is likely to provoke fatigue and headache to the audience, resulting to their being absent-minded. At this point, it should be stressed that common terms or words used in the specific domain of specialization would be more preferable in order to ensure a better and more to-the-point understanding. The speaker should carefully opt for the words to be used by taking into consideration the audience's domain of specialization as well as its cultural background. He/she should avoid idioms, contractions or metaphors in order to facilitate the comprehension of both non-native and native English speakers. Finally, one should never forget that the simple spoken word remains a cornerstone of communication (5). The success of those standing at the podium depends on what they say and how they say it (5). Verbal skills are therefore of vital importance for a competent lecturer.

2.1.4.2 Material Collection

Assessing the right amount of material is one of the greater challenges (5). What facilitates the material collection is always the identification and full understanding of the lecture's scope. The lecturer should not only have a professional knowledge about the subject to be presented, but also be able to decide on the main points and present them correctly in order to achieve the aim of the lecture. Moreover, according to the time available for the lecture the speaker should determine the "take home key points" of the presentation (2) and gather them in a concise form of presentation.

Practically “What” and “Who” will determine the learning objectives of the lecture. The speaker should first update his knowledge on the field to be covered by the lecture by bringing together related recent data and publications. From the established current state of the art, he then must select the key information to be included in his presentation. The choice of these key points should take into account the time allocated for the lecture, the relative importance of the various facets the topic as well as the relative importance and the quality of the different pieces of literature gathered.

Once the speaker has assessed the right amount of material, he should start creating illustrations in the form of slides, of which no more than 25-30 should be included in a 45-minutes lecture (5). Some authors are even more restrictive and recommend not more than 1 slide per 2 minutes of lecture (9). In addition, the speaker should prepare an amount of material that takes 10% less time than the allocated (5), in order to manage to conclude the presentation on time, by taking always into consideration the time needed for questions, answers, further clarifications and the conclusion. Finally, it should be stressed that knowing the material well and having sufficient material, gives the lecturer confidence and flexibility at the podium (2).

2.1.4.3 Visual Aids

A presentation is always more vivid and comprehensible when supported by visual aids. People assimilate messages quicker by hearing than by reading, but they see more rapidly than they hear (9). This is why visual aids are a precious asset to support a lecture. Visual aids can be effective in maintaining the attention of the audience, but only if they are used in judgment (10). There are many types of visual aids such as maps, graphs, charts, photographs, pictures, posters etc. The most commonly used visual aids for lectures, and more precisely for the scientific lectures, are the slides. The use of the “PowerPoint presentation system” has become very widespread. Preparing slides needs both a scientific and an artistic approach (4). Slides are considered to be quality visual aids for presentations with numbers, figures or new concepts (4). A combination of a scientific knowledge of the subject and an artistic taste on how to present and convey the intended message best is required. The use of slides in a presentation is widely practiced and has many advantages indeed.

Slides help the audience to concentrate, save the efforts and the time spent on explanation and improve effectiveness of the message (4). They improve comprehension as complex facts or processes can be visualized or visually broken down into simpler parts. The illustration of the talk by presenting real objects and situations and using simulations where appropriate will enhance understanding. Slides also facilitate memorizing the content of a lecture in many ways: by allowing the messages to be encoded visually as well as verbally, by structuring the presentation visually, by regulating the flux of information, by enhancing key points as well as by the increased attention and comprehension (9).

Slides not only help the auditorium, but also the speaker: they can serve the speaker's notes for important points and help in pacing the talk (4).

PowerPoint presentations are very convenient for adapting the content of each slide to the knowledge of the target audience or to the allocated time. The slides are used to support and reinforce the talk (4), but should never replace or even compete with the speakers discourse (9). If the listeners are fully occupied to read the slides, they will no longer follow listen to the speaker. The same observation applies to other types of visual supports and hand-outs.

A quality slide should be appropriate, accurate, legible, comprehensible, well executed, interesting and above all memorable (4). For each slide included into the presentation, there should be a good reason for using and showing it. Every slide should be relevant to the specific subject of the lecture and placed at the appropriate moment of the talk. The slides should be very carefully and cleverly chosen and should contain accurate information and spellings, including names of people, organizations, institutions etc., where deemed necessary. Moreover, they should be legible in order to avoid the fatigue of the attendees eyes and minds. Each slide should contain only one main message or idea and should be designed so as to be understood by its own. As visual aids, slides should be designed as such in order to be captured as graphics. The brain analyses audio-verbal and visuo-spatial input through different neuronal networks. People take in messages presented in form of simplified pictures, graphics, diagrams, drawings, etc. more rapidly than written texts. Therefore texts should be limited and laid out like graphics. As mentioned above, people are more rapid in hearing than in reading and there is no point replacing the spoken word by slides. Whole sentences should be avoided and citations only given if they are important or if they support a reflection (9). It is preferable to use an 18-point font for the slide text and a minimum font size of 36-points for the titles on each slide (4). Some authors even recommend a 24-point font or higher and advocate bold lettering for enhanced visibility (9). Simple fonts such as Arial, Tahoma, Zurich or Times New Roman should be used instead of a fancy calligraphic fonts or serifs that prevent the easy reading of the phrases. Different font sizes may be used to organize information into different levels of hierarchy. In general, there should be no more than six to seven lines of material per slide, including the title and subtitle. The individual lines should be no more than six to seven words in width. Upper and lower case letters are more preferable than capitals (9, 11). Additionally, considering that a first-rate presentation is an effective combination of verbal and visual elements, it is good to know that the use of illustrations, proper background colours and animations in a PowerPoint presentation help to explain and emphasize concepts that cannot be verbalized (4). Any illustrations to be used are likely to maintain the audience attentive as well as to clarify, restate, explain and interpret (4) all the important points or concepts of the lecture. The use of animations must be carefully measured. They can be used to explain a complex concept, fact or situation starting from a simple scheme and proceeding progressively to

more complexity. The evolution in time of a situation, a case or a concept can be presented step by step. They can be used to gauge the information on a slide or to highlight a part of an illustration. Fancy pictures and “aesthetic” inserts not related to the topic of a talk may hamper the communication rather than improve it (9). Should the speaker decide to use charts and graphs in its presentation he should carefully opt for the right ones, aiming to make the audience to absorb the point at a glance (4). For instance, the horizontal bar chart can be used for comparing items at one point in time, while for comparing parts of a whole, a pie chart is more appropriate (4). The charts should be made easy to read. For this purpose, colours, shading or arrows could be used in order to highlight the key words or concepts. The most important text could become larger or the most important data lines darker (4). Any element of a graphic which is not part of the message or needed for its comprehension distracts the public's attention and should be erased. As far as the illustrations are concerned, it should also be mentioned that tables are also effective if percentages are used for comparison (11). It should be stressed that crowding a lot of words, terms or concepts in small boxes or spaces should be avoided, in order to facilitate the clear view of what is written in a slide.

