Revista Médica del Instituto Mexicano del Seguro Social



volume 51

supplement 1

2013

ISSN 0443-5117



Clinical Research: Searching for Certainty Importance of the Dialogue Between Clinical Practice and Scientific Research Alberto Lifshitz

Medical Practice and Clinical Research: Keys to Generate Knowledge and Improve Healthcare Carla Martínez Castuera-Gómez et al.

Process Studies (Diagnostic Test) Juan O. Talavera et al.

How to Locate Articles to Answer a Clinical Question Rodolfo Rivas-Ruiz et al.

http://www.imbiomed.com

http://revistamedica.imss.gob.mx

Revista Médica del Instituto Mexicano del Seguro Social

2013

vol. 51

Supplement 1

EDITOR-IN-CHIEF Francisco Espinosa Larrañaga

EDITORS EMERITI

Francisco Olvera Esnaurrizar

EDITORS

Method and Statistics Arturo Fajardo Gutiérrez

Emilio García Procel

Abraham Majluf Cruz Ramón Paniagua Sierra

ONLINE EDITION

Gabriela Ramírez Parra

EMERITI ADVISORS

Silvestre Frenk Freund

Associate Editors

Martín Becerril Ángeles

Favio Gerardo Rico Méndez

José Luis García Vigil

Jesús Kumate Rodríguez

Alberto Lifshitz

Carlos Lavalle Montalvo

Roberto Medina Santillán

Aleiandro Treviño Becerra

Arturo Zárate Treviño

EDITORIAL BOARD

Armando Cordera Pastor Manuel de la Llata Romero Antonio Fraga Mouret David González Bárcenas

Australia

Colombia

USA

Paul Zimmet AM

Hugo Castaño A.

Fernando Arias

Jaime Davison

Horacio Jinich Brook

Luis Horacio Toledo Perevra

EDITORIAL COMMITTEES

INTERNATIONAL

Erlo Roth Finland Jaakko Tuomilehto England Graham R. V. Hughes Morocco Carlos Campillo Artero Uruguay Blanca Stéffano de Perdomo

Νατιοναι

José Dante Amato Martínez Octavio Amancio Chassin Francisco Avelar Garnica Patricia Atzimba Espinosa Alarcón Guillermo Fajardo Ortiz Ricardo García Cavazos Jaime García Chávez Fernando Laredo Sánchez Joaquín López Bárcena Gilberto Meza Rodríguez Armando Mansilla Olivares Oscar Arturo Martínez Rodríguez Lilia Elena Monroy Ramírez de Arellano Haiko Nellen Hummel Alejandro Pisanty

Manuel Ramiro Hernández Marco Antonio Ramos Corral Aleiandro Reves Euentes Hortensia Reyes Morales Enrique Romero Romero Ana Carolina Sepúlveda Vildósola Fortino Solórzano Santos Juan Osvaldo Talavera Piña Olga Vera Lastra Carlos Viesca Treviño Jorge Villegas Rodríguez Niels Wacher Rodarte María Elena Yuriko Furuya Lydia Estela Zerón Gutiérrez

GENERAL DIRECTOR José Antonio González Anaya

PRESTACIONES MÉDICAS' DIRECTOR Javier Dávila Torres

HEAD OF THE EDUCACIÓN. INVESTIGACIÓN Y POLÍTICAS DE SALUD UNIT Germán Enrique Fajardo Dolci

> EDUCACIÓN EN SALUD COORDINATOR Salvador Casares Queralt

ADMINISTRATIVE COUNCIL

Norma M. Palacios Jiménez María Edit Romero Hernández

> EXECUTIVE ASSISTANT Gloria Martínez Ferman

LIBRARY SCIENTISTS David J. Espinosa Almaguer Ana María López Jasso Alicia Zavala Delgadillo

GRAPHIC DESIGN

Ruth Jiménez Segura

MANUSCRIPT EDITING

Iván Álvarez Hernández

TRANSLATION FROM SPANISH TO ENGLISH Víctor Torrecillas

REVISTA MÉDICA DEL INSTITUTO MEXICANO DEL SEGURO SOCIAL es una publicación oficial de la Dirección de Prestaciones Médicas. Publicación bimestral editada por la Coordinación de Educación en Salud. Oficinas Administrativas: Centro Médico Nacional Siglo XXI, Av. Cuauhtémoc 330, Col. Doctores, Deleg. Cuauhtémoc, 06725 D. F. México. Revista Médica del Instituto Mexicano del Seguro Social está incluida en los índices MEDLINE del Sistema MEDLARS, ARTEMISA, ANUARIO BIBLIOGRÁFICO DE INVESTIGACIONES EN SALUD (ABISA), LILACS, PERIÓDICA, BIOSIS. Tiraje: 25 000 ejemplares en couché mate de 100 g, más sobrantes para reposición. Versión electrónica disponible a partir del 1 de diciembre de 2013. Número de Certificado de Reserva de Derechos al Uso Exclusivo de Título: 04-2009-012912585200-102, otorgado por el Instituto Nacional del Derecho de Autor. Número de Certificado de Licitud de Título: 2000. Número de Certificado de Licitud de Contenido: 1244. D.R. Composición tipográfica en Arial, Gotham, Times New Roman. Impresa en México.

CORRESPONDENCIA DE 2ª CLASE, REG. D.G.C. 015-015-0883 CARACTERÍSTICA: 229441116

Mylene Araiza Márquez

Gabriela Ramírez Parra

ISSN 0443-5117 Rev Med Inst Mex Seguro Soc. 2013;51(Supl):1-84 Los conceptos publicados son responsabilidad exclusiva de sus autores

Telephone and fax: 01 (55) 5761 2325 Email: revista.medica@imss.gob.mx

Contents

S3 Introduction

Clinical Research Juan O. Talavera

S4 Preface

Importance of the Dialogue Between Clinical Practice and Scientific Research Alberto Lifshitz

S6 Editorial

Medical Practice and Clinical Research: Keys to Generate Knowledge and Improve Healthcare Carla Martínez Castuera-Gómez, Juan O. Talavera