Should the speaker decide to use photos, he should carefully opt for those who will neither provoke any disgusts to the audience nor violate anyone's right of privacy. Otherwise stated, should a photo illustrate an injury, it would be preferable to opt for an illustration rather than a picture taken in a real life accident. Additionally, should a photo illustrate a related person but it is considered useful to be included in the lecture, it would be ethically proper to obtain that person's consent before using it.

Colors make slides easier to memorize in several ways. Colors attract attention, and can be used either to highlight key messages or to structure the content of a lecture by creating color-coded categories of information. Colors are associated with emotions which in turn facilitate memorization (Fig. 1).

Calm Security	Relaxing Harmony Stimulates interactions	Stability Strength / Power Clarity
Stimulating Optimistic Contrast	Warmth Enthusiasm Stimulates communication	Energy Danger Dominant Stimulates action

FIGURE 1. The emotional dimension of colors [modified from (9)].

Hence colors can help the auditorium to retain the content of the lecture. However there are a few principles to be taken into account (9): limit the number of colors to 3, maximum 4 including black, use cold colors rather for backgrounds and warm colors for text, use contrasts which make reading easy and are comfortable for the eyes, yellow text on blue background is very ergonomic but widely used. The dimension and the lighting of the conference room matters: colors tend to become brighter in small rooms and darker in big rooms. In dark rooms bright colors on a dark background give a better result than the opposite. In bright rooms dark colored text on bright background should be preferred (9, 12).

There are also some general rules about the lay out of a slide. In Europe people generally tend to skim through a slide from left to right, from the top to the bottom and clockwise. To take into account the human visual exploration pattern, the key elements of a slide should be put either in the center of the slide or at the intersection of the lines which divide the slide into thirds vertically and horizontally (Fig. 2).

The information should not occupy more than 50% of the space available on the slide (9). The consecutive slides of a presentation should be structured in the same way (12). The part of the slide reserved for the title and the part which carries the information should also be separated

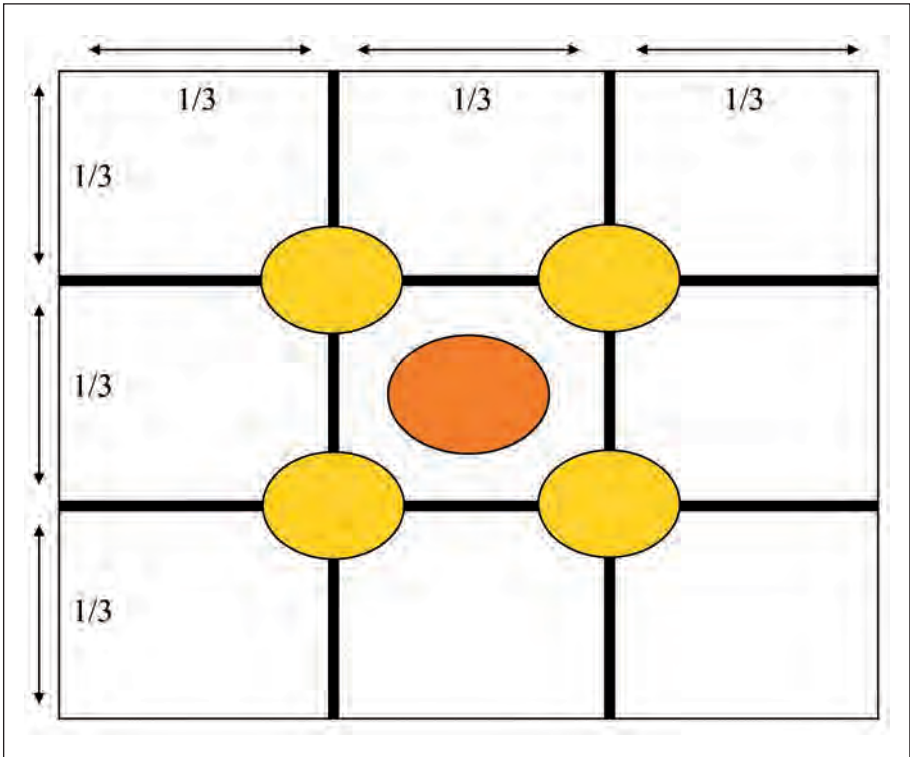


FIGURE 2. Slide zones which receive the most attention immediate attention by the viewer [modified from (9)].

graphically in a similar manner throughout the talk. These techniques minimize the cognitive effort to assimilate the content of the slide, as the brain does not lose time and neuronal resources to analyze the general layout of each slide. Knowing how the information on each slide is organized the listeners brain can focus its attention directly on the message the slide is intended to support.

All the above-mentioned should be included in a well-organised slide layout in order to provide a successful presentation. That is to say, the first slide should show the title, the presenter's name and the presenter's institution. This should be followed by an introductory slide which will be followed by the main contents. The last slide should be the summary or the conclusion, in conjunction with the references and the acknowledgments. Each slide should be self-explanatory and have a title (4). At the end of his work the author should assess the slides asking the following questions (9): Can the slides be understood rapidly? Do they naturally direct the visual attention of the audience to the key information? Do they capture the public's attention easily? Do they support the keypoints of the lecture? It should be underlined that, according to the rules, if a slide cannot be understood by the audience in four seconds, the slide is bad (13).

It is important for the speaker to know in advance the exact content and purpose of each slide and be able to explain it in his own words, without uncertainties and ambiguities. The experience and thorough knowledge on the presented subject increases the speaker's confidence, prepares him for possible questions on the slides included in the presentation and enables him to answer clearly, accurately and to the point. Finally, it would be helpful to provide a copy of the presentation and distribute it to the audience before starting the lecture.