S10

I. Research Designs Juan O. Talavera

S16

II. Process Studies (Diagnostic Test) Juan O. Talavera, Niels H. Wacher-Rodarte, Rodolfo Rivas-Ruiz

S24

III. Causality Studies Juan O. Talavera, Niels H. Wacher-Rodarte, Rodolfo Rivas-Ruiz

S30

IV. Appropriateness of the Statistical Test Juan O. Talavera, Rodolfo Rivas-Ruiz

S36

V. Sample Size Juan O. Talavera, Rodolfo Rivas-Ruiz, Laura Paola Bernal-Rosales, Lino Palacios-Cruz

S42

VI. Clinical Relevance Juan O. Talavera, Rodolfo Rivas-Ruiz, Marcela Pérez-Rodríguez

S48

VII. Systematic Research: How to Locate Articles to Answer a Clinical Question Rodolfo Rivas-Ruiz, Juan O. Talavera

S54

VIII. Structured Review of an Article Juan O. Talavera, Rodolfo Rivas-Ruiz

S58

IX. From Clinical Judgment to Clinical Trial Juan O. Talavera, Rodolfo Rivas-Ruiz

S64

X. From Clinical Judgment to Cohort Design Juan O. Talavera, Rodolfo Rivas-Ruiz

S70

XI. From Clinical Judgment to Case-control Design Juan O. Talavera, Rodolfo Rivas-Ruiz

S76

XII. From Clinical Judgment to Cross-sectional Survey Juan O. Talavera, Rodolfo Rivas-Ruiz

S80

XIII. Research Design in the Structured Review of an Article Juan O. Talavera, Rodolfo Rivas-Ruiz

S84 Authors

Introduction

Clinical Research

Juan O. Talavera

This series of articles is an attempt to provide to the clinical care physician a tool for better interpreting their day-to-day observa-L tions in order to solve the patients' health problems. This way, he will not depend on others' interpretation and he will also be able to identify unintended or intended misinterpretations that are observed in scientific publications. The series begins with a description of different approaches, out of which two have to be highlighted: the architectural approach, which is based on clinical judgment in order to describe the causality phenomenon and the process studies (diagnosis); and the hierarchical approach, the axis of which is the quality of information and where four basic designs are shown: the clinical trial, the cohort, the casecontrol design and the cross-sectional survey. Additionally, a strategy is referred, which allows for us to understand the reasons for the statistical testing and the size of the sample, followed by the difference between statistical significance and clinical relevance, with the latter determining the usefulness of the maneuver. Then, the systematic search procedure is described, a strategy aimed to find, in a fast and orderly manner, articles able to answer questionings generated in clinical care routine. The supplement concludes with a pair of examples: the first one, which integrates the elements proposed to be essential for a structured review of literature and the second one, which shows the combination of the architectural and the hierarchical models.

Preface

Importance of the Dialogue Between Clinical Practice and Scientific Research

Alberto Lifshitz^a

^aSenior advisor on education, research and health policies, Instituto Mexicano del Seguro Social, Distrito Federal, México. Contact: Alberto Lifshitz Telephone numbers: (55) 5623 2421; 5623 2300, extension 43038 Email: alifshitz@liceaga.facmed.unam.mx

ven though clinical practice is nourished by the results of scientific research and the latter is fed by the needs in clinical practice, the truth is that in recent times these two worlds have grown inconveniently apart. One seems to be the world of science and other the world of clinical practice. Even in the curricular structures of medical training, two clearly defined stages are differentiated: basic sciences and clinical disciplines, to such a degree that they appear as if they were two separate careers. All curricular programs have to make use of integrative activities since they are often seen as separate compartments. Furthermore, in many schools, basic science teachers are not clinicians anymore, but biologists or chemists; hence, they lack the perspective of the physician's professional practice, and many clinical teachers have forgotten, if not disregard or fear, basic sciences. Today, new basic sciences such as epidemiology, statistics, and communication and information technology have been added, and a trend towards geting out of the basic-clinical dichotomy and endeavor into the essential-applied dichotomy is rather perceived (Bandiera G, Boucher A, Neville A, Kuper A, Hodges B. Integration and timing of basic and clinical sciences education. Med Teach. 2013;35(5):381-7. doi: 10.3109/0142159X.2013769674. Epub 2013 Feb 27). Moreover, clinical practice is at risk of becoming an empirical, reflex, stereotyped activity when it drifts away from science, even from the so-called clinical science.

The movement of clinical epidemiology represented a change in the way the archetypal activity of physicians is seen by incorporating methods that are characteristic of science not anymore to the inquiry of basic aspects of medicine, but to clinical practice itself, and not only as a strategy for the generation of knowledge, but to take care of patients more adequately. From this proposal, many methodological advances emerged, several of which were grouped within evidence-based medicine. One of the most important achievements for the care of patients has been precisely the implementation of these methods in the search of better solutions for the diseased. This supplement is a contribution in this sense and not necessarily for the training of investigators but for the training of better physicians that integrate research activities to their routine clinical practice. Ultimately, patient care is an appropriate space for this integration of complementary visions: there is where the research needs to arise and there is where the results arrive as better solutions than the previous ones.

Evidently, traditional training of physicians does not cover sufficiently this ability to identify problems in daily routine that should be addressed using science, or to look for the appearance of solutions for their timely implementation, and even less the ability to judge the validity and reliability of everything that is published and disseminated. Unfortunately, the excess of information is riddled with pseudoscience, whether publicity appearing to be scientific information or well-intentioned results but with methodological flaws. Those who take care of patients should at least be able to tell apart the valuable from the superfluous, the promotional from the scientific, the applicable from the theoretical, the reliable from the questionable, and the valid from the non-valid information. The basic input for medical care is, certainly, information, and therefore, it has to have quality.

But clinical practice is also an appropriate setting for the creation of knowledge. The problem is that the motivation, discipline, curiosity or methodology required to make this potentiality effective are not widespread enough. This supplement is, therefore, a valuable tool to awaken the scientist clinicians carry within and to pour this capacity to the benefit of their patients and the progress of the profession. Much has been debated on whether clinical practice is a science or not. What we are able to state is that it is a space where knowledge generated by science can be put to test, a territory wherein scientific research needs emerge, an activity that follows a similar inquiry methodology to that of science, and a setting where patient-centered research can certainly be developed.

It is true that there are many and very good texts on research methodology and scientific literature critical analysis, but this supplement has the advantage of being aimed at those who are responsible for the care of patients in an institution like the *Instituto Mexicano del Seguro Social*, and it is written by healthcare professionals who have this kind of experience, in addition to their methodological training, which was focused on clinical research as well. The potentiality for finding questions that can be addressed by means of research and pursuing the results of investigations in order to apply them at the appropriate time on everyday patients has been poorly exploited. This *Revista Médica del IMSS* supplement is a tool to move forward along this path.

Editorial

Medical Practice and Clinical Research: Keys to Generate Knowledge and Improve Healthcare

Carla Martínez Castuera-Gómez, Juan O. Talavera

Medical knowledge that integrates clinical research into routine medical practice may directly impact in the quality of care. In the process of medical knowledge generation are four steps: posing a question related with medical practice, analyzing the knowledge published in specialized literature, developing a research protocol and publishing results. Enabling clinical research-based decisions is essential to favor the development of strategies that increase the quality of care.

Key words: clinical research, healthcare quality, periodic publications

Researching, creating and sharing knowledge are amongst the noblest activities that human beings can engage in, since their goal is invariably to improve the condition of life in general. This generousness is more evident in the field of medicine, since research results determine the quality of life that healthy people, as well as those affected by some disease, will have. Therefore, the importance of research in the medical area lies in its inherent social responsibility.

In view of the latter, this reflection seeks to contribute to the idea that it is possible to assume such responsibility when healthcare staff maintains a symbiotic relationship between medical practice, clinical research activities, and the publication of medical knowledge.

From Clinical Practice to the Generation of Knowledge

The process of medical knowledge generation may improve medical care quality when it begins in medical practice, it is enriched by clinical research and it ends up with its publication.

Medical practice can be understood as the strategy routinely followed by the physician when choosing the best care alternatives —within her means of knowledge and resources— in order to treat a specific health condition. When the physician faces situations that she is not able to solve in the usual way, she reaches the point to start generating medical knowledge.

The first step in this process is taken when the doctor poses a question trying to solve a problem arising from his professional practice, whether trying to establish a diagnosis, estimating the prognosis or deciding the cause of the problem or a better treatment. Questioning is a skill that physicians develop almost naturally. Routine activities like physical examination, history taking or review, prescription of a different drug upon complications or persistence of diseases, among others, involve a questioning. This questioning is followed by the search for causes, comparison of cases, and identification of irregular conditions, in order to make decisions on the best treatment for a certain health condition. Questioning, answering and deciding are inherent tasks to the medical profession, such as the creation of knowledge. When the physician gets involved in academic and research activities in parallel to his professional practice, questioning and assertive decision-making skills are refined and sharpened. In consequence, physicians who do not engage in research are wasting the opportunity to develop their professional skills and are neglecting their social responsibility by not using their knowledge and capabilities, in order to improve people's quality of life. Moreover, the development of clinical research must be included as a requirement in the design of healthcare systems and, therefore, administrative and medical tasks must exist in order to facilitate its execution.

The next step in the generation of medical knowledge is the search for answers consultating and critically analyzing specialized literature. The importance of this step is that it reduces the risk of investing time and human, financial and physical resources searching for answers to questions already posed, or even worse, ending up with inconclusive answers or answers that have already been proposed. Furthermore, comprehensive and critical review of literature is crucial because it ensures for the manuscript to be original and innovative, with appropriate scientific support and high feasibility estimation. When these factors are contained in a manuscript, it is more likely that it has accurately solved the posed question and that it will be able to turn into publication material, due to the relevance of the generated knowledge.

This step appears to pose two challenges: access to the sources of information and selective search. Actually, the challenge is only one: knowing how to search. Internet and PubMed are powerful sources of readily accessible information to all physicians, but if the use of search parameters is not known, they become an endless reservoir of low quality information that discourages research. For this challenge, a simple solution is proposed: teaching selective search strategies and constantly putting them into practice. This proposal is an aspect in which medical and administrative personnel can influence in order to maintain the *medical practice-clinical research-publication* symbiosis.

The third step in the medical knowledge generation process is to design and execute the clinical research protocol. The development, the contents, the characteristics and the execution of a protocol are widely discussed topics beyond the scope of this reflection, whose central interest is to state that medical knowledge is generated when clinical research is able to propose an answer to a question arising from medical practice. Nevertheless, it is important to emphasize that clinical research and the development of the protocol should follow quality control strategies in order to safeguard both methodological strictness and participating patients. This is achieved with the inclusion and observance of minimum ethical principles. Involvement of ethics committees, international registration of clinical trials, peer reviews and editorial boards counseling are some of the mechanisms to supervise adherence to ethical principles that warrant the development of quality research.

The execution of the research protocol generates an answer to the question. Even though the answer may be different from that what was inferred or expected, there is certainty that it was obtained collecting and testing evidence. Regardless of the answer, the fourth step of the process begins, and the time to select a journal to publish the obtained information.

Currently, there is a trend to select a journal considering mainly its impact factor: "today, too many of our postdocs believe that getting a paper into a prestigious journal is more important to their career than doing the science itself".¹

However, this decision should be based on the audience to whom the information is directed, the accessibility readiness offered by the journal to medical audiences, publishing requirements, and, ultimately, the impact factor. This order of selection priorities is ideal if the main objective of publishing is to disseminate clinical research results and encourage physicians to integrate them in their daily practice, in order to improve their practice and care.

Moreover, this order of priorities relieves the pressure imposed when trying to get published in a journal with impact factor and supresses frustration when that is not achieved. Although academic systems rely on parameters such as the impact factor for the assessment of scientific productivity, in the local setting, there is the possibility of creating assessment mechanisms and incentives that promote the publication of medical knowledge in prestigious journals that are easily accessible and widely available to the medical community, regardless of the impact factor. In our country, and especially in our Institution, the Revista Médica del Instituto Mexicano del Seguro Social is a unique and privileged space that has to be considered in order to encourage publication of medical knowledge.

According to an editorial published in "Proceedings of the National Academy of Sciences," numerous postdoctoral students state that they would choose publishing their academic work in their favorite journals, those in which they find writings they enjoy reading, if they were not assessed based on the impact factor.¹ Moreover, if —as it has been argued— published medical knowledge allows for the best practices to be shared and promoted, then, the selection of the journal to publish should not be defined solely by the impact factor.²

Taking this into account, it would seem convenient to promote publication of knowledge resulting from clinical practice research, in readily accessible journals, since this characteristic will favor its application in the medical area. For example, publishing in local journals increases the likelihood that the reader knows the author and vice versa. This could be an important stimulus to encourage more physicians, who perceive themselves on the same level as the authors, to feel attracted to create and share their knowledge through the process to generate knowledge. Furthermore, physicians who read knowledge published by colleagues may be more likely to integrate it into their own practice if the author is a person they respect, partly because the readers have the possibility of discussing with the author and because they are certain that the author knows the conditions of their medical service or, at least, their local or national circumstances. Such knowledge is perceived with authority and not as an imported recipe that cannot be applied to local circumstances. Selecting this kind of journals reduces the temptation to distort the results or the information in order to get published, contrarily when the publication is sought in a high-impact factor journal.¹

Finally, if we remember that researching is an act of social responsibility, the selection of the journal for publishing should not be made based on prestige but on the possibility of sharing knowledge. Therefore, promoting the improvement of medical practice is directly related to the promotion of publishing medical knowledge based on clinical research. The more integrated the medical activity into clinical research is, with the resulting publication of the generated medical knowledge, the greater the chances of influencing on medical care improvement will be, thus closing the virtuous circle of knowledge generation.

So far, we have tried to support the argument that the *medical practice-clinical research-publication* relationship has an impact on the quality of medical care. Like other authors, we believe that clinical research by itself has three positive effects:³⁻⁶

- 1. Patients who participate in a clinical research project receive better quality of care.
- 2. The physician's motivation and satisfaction at work increase.
- 3. Health systems benefit from the efficacy and efficiency shown by both physicians in their practice and patients in their treatment.

However, it is publication and dissemination of clinical research-derived knowledge that assures these benefits will be extended and reproduced by means of the medical practice-clinical research-publication relationship. The described pathway is ideal for maintaining this symbiosis and influencing on the improvement of healthcare. However, unfortunately, this is not the path that is always followed. It is possible, and more often than desirable, to find unoriginal or poorly substantiated and inconclusive clinical research publications, with very low quality control and, sometimes, disregarding relevant ethical principles. The consequences have not been negligible: eroded credibility of some journals; lack of interest in publishing knowledge, generated by clinical research and in conducting research; non-updating of physicians and a tendency to reduce their practice effectiveness; as well as low or non-existent creation of knowledge applicable to the patient's ailments.

Conversely, when the process to generate knowledge originated in clinical practice and clinical research is followed in an orderly manner, a virtuous environment is generated, and it stimulates the medical practice-clinical research-publishing symbiosis. A physician involved in medical care that performs clinical research and crystallizes the process with the publication in journals that are accessible to her colleagues becomes an authority and a role model. Anyone who solves the needs of medical practice through clinical research develops good care habits and makes it easy for this attitude to be reproduced among the healthcare personnel she works with. In summary: an immediate improvement in the care of patients is estimated.