2.1.4.4 Writing the Talk

Writing down the talk to be delivered shows the professionalism and indulgence of the lecturer and is a step that should never be ignored. The writing-down procedure has several benefits for the speaker and ensures a successful delivery of the lecture. It helps the speaker not only to clarify what will be presented, but also to see clearly the sequence of the points to be mentioned. This will lead to the identification of any gaps in logic or content (2). One of the best approaches of writing a lecture is to start with the "whole text" and then highlight or underline the core points, before going on to the short version (5). An ideal writing of a lecture should end up with a "catchword manuscript" (5), in the form of cards. The cards could contain words, concepts and phrases to guide the lecturer. At this point, it should be stressed that the speaker should always use the same words in writing as he used in the slides (11), in order not to be confused with the specific purpose of each slide. However, notes should not be read but memorized only (2).

The speaker's being familiar with the subject is the best way to avoid nervousness when at the platform (10). The written form of the lecture should not be taken to the podium, however some relevant important notes

would be of vital importance for the speaker. For instance, it would be helpful for the speaker to try to invent some methods that are likely to facilitate the delivery of the lecture. He could highlight specific terms, in order to remember to mention and explain them at the time of presentation.

2.1.4.5 Prepare and Practice

What is more, a speaker should never forget that practice and more practice is the only ticket for an outstanding presentation. Preparation and practice will help the lecturer arrange his speech according to the time that has been allotted. All speakers should keep in mind that it always takes longer to present that which they have written than they think (14). It should also be stressed that going beyond the assigned time shows the speaker's poor planning and it is usually interpreted as arrogance (14). The lecture practice could be in front of a mirror, a colleague or a family and is surprisingly helpful for the speaker's personal progress. The speaker should pay attention to the pitch of the voice, the speed of delivery and the breathing (2). It should also be highlighted that rehearsing always boosts the speaker's confidence, despite the fact that everyone is always anxious before walking to the podium. However, the most important thing is that when the speaker is properly prepared everything will be under control and he will manage to handle the stress in the very first minutes.

Finally, it is worth mentioning that writing down the speech helps the lecturer to convert the presentation in a manuscript for publication purposes, fact that is of vital importance for him and his profession (2). A tip for preparing the manuscript is to incorporate the audience's feedback and queries in order to enhance the message (2).

3. DELIVERING A LECTURE

3.1 The during-lecture Phase

As mentioned earlier the actual talk comprises the 20% of the whole preparation. At this stage, it is time for all theoretical aspects that have been studied and considered to be put into practice. During the talk, the speaker must be fully concentrated and think various things of vital importance at the same time. The main steps towards a successful delivery of the lecture will be analyzed in the following pages.

3.1.1 Time Preceding the Talk

First and foremost, it is clear that a skillful and professional lecturer should verify that he has all necessary equipment before leaving, such as the slides (a PowerPoint presentation in a USB), notes and the copies of the presentation, a pointer, his business cards and a watch. The slides and notes will be the supporting materials during the lecture and the copies will be distributed to the audience in order to have the key points already indicated, gaining thus more time for listening than for

writing. The pointer will facilitate the presentation in cases of emphasizing a word or phrase in a slide. As far as the business cards are concerned, they are likely to be used at the very first meetings of the lecturer with new colleagues, since they include the lecturer's contact details. Finally a watch is undoubtedly useful for keeping up with the time that passes so quickly without realizing it. In other words, a watch will remind the lecturer about the hour, in terms of the lecture's start time, the break's time or whether he will have to accelerate in order to finish the speech on time. It would also be recommended for the lecturer to have a light meal before the lecture in order not to feel sleepy and tired.

When arrived at the place of the lecture, the speaker should firstly check out the room (11). He/she should ensure whether the visual aids function properly, in other words to check out the microphones and the projector. He/she should also determine how to get to the podium and check the podium's height so that the audience can see him/her (2). Then he/she should load his/her PowerPoint presentation on to the computer utilized as one placed slides into the carousel in the past (11) most preferably by himself or with the co-operation of the technician in charge.

The lecturer should also expect to be a little nervous (11). Any signs of panic, such as perspiration, tachycardia or thirst should not worry him, since they are totally normal during the minutes before a lecture. However, he/she should try to stay calm and concentrated in his/her scope. It would be recommended to use the restroom during the time preceding the talk and avoid drinking alcoholic beverages and large quantities of liquids. The only thing that the lecturer should think of, is that the lecture to be delivered is significant for his/her future career and though that he/she will manage to prove what he/she is able for.

3.1.2 During the Actual Talk

First of all, the speaker should take into consideration that the oral communication plays an important part in the exchange of scientific information. The main purpose of a congress-oral discussion between participants can be achieved only if a contribution is heard and understood (10). Traveling all around the world in order to attend a seminar or take part in a seminar as a lecturer is usual for many scientists, such as doctors, university professors etc. and they should all feel duty-bound to be informative, interesting and concise (10). Therefore, it is clear that the lecturer should be well aware of what he/she is committed to do and take into consideration the following significant factors.

3.1.2.1 At the Podium

First and foremost, the speaker should always remember that once he/she is at the podium, he/she is not only a researcher, clinician or administrator; he/she is a teacher (8). He/she should keep in mind that the audience has come for him/her, waiting to glean new information and broaden its knowledge, and that is enough for him/her to be motivated and do his/her best. It would be polite enough to introduce him/herself

and his/her topic, being smiling and calm and let the audience know that he/she is glad to be there. After three deep breaths the lecture may start and the only thing left for the speaker is to enjoy it!

3.1.2.2 *The Introduction*

The introduction of a lecture is arguably the most important part of a lecture (5). The opening words of a speech must be simple, easily understood and carefully slanted towards the interests of the audience (10). During the introduction, the speaker should firstly define the goals of the lecture, state the major tasks he wishes to fulfill and briefly explain how the latter will be accomplished (6). According to a general rule, no more than three major goals should be undertaken in a single presentation (6). Moreover, the speaker should explain why the lecture is important to his audience, by drawing from specific examples within their background, in order to convince them to listen seriously to him/her (6). The main goal while talking is to engage the audience, share the collected material and keep everyone awake (8). There are several methods to achieve a successful introduction. Such methods are the following:

- Opening with a *narration* may arouse the interest of the attendees, considering that people like hearing stories (10).
- Opening with a *quotation* can be effective if the one chosen is relevant to the speech and points toward a direct statement of the speaker's purpose (10).
- Opening with a *rhetorical question* will center the attention on the purpose of the study and make the audience think about the main subject (10).
- Opening with a *negative statement* will enhance the "suspense" (10) and make the audience willing to see what comes next.
- Opening with a *comparison or contrast* makes a neat opening (10) and motivates the audience to think.
- Opening with a *descriptive statement* (3) helps the audience get an exact idea of what will be mentioned.
- Moreover, if the speaker is aware of the dominant interest of his audience, he could use it to establish an understanding with them (10).
- Finally, opening with a *funny story* (10), where deemed necessary usually establishes a friendly relation between the speaker and his audience and releases the tension.