Conclusions

The impossibility of a physician to address part of his social responsibility by not getting involved on academic and research activities could be considered overwhelming. However, there is no reason for such an interpretation when it is understood that the responsibility of this professional is the generation of medical knowledge and its use for the improvement of patient care. It is the responsibility of administrative personnel and healthcare systems designers to promote favorable environments to engage physicians in clinical research and publish their results. With this in mind, there are four aspects that are worth thinking of:

- Not all medical practice should become research material, but all research must turn into decision-making material in clinical practice.
- Training in information search techniques and adequate analysis of literature are simple and inexpensive alternatives that will help doctors to refine their questioning and decision-making skills in favor of better patient care. Evidently, this requires basic training that allows assessment of quality information and preventing its acceptance without critical reflection.
- Support to the publication and dissemination in local medical journals can be a mechanism for

stimulating the medical practice-clinical researchpublication symbiosis.

The creation of a favorable environment for physicians to conduct clinical research is an opportunity for healthcare systems administrators and decision-makers to facilitate the generation of medical knowledge that impacts on the quality of care.

Consequently, stimulating academic and research activities in discussion sessions between physicians and residents is suggested, since literature search tools and critical analysis are thereof transmitted, in order to solve questions arising from medical practice. Since many healthcare centers are also teaching centers, this task would only imply time organization, setting up a classroom or a meeting room with computing equipment, access to Internet and interactive communication systems, which allow for real-time medical literature

References

- Marder E, Kettenmann H, Grillner S. Impacting our young. Proc Natl Acad Sci USA. 2010;107(50):21233. doi: 10.1073/pnas.1016516107
- McIntyre E, Eckermann SL, Keane M, et al. Publishing in peer review journals. Criteria for success. Aust Fam Physician. 2007;36(7):561-2.
- Jowett SM, Macleod J, Wilson S, et al. Research in primary care: extent of involvement and perceived determinants among practitioners from one English region. Br J Gen Pract. 2000;50:387-9. Free text http://www.ncbi.nlm.nih.gov/pmc/articles/ PMC1313704/pdf/10897537.pdf

searches and promote communication between physicians from different healthcare centers.

Finally, the promotion and support to local journals can be achieved if physicians ask for those publishing spaces to be opened, and, at the administrative level, if their production and distribution is facilitated.

Knowledge that is generated but not shared is useless knowledge because there is no possibility of applying, reproducing and improving it. Publication is the most powerful mechanism to share knowledge since, on one hand, it forces its generators to structure and order it in an accessible way and, on the other hand, because publishing crystallizes knowledge for its recall and consultation. The publication of medical knowledge, supported by medical practice and clinical research, is useful knowledge that will allow improvement of medical care quality and the fulfillment of the social responsibility inherent to medicine.

- Starfield BL, Macinko J. Contribution of primary care to health systems and health. Milbank Quarterly. 2005;83:457-502. doi: 10.1111/j. 1468-0009.2005.00409.x.
- Sullivan F, Butler C, Cupples M, et al. Primary care research networks in the United Kingdom. BMJ. 2007;334:1093-4. doi: 10.1136/bmj.39190.648785. 80.
- Soler-González J, Ruiz C, Serna C, et al. The profile of general practitioners (GPs) who publish in selected family practice journals. BMC Res Notes. 2011 May 26;4:164. doi: 10.1186/1756-0500-4-164. Free text http://www.biomedcentral.com/ 1756-0500/4/164

I. Research Designs

Juan O. Talavera

Clinical research takes care primarily of the study of groups of diseased individuals in order to establish a diagnosis, estimate a prognosis and start a treatment. With this purpose, it uses the scientific method from different points of view: architectural, which is divided in cause-effect and process studies; methodological, which includes clinical trials, cohort-case-control-studies and surveys; and by objectives, which comprised diagnostic test, prognosis and treatment studies, as well as risk factors or etiologic agent studies. These designs are considered to be primary, i.e., they use information obtained directly from the subject under study; however, there are other that use information from primary studies, which are known as secondary or integration designs.

Key words research research projects clinical trial

This article was originally published in Rev Med Inst Med Seguro Soc 2011; 49 (1): 53-58 and it has been reviewed for this issue.

Introduction

Clinical research, known as clinical epidemiology -a term that under the current sense was quoted by Alvan R. Feinstein (previously, it had been used by John R. Paul, to refer to what we currently know as social epidemiology and community-based medicine)- takes care of the study of groups of individuals in order to obtain decision-making evidence in patient care; i.e., it deals with the study of the structure and function of research performed in diseased subjects. However, sometimes it overlaps with classical epidemiology and studies the subject before the development of the disease. On the other hand, knowledge acquired in clinical epidemiology applies to the patient as an individual entity, whereas in most cases, knowledge obtained in classical epidemiology applies to a group of subjects.

The research method in clinical epidemiology is unique and it is consistent with the scientific method. However, for educational purposes, classifications have been made from different points of view, out of which three are the most common.

The first one, called *architectural*, is based on the most accurate description of the real event and includes cause-effect and process studies. The second one, known as *methodological*, is characterized for hierarchically categorizing the quality of the information obtained from the groups under study; it comprises clinical trials, cohort –case-control– studies and surveys. The third one uses the purpose it entails in everyday clinical practice and is known as *approach by objectives*; it is divided in diagnostic, prognostic, treatment and risk factors or causative agent (causality) studies.

Studies not considering a maneuver imposed by the investigator and that, therefore, are not experiments but observations, follow the principles of the scientific method and replace the experimental maneuver with a naturally-occurring or an imposed maneuver with purposes unrelated with the research.

Architectural Approach

When we talk about cause-effect studies, we refer to the change suffered in the subject's baseline state when receiving a maneuver, for example: when estimating, in a previously healthy patient (baseline state) who suffers a head injury (observational maneuver), the probability of dying or being left with sequels (outcome); or when assessing, in a patient with headache (baseline state), if a prescribed analgesic (maneuver) reduced the pain (outcome). This means that cause-effect studies not only include the search for an etiologic agent or risk factor, but also for prognostic factors and even therapeutic actions. On the other hand, process studies assess the quality of procedures, either by comparing the procedure to be analyzed with a standard or with another execution of it; for example: to estimate the sensitivity and specificity of neck ultrasound (procedure under study) it is compared in patients with carotid obstruction (against carotid arteriography). In cases without gold standard, the study is compared with another execution of the same study assessing the same lesion by two radiologists in order to evaluate the coincidence beyond that expected by chance (Figures 1 and 2).

Methodological Approach

Based on the quality of the obtained information, the methodological approach attempts to hierarchically categorize the different designs in a way that it allows for deciding which study on the same matter is more reliable by being less likely to have biases present and, therefore, in which the decisions related with patients should be based.

It is important to consider that designs in lower hierarchical levels carried out adequately can outperform others with higher levels but poorly structured; furthermore, studies at lower hierarchical levels may be sufficient to answer a research question; moreover, not rarely, these are the only ones that can be performed.

In the description of the designs it is necessary taking into account four basic characteristics and the measurement of the outcome occurrence.

Basic Characteristics

 Imposition or not of a maneuver with research purposes. A study is considered experimental if the maneuver was imposed by the investigator, and observational when such maneuver is natural (e.g., the presence of some disease) or imposed with purposes unrelated with the research (smoking, alcoholism, etc.).

- 2. Follow-up of the patient over time or not. A study is considered to be longitudinal when the patient is assessed in some of his/her characteristics of interest over time (more than once); in most cases, the change from baseline state to that of the result or outcome is referred, for example: follow-up of a group of physicians with no history of ischemic heart disease (baseline state) for five years and measurement of the onset of coronary heart disease during this period (outcome). The research is cross-sectional when the patient is assessed in a stationary manner (only on one occasion), for example: measurement of hypertension in a group of diabetic patients trying to find an association of lack of metabolic control with hypertension. While longitudinal studies allow for the assessment of different factors as sources of change from baseline to the subsequent state with certainty of the temporality of exposure to them, in transversal studies, often there is no certainty of a temporal relationship, even when associations are established between variables where which is the maneuver and which the outcome is artificially assumed.
- 3. Directionality in the collection of information. A study is prolective when the collection of information relates to the baseline state, as well as to the maneuver and the outcome. It is performed in real time with investigational purposes, i.e., simultaneously with the exposure to the maneuver and the occurrence of the outcome. It is retrolective when the information is obtained once the exposure to the maneuver and the outcome have occurred. It is possible for a study to be retro-prolective if at the moment at which the information is obtained the maneuver has already occurred, but not yet the



Figure 1 A cause-effect study seeks to establish the association between the maneuver and the change in the subject's baseline state, which generates a result. Three components must be considered: the subject's baseline state, the principal maneuver and the outcome or result; according to the question, the comparative maneuver may be necessary or not.



Figure 2 Process studies try to assess the reliability of the procedure, for which input information (substrate) is required, as well as the execution of a procedure to be compared with the gold standard or with other execution of the procedure, which yields as a result output information.

result, and therefore, its measurement is performed at the moment it occurs (Figure 3).

4. Search or not for an association between two variables. A study is descriptive when the purpose is to show the range of characteristics of the group under study. Frequently, the results of descriptive studies are used for comparative purposes; for example: when the prevalence of certain disease in a given population is compared with the prevalence of the same disease in a previously analyzed population. Conversely, a study is comparative when the association between the maneuver and the outcome or between a standard and the quality of a product or procedure (when it is a diagnostic study) is searched. An example of a comparative study is the search for association between obesity (natural maneuver) and insulin resistance (outcome), or when comparing an acute cholecystitis ultrasonographic diagnosis (procedure) with surgical findings (gold standard).

Measurement of Outcome Occurrence

Measurement of the outcome frequency can be performed in two ways according to the methodological design:

- 1. *Incidence (cumulative incidence)* refers to the number of new cases occurring in a certain period and population; it is characteristic of studies with follow-up, i.e., of cohorts (either observational or experimental). It can have different names: when mortality is studied and not the occurrence of a disease, it is known as mortality rate.
- 2. *Prevalence* or number of existing cases at a given moment in a given population; it is typical of cross-sectional studies, except for case-control studies.

The *case-control ratio* is not a way to measure the occurrence of the outcome but rather an artificially-created simple case-control relationship.

Basic Designs

Hierarchical order, assigned by the quality of the obtained information, places the *clinical trial* at first place, since it allows for information to be obtained directly and with control over the maneuver and, consequently, with the least amount of errors. It is followed by the cohort study, then the case-control study and, finally, the survey.

The *clinical trial* is characterized for being a prolective and longitudinal study, where the application of the maneuver (experimental) to which the change in the baseline state wants to be attributed to (comparative) is planned; a clinical trial is experimental when it has a comparative group, with randomization to the maneuver and blinded assessment of the outcome. However, sometimes there is no comparative group available, and baseline state is the characteristic that has to be compared with the result (before-and-after study), or randomization of the maneuver or a blinded assessment of it are impossible to perform, which defines the clinical trial as being quasi-experimental. The clinical trial can be defined as an experimental cohort, since it has all the characteristics of a cohort with allocation of the maneuver. Being a longitudinal study, it allows for the incidence to be estimated as a measure of occurrence of the disease.

The *cohort* is the ideal design among observational studies. It is characterized for having a group of subjects selected according to common characteristics at a given moment and that are followed over time in some of their characteristics (*longitudinal*), where the collection of information (*prolective*, *retrolective* or *retro-prolective*) may or may not coincide with the occurrence of the maneuver or the result, and in which the association between the maneuver and the result is always sought (*comparative*). Even when the design may be retrolective, a situation in which it is termed historical cohort, the direction goes from the cause (maneuver) to the effect (result). For example, a prognostic study can be conducted to find out which



Directionality in the collection of information

Figure 3 When the capture of information starts at baseline state before the maneuver and the result, the study is considered to be prolective (a); when the capture is carried out once the maneuver and the result have occurred, it is considered to be retrolective (b); and when the capture is performed once the maneuver has occurred but before the result, it is a retro-prolective study (c).

stroke patients will die within the first few days after the event, for which the information on the charts of all patients admitted to the hospital during the year preceding the study is reviewed; since the maneuver (characteristics present within the first hours of the stroke, known as prognostic indicators) and the result or outcome have already occurred (death within the first seven days of the event), it is a retrolective study; however, the analysis and capture of data should be done with all patients, starting with clinical manifestations present at admission and then measuring the outcome. Unlike case-control studies, which may cover these same characteristics, the cohort provides information of all the patients that suffered the stroke during the year and, therefore, the incidence of the outcome is available, whereas in case-control studies, the whole population is not available but rather an artificial rate of case-controls is used, as outlined below.

Conversely to the aforementioned designs, the case-control design is characterized for going from the effect to the cause. It starts with a group of subjects with the outcome of interest (*result*), which corresponds to the cases, and a witness group that did not suffer the outcome (*controls*) is selected; afterwards, the association between the maneuver and the outcome (*comparative*) is searched. Therefore, it is a *retrolec*-

tive and *observational* study. There is controversy regarding the follow-up of variables or not, with some authors considering this to be a cross-sectional study, since all the information is obtained at one time-point, whereas for others, it is *longitudinal* because a recapitulation of the maneuver temporality is feasible until the moment of the outcome. In this design, there is no outcome occurrence measurement; there is simply an artificially-created case control relation.

The survey is the simplest among observational designs but also the most limited in its assertions; it is carried out on a representative sample of the study population and the most common objective is outlining the population characteristics (descriptive); however, it can also be used to establish an association between two or more variables (comparative). Frequently, it is impossible to determine whether the maneuver precedes the outcome, since the gathering of information happens after both the maneuver and the outcome have occurred (retrolective) and at one single time (transversal). Unlike case-control studies, there is no predetermined ratio of the number of cases and controls; in fact, there is no selection of the population based on the outcome, but instead, once the population is selected (whatever the criteria are), exposure to the maneuver, which in this case is observational, and the

Table I Designs according to the methodological approach									
Design	EXP/OBS	LONG/TRANS	PROL/RETROL	COMP/DESC	MEASURE				
Clinical trial	Experimental	Observational	Prolective	Comparative	Incidence				
Cohort	Observational	Longitudinal	Prol/Retrol/RP	Comparative	Incidence				
Case-control	Observational	Long/Trans	Retrolective	Comparative	Prop. C-C				
Survey	Observational	Transversal	Retrolective	C/D	Prevalence				

The methodological approach considers four features: 1. Imposition or not of the maneuver for investigational purposes: experimental (EXP) or observational (OBS) study, respectively. 2. Patient follow-up (LONG) or not (TRANS) over time. 3. Directionality in the collection of information: prolective (PROL), retrolective (RETROL) and retro-prolective (RP). 4. Search or not of association between two or more variables: comparative (C) and descriptive (D), respectively. Measurement of outcome occurrence (MEASURE), either through incidence, prevalence, or simply the case-control ratio (C-C ratio)

outcome are measured. Therefore, the obtained result is the prevalence of the outcome.