The afore-mentioned methods may be used according to each lecturer's personal way of conducting a speech as well as to his personality.

3.1.2.3 *Main Body of the Lecture*

Once the purpose has clearly stated, it is more or less known what is likely to come next. The body of the talk is always based on the organisation of the material already collected. The speaker should start talking based on the slides' order; by trying to convey his/her ideas in short and clear statements. He/she should also use linking words and phrases in order to switch from one slide to another, trying not to make long pauses.

The statements should be clear and reasonable. Every word should be chosen and used appropriately and precisely (10), always according to each slide. However, the speaker should not stick on the slides and be able to develop the slide's content in his own words. Should he/she feel that definition of a term is necessary, he/she should give the definition when the term is first mentioned. Moreover, if there are experiments to share, they should be closely related to the lecture's purpose and explained in brief words, due to the time limit. Every minute of the talk is precious and should be exploited to the maximum.

3.1.2.4 *The Audience*

The speaker should never forget his/her audience and leave it in the dark throughout the talk (10). The speaker should face his/her audience and never turn his/her back to it during the presentation. The speaker's challenge at this stage is to identify the learning needs of the audience in order to modulate his/her message. That is to say, the audience may need to collect new information about a specific subject or get informed about the results of the speaker's experiments. Therefore, the speaker will adjust his/her speech accordingly. It would also be advised to observe the activities or facial expressions of the audience and occasionally ask "Am I going too quickly?" (6).

What is more, the speaker should take into consideration the attention rate during lectures. According to scientific studies, as the lecture proceeds attention spans become shorter and often fell to three or four minutes towards the end of a standard lecture. Adult learners can keep tuned in to a lecture for no more than 15 to 20 minutes at a time, and this at the beginning of the class (15). In 1976, A. H. Johnstone and F. Percival identified a general pattern according to which "after three to five minutes of "settling down" at the start of class, the next lapse of attention usually occurs some 10 to 18 minutes later, and as the lecture proceeds the attention span becomes shorter and often falls to three or four minutes towards the end of a standard lecture" (15). The said general pattern is illustrated in Figure 3 (16).

After twenty minutes about 50% of the audience will no longer listen to the speaker. That is to say, that the speaker should try to state the most important points during the first 20 to 25 minutes of the lecture and most preferably proceed with examples for the next 10 minutes, leaving some time available at the end for a conclusion that summarizes the talk and reinforces the key messages as well as for discussion or potential queries. Another technique is to relaunch the public's attention after about 20 minutes by asking a question to the public or rendering them active in an other way.

As far as the queries are concerned, the speaker should devote some time at the end of a lecture for questions. He/she should listen carefully when a person asks a question and not interrupt (2). He/she should then repeat all questions in order to verify that everyone has heard the question (2), and then give a clear and brief answer, using the same style as in making the presentation (2). Finally, once the speaker makes sure that

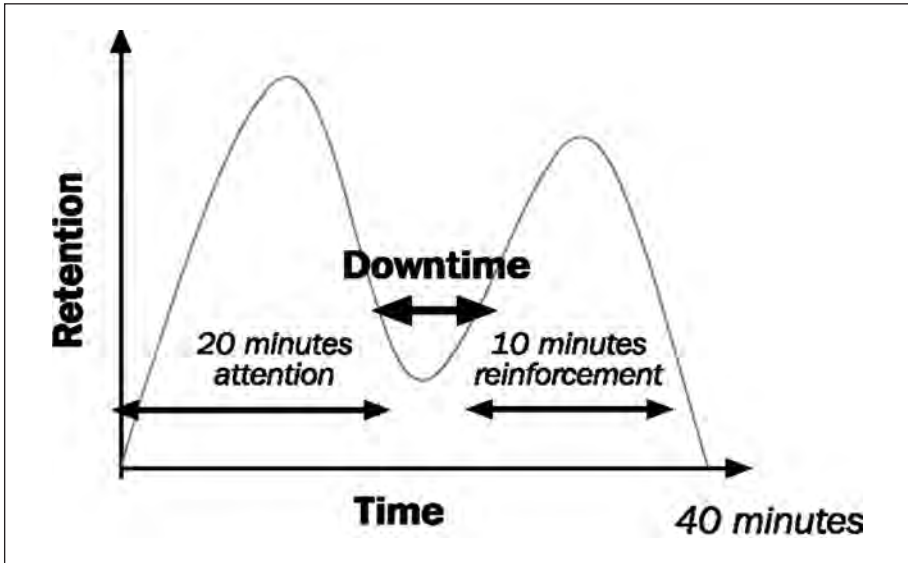


FIGURE 3. Attention rate during a lecture.

there will be no more questions, he/she should thank the people for their questions. He/she should always have in mind that the audience questions give him/her the chance to prove his/her thorough knowledge on the subject and even to provide him/her with ideas for further research.

3.1.2.5 Speaking, Rate and Pitch

The speech comprises the vocabulary, the syntax which structures them, and the set of speech sound units that differ, creating the existence of many thousands of different types of mutually unintelligible human languages (17). The language and expressions of every human being reflect his/her inner world, personality and manners. A lecturer can be very easily judged for the way he talks. Practicing the lecture before the actual day of delivery makes better use of the voice mechanism(10). During a lecture a proper speaking rate is of vital importance. The speaker should talk at a rate of no more than 80-100 words per minute (5). It takes practice to avoid talking quickly that the words are slurred (10). Moreover, both force and pitch of the voice should vary in order for the lecturer to emphasize the main ideas and important parts of the talk (10). The speaker should also make pauses and take breaths to continue. A useful advice which was occasionally given by professional speakers is that long pauses are preferably used at the end of principal thoughts (10). That time period allows time for the audience comprehension. In addition, the speaker should avoid word and thoughts repetition. A monotone way of talking is likely to deflect the audience's attention, provoke slaps easier and fail to convey the intended message.

3.1.2.6 *Eye Contact*

The eye contact and facial expressions provide important social and emotional information and people, perhaps without consciously doing so, probe each other's eyes and faces for positive or negative mood signs (18). An experienced speaker is usually more confident and looks his/her audience straight into the eyes. The eye contact is absolutely essential to effective speaking (10). However, it is absolutely clear that it would be improper to stare at somebody. A useful advice to novice speakers is to try to establish the eye contact from the very first rows to the very last ones and occasionally pick out one to three persons among the audience and speak directly to them (10). This way may urge the other attendees to turn their heads and see who is speaking, listen carefully to his query or remarks and why not get involved in the discussion.