Table I summarizes the distinctive characteristics of each design. It is worth mentioning that there are combinations of these designs and sometimes it is difficult defining them.

Approach by Objectives

The approach based on clinical practice is the one that we are more used to; furthermore, in it, it is possible to distinguish the largest difference between clinical epidemiology and classical epidemiology. In clinical epidemiology, which studies groups of patients, the primary objective is to solve an already existing problem in a group of people for which a diagnosis must be established (diagnostic study), a prognosis has to be estimated (prognostic studies) and a therapeutic maneuver has to be initiated (experimental or quasi-experimental clinical trial). However, as we mentioned earlier, it is common for clinical epidemiology to overlap with classical epide-

Bibliography

- 1. Cañedo DL. Investigación clínica. México: Interamericana; 1987.
- Feinstein AR. Clinical epidemiology. The architecture of clinical research. Philadelphia: WB Sanders; 1985.
- Feinstein AR. Directionality and scientific inference. J Clin Epidemiol. 1989;42:829-33.

miology and to address risk factors problems, such as cardiovascular disease (risk factors or etiologic agent study, the latter when the agent is single).

Complementary Studies

So far, we have mentioned only studies that use primary information; however, there is a group known as "integration studies," characterized by the pooling of data obtained in primary studies. These comprise four designs: review studies (meta-analyses and systematic reviews), clinical practice guidelines, decision analyses and economic analyses.

Acknowledgements

We are grateful to Doctors Niels H. Wacher-Rodarte, Susana Castañon Robles, Rodolfo Rivas-Ruiz and Jorge Salmeron-Castro for their suggestions, which allowed for this manuscript to be substantially improved.

- Feinstein AR. Scientific standards in epidemiologic studies of the menace of daily life. Science. 1988;242:1257-63.
- Hernández-Ávila M, Garrido-Latorre F, López-Moreno S. Diseño de estudios epidemiológicos. Salud Publica Mex. 2000;42:144-54.
- Hughes M, Williams P. Challenges in using observational studies to evaluate adverse effects of treatment. NEJM. 2007;356:1705-7.

- Hulley S, Cummings S. Designing clinical research. Philadelphia, USA: Lippincott Williams & Wilkins; 1988.
- Kelsey J, Whitemoore A, Evans A, Thompson W. Methods in observational epidemiology. Second edition. New York, USA: Oxford University Press; 1996.
- Meinert C. Clinical trials: design, conduct and analysis. New York, USA: Oxford University Press; 1986.
- Moreno A, Valle C, Romero G. Epidemiología clínica. Segunda edición. México: Interamericana-McGraw-Hill; 1994.
- Petitti D. Meta-analysis, decision analysis and costeffectiveness analysis: methods for quantitative synthesis in medicine. Second edition. New York, USA: Oxford University Press; 2000.
- Schlesselman J. Case control studies: design, conduct, analysis. New York, USA: Oxford University Press; 1982.
- Wacher N, Lifshitz A. Qué es la epidemiologia clínica y para qué le sirve al clínico. Rev Med IMSS. 1989;27:171-4.
- Walker AM. Observation and inference. An introduction to the methods of epidemiology. Chestnut Hill, MA: Epidemiology Resources Inc.; 1991.
- Weiss NS. Scientific standards in epidemiologic studies. Epidemiology. 1990;1:85-6.

Recommended readings of examples

Case-control

 Cruz-Anguiano V, Talavera J, Vázquez L, Antonio A, Castellanos A, Lezana M, et al. The importance of quality of care in perinatal mortality: a populationbased case-control study in Chiapas, Mexico. Arch Med Res. 2004;35:554-62.

Cohort

 Brea-Andrés E, Aburto-Gudiño E, Vázquez-Estupinán F, Nellen-Humel H, Talavera-Piña JO, Wacher-Rodarte N, et al. Incidencia de delírium y morbilidad asociada en medicina interna. Acta Psiquiátrica y Psicológica de América Latina 2000;46:359-62.

Diagnosis

Talavera J, Wacher N, Laredo F, López A, Martínez V, González J, et al. A rating system for prompt clinical diagnosis of ischemic stroke. Arch Med Res. 2000;31: 576-84.

Survey

 Gómez-Díaz R, Martínez-Hernández A, Aguilar-Salinas C, Violante R, Alarcón A, et al. Percentile distribution of the waist circumference among Mexican pre-adolescents of a primary school in México City. Diabetes Obes Metab. 2005;7:716-21.

Randomized clinical trial

 González-Ortiz M, Guerrero-Romero JF, Violante-Ortiz R, Wacher-Rodarte N, Martínez-Abundis E, Aguilar-Salinas C, et al. Efficacy of glimepiride/ metformin combination versus glibenclamide/metformin in patients with uncontrolled type 2 diabetes mellitus. J Diabetes Complications. 2009;23: 376-9.

Process studies

- Gómez R, Aguilar-Salinas CA, Morán-Villota S, Barradas-González R. Herrera-Márquez R, Cruz M, et al. Lack of agreement between the revised criteria of impaired fasting glucose and impaired glucose tolerance in children with excess body weight. Diabetes Care. 2004;27:2229-33.
- Pérez-Cuevas R, Reyes-Morales H, Flores-Hernández S, Wacher-Rodarte N. Efecto de una guía de práctica clínica para el manejo de la diabetes tipo 2. Rev Med Inst Mex Seguro Soc. 2007;45(4):353-60.

Prognosis

 Cruz M, Maldonado-Bernal, C, Mondragón-González R, Sánchez-Barrera, Wacher N, Carvajal-Sandoval, et al. Glycine treatment decreases proinflammatory cytokines and increases interfeon-g in patients with type 2 diabetes. J Endrocrinol Invest. 2008;31:694-9.

Risk

 Cruz M, García-Macedo I, García-Valerio Y, Gutiérrez M, Medina-Navarro R, Durán G, et al. Low adiponectin levels predict type 2 diabetes in Mexican children. Diabetes Care. 2004:27:1451-3.

Treatment

25. Nellen H, Flores G, Wacher N. Treatment of human immunodeficiency virus enteropathie with a gluten-free diet. Arch Intern Med. 2000;160:244.

II. Process Studies (Diagnostic Test)

Juan O. Talavera, Niels H. Wacher-Rodarte, Rodolfo Rivas-Ruiz

The purpose of a diagnostic test is to establish the presence of health or disease, it can even graduate the degree of illness. Diagnostic tests are usually assessed mathematically. Thus, sensitivity and specificity are estimated once the existence or not of disease is known; in clinical practice, the course of action is often reverse: from positivity or negativity of a test for the presence or not of the disease and, therefore, positive and negative predictive values are used. Mathematical strategies allow for an observation to be quantified, but clinical judgement is required in order to establish the quality of that observation; in consequence, some characteristics have to be considered: a) selection under the same criteria for cases and witnesses; b) inclusion of the entire spectrum of severity of the disease (trying that all the strata include an important number of subjects); c) the interpretation of the gold standard and the test under study must be blinded and done by experts; d) the interpretation of the results must show the applicability of the test in everyday practice; e) reproducibility of the test must be proven. It is important not to forget that, usually, only a patient is seen at a time; therefore, full knowledge of the diagnostic test performance is essential, as well as considering the clinical aspects for its correct application.

Key words research research projects diagnostic techniques and procedures

This article was originally published in Rev Med Inst Med Seguro Soc 2011; 49 (2): 163-170 and it has been reviewed for this issue.

Introduction

Part 1 of this series [Rev Med Inst Seguro Soc 2011; 49(1):53-58] mentioned the different approaches for addressing clinical problems: *architectural approach*, based on the natural phenomenon; *methodological approach*, based on the hierarchy of the information; *clinical approach*, based on the aims of medical practice. Methodological approach key features were analyzed in detail, and integration studies were also mentioned.

However, in clinical practice, questions use to be related with the need to establish a diagnostic or ascribe causality either through a prognostic study, a treatment, or by trying to identify whatever provoqued a certain disorder or disease. This is where the architectural approach fits together with the objective-based approach.

Among the process studies, according to the architectural approach there is the diagnostic testing (objective-based approach). Additionally, causality studies include the prognostic, treatment and risk factors or causative agent studies (objective-based approach). In this article, we describe the most commonly used tools in diagnostic testing.

In clinical practice, a diagnostic test aims to identify the health or disease status of the subject under study. Frequently, in the presence of a disease, it allows for the severity of the condition to be established; for example: in a patient with sudden neurological deficit, tomography allows for the diagnosis to be defined (ischemic stroke), whereas if the diagnosis is already available, tomography allows for the extent of the lesion to be known.

The use of mathematics during the diagnostic process has the purpose of estimating the degree of efficacy and certainty of the tests in clinical practice. Below, the main features of every diagnostic test, using both clinical data and laboratory and imaging findings, are described.

Characteristics of a Diagnostic Test

The way to assess the efficacy of a diagnostic test depends on the type of data (variable) to be used. Therefore, it is important to identify the type of variable. Basic variables are those that we know as *qualitative of the nominal* or *dichotomic type*, and they refer to those for which we only notice its presence or for which only two options exist (e.g., nationality, presence or not of disease, male or female). *Ordinal qualitative* variables are those in which it can be identified only the place occupied in the group by the evaluated characteristics, but we do not know the size of the difference between each other (e.g., the degree of severity of a disease — mild, moderate or serious—, or the intensity of a clinical piece of information identified by a cross mark, where, even when + is acknowledged to be lower than ++ and, consequently, lower than +++, ++ can not be stated as being double to +). And, finally, *quantitative variables* are those in which the distance between two levels of intensity is known; and in this variables the distance between two units is always equidistant. They are known as *discrete* or *discontinuous* when they can not be fractionated (e.g., how many children has a family [0, 1, 2, 3]), and *continuous* when fractions can be identified between one value and another (e.g., 52.0 kg, 52.2 kg or 52.250 kg weight).

Sensitivity and specificity are distinctive characteristics of every diagnostic test and indicate their efficacy. *Sensitivity* refers to the proportion of diseased individuals with a positive test. *Specificity* refers to the proportion of non-diseased individuals with a negative test.

The calculation of sensitivity and specificity uses *nominal* or *dichotomic* data and it is based on the use of a 2×2 table, in which the tested data is contrasted against the final diagnosis obtained by means of an ideal parameter named *gold standard*, which represents the test with the highest reliability for demonstrating a disease, e.g., histopathological results (testicular seminoma), surgical findings (cholecystitis), imaging studies interpretation (stroke by tomography or magnetic resonance imaging), interventional imaging studies (type of congenital heart disease by cardiac catheterization) or laboratory findings (renal failure by creatinine clearance).

Figure 1 shows the calculation of sensitivity and specificity of neck stiffness for the diagnosis of subarachnoid hemorrhage in patients with sudden onset neurological deficit, likely of vascular cause. A sensitivity of 59 % with a specificity of 94 % is observed, which means that 59 % of the patients with subarachnoid hemorrhage may show neck stiffness and among those without subarachnoid hemorrhage, 94 % do not have neck stiffness.

Sensitivity and specificity calculations are directed from the presence or absence of a particular disease, towards the probability of experiencing or not certain data. However, in clinical practice, the approach is often in the reverse direction: it goes from a positive or negative test result to the likelyhood of having or not a specific disease. This type of orientation corresponds to what we know as predictive values. The *positive predictive* value represents the probability that a patient with a certain positive test (sign, symptom, laboratory or imaging result or some index) has of suffering a particular disease; the *negative predictive value* is the probability that a patient, with a certain negative test, has of being free from a particular disease.



Figure 1 shows a positive predictive value of 57 % and a negative predictive value of 95 %; this means that among the patients with clinical symptoms of stroke, a subject with neck stiffness has a 57 % probability of suffering from subarachnoid hemorrhage, whereas a patient without neck stiffness has a 95 % probability of not having subarachnoid hemorrhage.

While sensitivity and specificity values are considered to be constant, which is not true as we will explain later, predictive values are affected by disease prevalence. For example, in Figure 2, where the disease prevalence increased only from 11 to 56 %, maintaining the proportion of diseased subjects with positive and negative tests, sensitivity and specificity are shown to be preserved, whereas predictive values change: the positive predictive value is 93 % and the negative predictive value is 65 %. Thus, a prevalence increase causes an increase in the positive predictive value, with a decrease in the negative predictive value (a positive test in a population with high prevalence of the disease practicaly establishes the diagnosis; a negative test, however, does not rule it out); conversely, a decrease in prevalence produces an increase in the negative predictive value and a decrease in the positive predictive value (a negative test in a population with low prevalence of the disease almost rules the disease out).

If prevalence of the disease in the population from which predictive values of the diagnostic test were obtained is different from the prevalence of the disease in our population, these predictive values cannot be used. However, Bayes' theorem allows for predictive values to be estimated by using the sensitivity and specificity of the test, as well as the prevalence of the entity under study in our population. Table I shows how the increase in prevalence from 11 to 56 % produces a 57 to 94 % increase in the positive predictive value. This example shows clearly how a positive test in a population with low prevalence (11 %) has an approximate probability of 50 % for diagnosing the disease, whereas with a high prevalence (56 %), it practically establishes the diagnosis.