3.1.2.7 *Posture and Gestures*

The speaker's stance and especially his upper body stance indicate his/her self-confidence and more or less his/her experience of lectures. A good posture can contribute much to the impression of relaxation and the more relaxed the speaker appears to be, the more the audience will be at ease (10). Walking should be planned and purposeful but aimless wandering should be avoided. As far as the arm gestures are concerned, they should be let hang naturally and move fluently into the most frequently used gesture, pointing to the slides (10). Being able to locate and point to a portion of a slide while continuing to face the audience and to speak smoothly and coherently is fundamental to good oral presentation (10). On the other hand, looking constantly down to the floor or outside the windows, nervous gestures or visibly quaking knees are definitely pitfalls to avoid.

3.1.2.8 *Timing*

The purpose of the timing is to set limits of the lecture's length. The lecture should start and finish according to the time allotted and this is where the watch will be used. Considering that timing is important for a successful lecture, it could be advised to insert timing marks into the organizational plan of the talk (10). The timing marks could be words like "pause", "verify time", written in specific parts of the cards, where deemed necessary in order help the speaker delivery the talk according to the organizational plan.

3.1.2.9 *The Conclusion*

When concluding a lecture, the wheel has come full circle (5). The speech should progress from an attention-getting, tell-me-more introduction to a solid, relevant and strong ending (2). The end of the speech should be the strongest point considering that it constitutes the last impression of the audience. It should be neither too short nor too long. It should be delivered in a firm voice, mentioning briefly for once more the main points of the lecture and the arisen key results, helping those who were attentive

only at the start to receive the intended message. The attendees should leave the lecture with the important “take-home points” written in their notes. The speaker should not forget to name his/her co-authors, where applicable and also propose closely related subjects that could be presented in a form of lecture at a future stage.

Finally, a well prepared, practiced and delivered presentation may lead to the increased interest in and comprehension of the speaker’s work. This may offer opportunities for networking, collaborating, funding and publishing. As a consequence, a well-delivered lecture is rewarded accordingly.

4. CONCLUSION

A lecture may constitute a benchmark for the career development of many scientists. It offers them opportunities not only to share their knowledge but also to propose innovative ideas in their domain of specialization. It could also be considered as an occasional reunion of a large group of specialists of the same domain, who seek solutions to dominant problems of their domain as well as proposals for further research. That is a considerable proof of the significant technological and scientific development of the 21st century. Consequently, the more the science develops the more innovative and original should the lectures be. The lecturer should be constantly up-to-date with the progress of their field as well as the new methods of presentations. The truth is that the quality of a lecture reflects the passion of the lecturer towards his/her subject of specialization.

Preparing and presenting a lecture is a great challenge indeed for specialists who seek evolution and professional development. However, it requires many times of serious work and concentration in order to achieve a compilation of knowledge, findings and collected material, properly gathered in a concise text. It also requires professionalism, indulgence and full dedication on the lecture’s purpose. It is of vital importance that the lecturer is honest, informative and certain during his presentation. It is always easier to convince an audience of the message, if it is confidently and well delivered. Last but not least, the skills for achieving an eye catching and interesting lecture are not inborn; however, they may attain a high level through continuous practice.

REFERENCES

1. Gross Davis B. Tools for teaching - Delivering a Lecture. [cited 2010 June 3]. Available from: <http://teaching.berkeley.edu/bgd/delivering.html>
2. McConnell EA. Giving an outstanding presentation. *Am J Nurs* 1997; 97: 62-4.
3. Estrada CA, Patel SR, Talente G, Kraemer S. The 10-minute oral presentation: what should I focus on? *Am J Med Sci* 2005; 329: 306-9.
4. Sahu DR, Supe AN. The art and science of presentation: 35-mm slides. *J Postgrad Med* 2000; 46: 280-5.

5. Blomhoff HK. Successful Lecturing. In: Lakke P, Benestad H, Olsen BR (eds). Research methodology in the medical and biological sciences, chapt. 14. Amsterdam: Elsevier, 2007, pp. 415-432.
6. Findley LJ, Antczak FJ. How to prepare and present a lecture. JAMA 1985; 253: 246.
7. Wikipedia - the free encyclopedia. Language. [cited 2010 March 31]. Available from: <http://en.wikipedia.org/wiki/Language>
8. Jacobs JL. Keep Attendees Awake, Writing Effective Presentations for International Conferences. Chest 2008; 134: 204-6.
9. Lanarès J. Pourquoi utiliser des visuels et des logiciels de présentation?. Lausanne: Centre de soutien à l'enseignement, Lausanne University; 2009. [cited 2010 July 5]. Available from: http://www.unil.ch/webdav/site/cse/shared/brochures/UNIL-CSE_visuels_presentation.pdf
10. Woodford FP. Oral presentation of a scientific paper. Scientific Writing for Graduate Students: a Manual on the Teaching of Scientific Writing, chapt. 13. Bethesda (MD): CBE, 1986: 150-166.
11. Garson A, Gutgesell HP, Pinsky WW, McNamara DG. The 10-minute talk: Organization, slides, writing, and delivery. Am Heart J 1986; 111: 193-203.
12. Berthiaume D. Communiquer à l'aide de présentations orales. Lausanne: Centre de soutien à l'enseignement, Lausanne University; 2009. [cited 2010 July 5]. Available from: http://www.unil.ch/webdav/site/cse/shared/brochures/UNIL-CSE_presentations_orales.pdf
13. Evans M. The abuse of slides. Br Med J. 1978; 1(6117): 905-8.
14. Emerson RW. The Art and Science of Presenting a Lecture. [cited 2010 June 1]. Available from: http://dr.pierce1.net/PDF/chpierce_presenting_2001_06.pdf
15. Middendorf J, Kalish A. The 'Change-Up' in Lectures. The National Teaching & Learning Forum © 1996-2010. [cited 2010 June 2]. Available from: <http://www.ntlf.com/html/pi/9601/article1.htm>
16. Milsom C. A road map for lecture writing. [cited 2010 May 31]. Available from: http://www.psychology.heacademy.ac.uk/docs/ppt/p20080507_Clare_Milsom_lecture_writing.ppt
17. Wikipedia - the free encyclopedia. Speech. [cited 2010 June 3]. Available from: <http://en.wikipedia.org/wiki/Speech>
18. Wikipedia - the free encyclopedia. Eye Contact. [cited 2010 June 2]. Available from: http://en.wikipedia.org/wiki/Eye_contact

ANDREA GIORDANO¹
GIORGIO FERRIERO²
ANTONIO NARDONE^{3,4}

HOW TO PREPARE A SCIENTIFIC PRESENTATION AS A POSTER

¹ Unit of Bioengineering, ² Occupational Rehabilitation and Ergonomics,
³ Dept. of Physical and Rehabilitation Medicine, Salvatore Maugeri Foundation, Clinica del Lavoro e della Riabilitazione, IRCCS, Veruno (NO), Italy

⁴ Department of Clinical and Experimental Medicine, University of Eastern Piedmont, Novara, Italy

Poster sessions are becoming an increasingly important part of scientific conferences, and usually represent the first occasion to gain experience with scientific communications.

Since guidance in preparing a poster presentation is widely available from scientific societies (1-4), published recommendations (5-10) and the Internet (11, 12; a Google search for poster preparation guidelines in scientific society sites reports more than 16,000 hits), our aim here is to synthesize a few simple suggestions to prevent the most common mistakes that could detract from poster effectiveness and to avoid, as Wolcott says “to give the poster presentation no one remembers” (13).



The aim of a poster is to outline some findings in a visual form that is easily assimilated and able to stimulate interest and discussion. The poster should be simple, so that any reader can grasp its main message in 1-2 minutes, but graphically appealing to capture people's attention. In fact, the average person spends just a few seconds scanning each poster and then, only if there is something of interest, reads accurately the text (6). During poster sessions, part of the efficacy of the communication relies on visual impact and part on oral presentation of the poster to peers or colleagues. Consequently, both aspects of this kind of communication have to be prepared accurately (14).

1. VISUAL PRESENTATION OF THE POSTER

1.1 Design

A poster is not a research paper stuck to a board but a different mode of presenting research using visual logic to express its essential content in a graphic way. A careful design phase, before actually composing the poster, is helpful to evaluate the effectiveness of different graphic solutions in involving the viewer.

The poster audience is composed of persons with different interests, that can be divided into 3 main groups: people who personally know the authors and their work; scientists working in the same general area as the poster's authors, but sometimes in different sub-specialties; and researchers whose work has little or no relationship to those of the authors. Poster layout has to be aimed in particular to attract the second group of poster audience, because these may provide valuable suggestions and insights about the presented research (14).

1.2 Structure

The poster is composed of different sections, usually: Title and Authors' names with affiliations, Introduction, Aims of the study, Materials and methods, Results, Conclusion, References and Acknowledgements. To save space, if not explicitly required by the organizing committee, an abstract section is not necessary, since it is already present in the abstract book of the scientific event.

Title - The title must quickly orient the reader to the main result of the work presented. To do this, it must be assertive and direct. It is better not to typeset the title in capitals because they are difficult to read and do not allow to highlight acronyms. The title, and the authors' names, affiliations, and email addresses, should be prominent and placed in the upper center portion of the poster. The Institutional logo enhances the aspect

and communicates quickly who the authors are: its size must be consistent with that of the title, and should be placed on one side of the title (4). Not necessary but a nice touch if you want to remember the occasion in which poster was presented, the logo of the scientific event can be positioned on the other side of the author's Institutional logo.

Introduction - If necessary, it will contain the background (rationale, previous research) on which the study is based.

Aims of the study - This section is short in a poster and describes the goal of the research.

Materials and Methods - Here are quickly and precisely described all the elements necessary for the specific research and essential to reproduce the study. Details should be kept to a minimum, unless the research is on a new methodology. It is a good idea to resort to drawings, flow-charts, tables or photographs whenever possible. Remember to mention statistical analyses that were used and how they were used. A more complete description could be provided verbally, or presented to interested viewers in handouts (an A4 paper reproducing the poster on display) or further material.

Results - Since it is probable there will not be room for every result, it must be considered a summary. Often this section represents the largest portion of the poster because here the reader can find the end product of the research. Quantitative results should be presented in the format of a graph wherever possible. One or more tables may be used to present some data, e.g. demographic and clinical data, but lengthy or complex tables should be avoided.

Conclusions - The last section of the poster should draw brief conclusions, giving the reader a clear take-home message. It is not a Discussion section as in a paper - in a poster there is no room to go into depth, e.g. comparing the results with those of other authors.

References - Many readers like to get information about previous results or methodological aspects described in other studies. Therefore, references are generally welcome, but should be limited to the minimum necessary, i.e. no more than 5.

Acknowledgements - Specific contributions to the project such as equipment donation, statistical advice, laboratory assistance, and funding should be mentioned here.

Each section should not be longer than 200 words, except for the title (1-2 lines), Aims of the study (max. 100 words), Acknowledgements (max. 40 words), and References.

In general, all the sections should be allocated harmoniously and in an orderly way, separated one from the other in such a way that the reader can easily and quickly find the part of interest of the communication; blank spaces can be used to highlight section breaks. It is also important not to leave large parts of the poster blank. The content (text, graphs and images) should be placed in columns, top to bottom, left to right, especially in horizontal posters. It is useful to supply clues, such as numbers and arrows, in order to guide the viewer through the poster.

Authors have to resist the temptation to include every possible detail, leaving them for the oral presentation of the poster: the content has to be concise and focused in clear messages. Given the visual nature of the poster, the text should be kept to a minimum: a sentence should not exceed 50 words, wherever possible use phrases instead of complete sentences.

1.3 Style

The style of a poster should adopt a consistent visual grammar (13) based on composition rules which are adhered to throughout the poster: e.g. the typeface of headings should be used only for headings; to highlight important information either color or bold characters should be used, not both, etc.

Text

The text should be easily readable by a person standing about 1 meter away, while the title and major headings must be readable at a distance of 3 meters. In fact, a crowd of readers could be gathered in front of the poster or some might not have perfect eyesight. Hence, the size of the typeface is essential. To improve the poster's aesthetic appeal, it is better if the text is written using different character sizes according to the importance of the sentence (e.g. title, subtitle, text) or to the specific section (e.g. introduction or references). Table 1 summarizes the suggested typeface characteristics.

Serif fonts (e.g. Times Roman, Garamond, Palatino and Century Schoolbook) are recommended for their readability; even if the Comic Sans makes the poster friendly, it is not regarded as a professional font by designers. The use of multiple typefaces should be avoided. As a maximum, for further emphasis, two fonts can be used, for instance non-serif (e.g. Arial) for text in figures and serif for the remaining text. When a fixed width font is absolutely necessary (as in reproduction of old printouts or in outputs of statistical programs), Courier should be used.

It is better to avoid capital text in whole sentences, as this can be visually distracting and annoying to the reader (15).

Colors

Color should be used for emphasis and in a consistent way. It is better to keep the number of different colors to a minimum to prevent visual confusion. The color scheme is particularly important (over and above personal

TABLE 1. Suggested typeface characteristics depending on reading distance, adapted from (14)

	Suggested Typeface Characteristics		
	Typeface size (points)	Type weight	Reading distance (meters)
Title	60-80	Bold	3
Name of authors	44-54	Bold	3
Department and Institution	40	Bold	3
Heading	44	Bold	3
Text	36	Normal	1
Captions	32	Normal	1
References	24	Normal	1
Text in figures	24	Normal	1

taste) since the combination text-background colors determines contrast. The combination dark background-light text color should be avoided since it is visually tiring. It is better to avoid green on red/pink or red on black combinations and vice versa as these can cause confusion for color-blind people. Different colors for different characters can be used to highlight sections, words or concepts such as, for example, the title or the name of some therapy or device. The background is better if it is a single, solid color; a light background image, if it does not interfere with text readability, can be used. Color can be used also to draw attention to specific areas or sections. Also colored arrows, bullets, lines etc. can be used to enhance meaning through graphic cues.

Figures and Graphs

Good graphics are essential for an interesting poster. The poster should be as attractive as possible and photos, images and graphs can help make it more pleasing to eyes that have been scanning a multitude of posters. Diagrams and pictures should be of sufficient size to be visible from a distance. Photos should have a 10x12 cm size, while drawings should be at least 20x25 cm in size. Use proper resolution in digital pictures, since enlarging low resolution images results in poor quality figures.

Charts are often the preferred way to present data and results (5). They allow long descriptions to be reduced and avoid use of complex tables. Distribution and line charts are frequently used to suggest trends and correlations; pie charts can be utilized to show sample composition while bar charts can be used to show changes in the dependent variable across groups (16). Graphs that include confidence intervals around point estimates are an effective way to present effect size and statistical significance. In multivariate analyses, it could more practical to present the results for the main variables of interest, listing the other variables in a footnote and relegating complex statistical tables to the handout (9).

Graphs usually need to be edited for poster use: for instance a short title - maybe a rhetorical question or summary of the main finding - should be provided, and the text has to be resized and lines fattened and eventually colored to make them more readable from a distance. While each figure needs to be self explanatory, do not clutter it with too much “nondata ink” such as unnecessary or excessive labeling, arrows, grids, bullets or other visual objects on the chart: see these suggestions at work (7).

2. INSTRUMENTS

A few years ago authors tended to print their posters in separate sections on many A4 sheets which they would paste together on a board. This is now considered not professional. In fact, at present authors prepare the poster on a single page electronic document, that can be printed in poster size on glossy heavy bond paper at most copy centers. Page layout applications such as InDesign or QuarkXPress can be used to design and produce large format posters, along with the most popular graphics packages such as Illustrator, CorelDraw and Freehand, but the most common software for preparing posters is PowerPoint, due to the wide diffusion of related Microsoft Office products and its availability on Mac and PC platforms. These programs allow the authors to create an image in which different components of the poster (texts, pictures, and so on) can be freely composed, placed, moved and modified until a satisfactory result is obtained; all these programs can produce an electronic document (in PDF or other formats) that can be sent to a printing or service bureau for the actual printing of the poster.

First of all, it is necessary to define the working surface of the poster once it is printed; for instance, in PowerPoint this result is obtained in “Page Setup” under the main File Menu (Fig. 1), choosing Custom Slide Format and setting layout, height and width according to the specifications of the conference organization, which are usually reported in the author section of the conference program.

Reducing the zoom, it is easy to ‘paste’ all the contents already prepared or compose them directly on the page as defined, in the first instance without too much regard for precision or effect. This first draft is useful to judge the ratio between text, images and blank spaces and the overall visual effect (Fig. 2). The necessary successive adjustments (such as text reduction

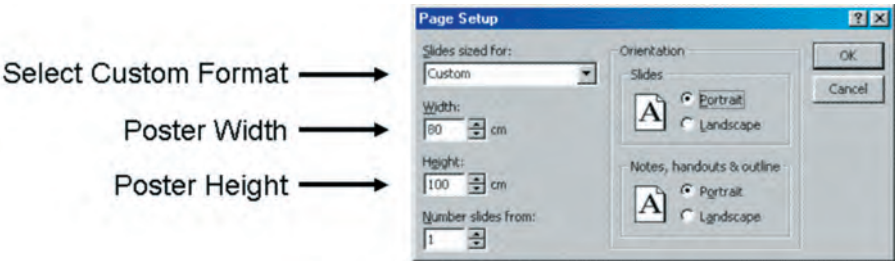


FIGURE 1. How to dimension a poster in PowerPoint.

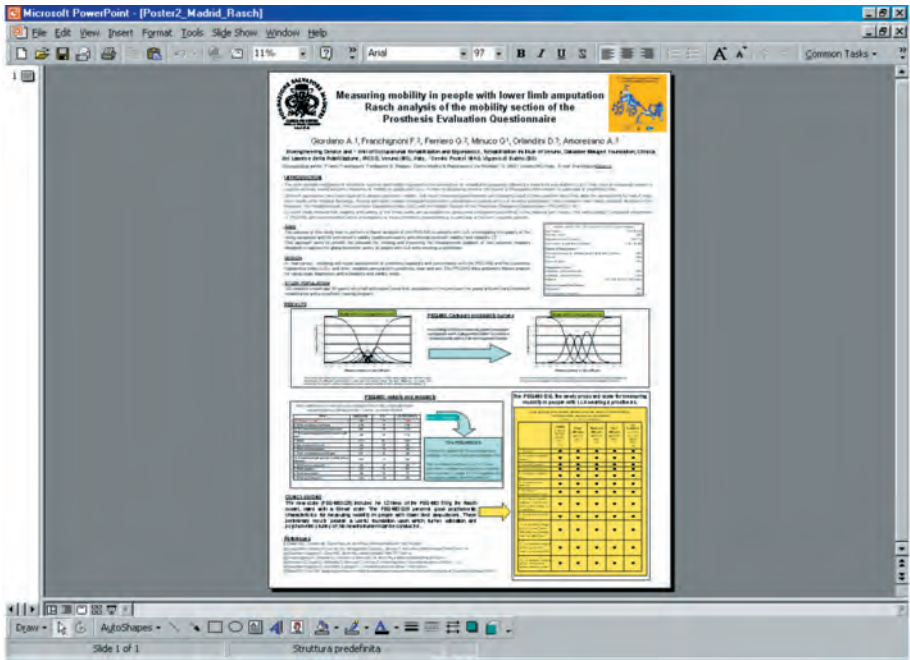


FIGURE 2. PowerPoint workplace with a sample poster

or reformatting, image enlargements, section placement, etc.) can be readily carried out on the draft, with immediate visual feedback, using the instruments available in the software.

A test printout of the poster adapted to an A3 format sheet, while possibly constituting a practical handout, is useful to judge the overall effect of the poster; moreover, if the text is readable on the sheet, it means that it is of adequate dimension to be read on the real poster.

One way to increase the reader's interest in the poster message is to offer the possibility to access online more in-depth data, such as films or publications. This innovative and interactive approach with the reader is becoming more and more common in magazines, and its use in other fields, such as poster sessions, can find new application. It consists in the presence on the paper of a two-dimensional particular kind of bar code, named Quick Response (QR) Code, printed for example near a picture or a reference of the poster. Readers equipped with a smartphone and appropriate scanning software can directly access via the phone's internet browser remote resources, as movies, photos or references arranged beforehand in an accessible repository.

3. ORAL PRESENTATION OF THE POSTER

Prepare a short synopsis of the poster according to the allotted time (generally, 3 min) and rehearse the talk. During the oral presentation of



FIGURE 3. Accessing interactive information via QR codes.

the poster, do not read the poster, but simply use it as a visual aid for the sake of clarity. When explaining the results, refer to graphs and figures but do not forget to face the audience.

Be ready to supply details left out of the poster for space considerations. These details might be made available in extra material (e.g. handouts, reprints, latest data). Sometimes readers are so interested in the poster that they write down some sentences from the poster or take pictures of it. The preparation of a handout to be freely distributed during the conference could enhance the efficacy of the poster intervention. Many copies can be placed near the poster, in a sheet protector, so that readers can take them as a memo. It is very important that the handout also contains the authors' names, affiliations, and addresses (email) for follow-up questions after the conference.

Readers may be interested not only in obtaining additional information but also in making contact with the authors for future cooperation. It is important for the authors to bring their business cards to the conference or ask for the reader's business card (on the back of which they can write some notes about the reader, e.g. questions raised, or request for scientific material). This common marketing tool may represent a good channel for creating and maintaining contacts.

4. PRACTICAL CONSIDERATIONS

Posters must be transported safely to prevent any damage. The best packaging is a poster tube with a carrying handle, of adequate diameter and length so that the poster fits comfortably in with extra room at the

ends. The poster should be rolled up gently and placed carefully inside the tube without bending or ripping it in the process. Finally, the plastic caps of the tube should be properly closed. In the case of distant meetings reached by plane another risk is the loss of the poster tube. This can occur if the poster has been packaged separately for airline transportation. In fact, the safest way to transport the poster is to carry it on the plane as hand luggage, storing it in the overhead compartment or asking the cabin crew to stow it safely somewhere inside the cabin. It is advisable to bring separately the electronic (PPT, PDF or TIFF format) version of the poster in order to be able to print a "last-minute poster" on the spot if the poster gets lost or, if not possible, to print a smaller version of the same acknowledging that this is a copy of the missing poster.

Finally, another good practical suggestion is to bring to the conference items useful to fix the poster to the display board, such as strong double-sided adhesive tape and push-pins, according to the instructions of the meeting. Do not rely on the conference secretariat to supply you with what is necessary to attach the poster.

REFERENCES

1. American College of Physicians. Preparing a Poster Presentation. Available from URL: http://www.acponline.org/srf/abstracts/pos_pres.htm
2. Campbell RS. How to present, summarize, and defend your poster at the meeting. *Respir Care* 2004; 49: 1217-21.
3. Durbin CG Jr. Effective use of tables and figures in abstracts, presentations, and papers. *Respir Care*. 2004; 49: 1233-7.
4. Shelledy DC. How to make an effective poster. *Respir Care* 2004; 49: 1213-16.
5. Block S. The DOs and DON'Ts of poster presentation. *Biophys J*. 1996; 71: 3527-9.
6. Boullata JI, Mancuso CE. A "how-to" guide in preparing abstracts and poster presentations. *Nutr Clin Pract*. 2007; 22: 641-6.
7. Erren TC, Bourne PE. Ten simple rules for a good poster presentation. *PLoS Comput Biol* 2007; 3(5): e102.
8. Hess G, Tosney K, Liegel L. Creating Effective Poster Presentations. Available from URL: <http://www.ncsu.edu/project/posters/NewSite/>
9. Miller JE. Preparing and presenting effective research posters. *HSR* 2007; 42: 311-28. Available from URL: <http://www.sph.umn.edu/news/nphw/assets/PreparingResearch-Posters.pdf>
10. Franchignoni F, Benevolo E, Martino MT. Preparare una comunicazione scientifica: il poster. *Il Fisioterapista* 2000; 3: 63-5.
11. Society of General Internal Medicine, Poster Tips. Available from URL: <http://sgim.org/userfiles/file/PosterTips.pdf>
12. Purrington CB. Advice on designing scientific posters. Swarthmore College. Available from URL: <http://www.swarthmore.edu/NatSci/cpurrrin1/posteradvice.htm>
13. Wolcott TG. Mortal sins in poster presentations or, How to give the poster no one remembers. *Newsletter Soc Integr Compar Biol* 1997. Fall: 10-1.
14. Woolsey JD. Combating poster fatigue: how to use visual grammar and analysis to effect better visual communications. *TINS*, 1989; 12: 325-32.
15. Ranger P. A Dyslexic Perspective on e-Content. *Techdis*; 2003. Available from URL: <http://www.techdis.ac.uk/resources/files/dyslexia.pdf>
16. Miller JE. *The Chicago Guide to Writing about Multivariate Analysis*. Chicago Guides to Writing, Editing and Publishing. Chicago: University of Chicago Press; 2005.

FINITO DI STAMPARE
NEL MESE DI OTTOBRE DUEMILADIECI
PRESSO LA TIPOGRAFIA PI-ME EDITRICE S.R.L.
DI PAVIA



ISBN 978-88-7963-256-0