Another practical strategy for estimating the probability of the disease in case of a positive test, but at different prevalence values, is the use of Fagan's nomogram and the likelihood ratio (LR). The positive LR (PLR) is obtained from the ratio sensitivity/1-specificity. In turn, the negative LR (NLR) is obtained from the ratio 1-sensitivity/specificity. Fagan's nomogram is divided in three parts. In the first column appears the pre-test possibility (prevalence). In the middle, there are the values of the LR and in the last column, the post-test probability. The post-test probability for a PLR refers to the probability of obtaining a positive result when the test is positive and it corresponds to the PPV; the post-test probability for an NLR refers to the probability of obtaining a positive result when the test is negative, which is equivalent to 1-NPV. Examples for a prevalence of 11 and 56 % are shown in Figure 3.

It was mentioned previously that the sensitivity and specificity of a test are not dependent on



the prevalence of the disease; however, the values vary according to the predominant disease severity degree in the group under study. For example, diagnosing lung cancer at an advanced stage with a chest x-ray is simple and it will rarely go unnoticed, i.e., false negatives will rarely exist and sensitivity will be high; however, it will be hardly detected if we try to diagnose it in asymptomatic individuals, at an early stage, which will provoke a high percentage of false negatives and low sensitivity. Therefore, considering that the sensitivity obtained from a test in a population is applicable to other population implies that the distribution of disease severity is the same in both samples, since if in the first one the proportion of subjects in advanced stages is predominant, sensitivity will be high, and if in the second prevails an early stage, sensitivity will be low. Having the same inclusion criteria between different studies of different populations does not guarantee that the distribution of subjects will preserve a similar proportion of subjects at every stage of the disease and, consequently, sensitivity may be different.

Use of Ordinal and Quantitative Data

Unlike nominal data, when the test under study corresponds to ordinal or quantitative data (with more than one cut-off point), a ROC (receiver operator characteristic) curve has to be plotted, which enables to determine in which of the cut-off points the highest diagnostic certainty is obtained.

Figure 4 shows the different value ranges of creatine phosphokinase in cerebrospinal fluid expressed in μ U/mL, with their respective frequencies, and the calculation of sensitivity and specificity is outlined according to the different cut-off points by elaborating 2×2 tables. In these tables, intervals are constructed with the different values of the test under study and tabulated in two columns; the first shows the frequencies of subjects with the disease in each of the intervals and the second shows the frequency of subjects without the disease within the same intervals. The most altered values appear above (first intervals) and the less altered below. The cumulative percentaje is calculated upwards and downwards of each cut-off point, in both columns. In the column of diseased subjects, sensitivity is estimated from the cut-off point upwards, and in the column of controls, the percentage of false positives (1-specificity).

The results are plotted with the sensitivity values and the percentage of false positives: sensitivity values on the ordinate axis (Y), and the ratio of false positives (1-specificity) on the abscissa axis (X); a specificity value of 90 % corresponds to 10 % of false positives (Figure 5). The best cut-off point corresponds within the ROC curve to the closest point to the left superior angle of the curve, or to the point within the table that contains the lowest b + c value (values that belong to the sum of false positives and false negatives) or the highest value for a + d (values that belong to the sum of true positives and true negatives). In this case, the cut-off point is $\geq 16 \,\mu$ U/mL, which allows for 79.6 % of patients to be correctly classified as diseased or healthy, with a sensitivity of 61.5 % and a specificity of 96.5 %. However, according to the use given to the test, more than one point can be selected: where sensitivity or specificity is favored (higher negative or positive predictive value).

There are cases in which not only the test under study contains more than two strata, but even the gold standard. In these cases the percentage of success and error can be estimated. Figure 6 compares clinical diagnosis of pulmonary embolism considering the diagnosis by ventilation/perfusion scan as the gold

Table I Bayes' theorem										
		<i>р</i> (<i>P</i> +/ <i>E</i> +) р (<i>E</i> +)								
	p(E+/P+) =	р (P+/E+) р (E+) + р (Р+/Е–) р (Е–)							
p (E+/P+)) =	a posteriori pro in case of a pos tive predictive	bability of having a certain disea sitive test; corresponds to the <i>pc</i> <i>value</i> (PPV).	ise)si-						
p (P+/E+)) =	probability of a has the disease	positive test result when the pat e; corresponds to <i>sensitivity</i> .	ient						
p (E+)	=	<i>a priori</i> probabi ding to the pop corresponds to	ility of having the disease accor- oulation that the subject belongs to prevalence.	to;						
p (P+/E–)) =	probability of a does not have positives or 1-s	positive test result when the pat the disease; equivalent to false specificity.	ient						
p (E–)	=	a priori probabi corresponds to [1 – p (E+)].	ility of not having the disease and 1-prevalence.	d						
	Prevalence Sensibility Specificity PPV NPV	11 % 59 % 94 % 57 % 95 %	56 % 59 % 94 % 94 % 64 %							

The negative predictive value is estimated in the same way reversing the signs of the formula [e.g.: p (E+/P+) changes to p (E-/P-)]

standard; the percentage of accuracy corresponds to the cells where both clinical diagnosis and the gold standard match, i.e. in cells a, e, i (40 + 90 + 70), with this being 66.66 %, and our percentage of errors overestimating the diagnosis corresponds to cells b, c, f(30 + 20 + 10), with this being 20 %; finally, the percentage of error underestimating the diagnosis is comprised by cells d, g, h(7 + 30 + 3), with this being 13.33 %. However, there is the possibility of wanting to handle the outcome only with two possibilities; in this case, the scans with low and moderate probability could be grouped and talk about a scan with high probability of pulmonary embolism or without high probability, or grouping those with high and intermediate probability and leaving those with low probability in a single group. This same procedure can be performed with the clinical scale, so that by having only four cells, the traditional usefulness estimators of a diagnostic test can be used, or preserving the three strata of our test under study and calculate a ROC curve.

Diagnostic Test Applications

It should remain clear that the application of a test may have different purposes:

- If a screening test is wanted, a high sensitivity test should be used, even if it has low specificity (e.g., test strips to measure blood glucose, to search for suspected diabetes mellitus).
- 2. If ruling out a given disease is wanted, a test with high sensitivity and, if possible, high specifity is used (high negative predictive value, e.g., ELISA for HIV), since, although when positive it is not diagnostic, when negative it does rule it out.
- 3. If we want to confirm a diagnosis in a patient suspected of having a certain disease, a test with high specifity and, if possible, high sensitivity is used (high positive predictive value, e.g., Western-Blot for HIV), since, although when negative it does not always rule the disease out, if positive, it establishes the diagnosis.



Figure 3 Use of Fagan's nomogram and likelihood ratios

Þ.

đ

Creatine phosphokinase (µU/mL)	Dar	mage	123	-		222	100		Γ.			14		20
	Organic	Metabolic	1		13	51	10	46	1	31	1	19	3	0
≥ 76	3	0	2											
61-75	(1)	0	25	21	3	4								
46-60	2	0			24	22	5	(0)						
31-45	9	0					22	27	t	0				
16-30	8	- Ť							16	27	17	18		
≤ 15	-10	26									10	20	27	27
		_											0	100

Sensitivity (a/a + c) %	3 7 0	211	19	41	63	100
Specificity (d/b + d) %	100	100	100	100	96	0
False positives %	0	0.	05	0	4	100
Error	25	24	22	15	11	27

Figure 4 Estimation of sensitivity and specificity at different cut-off points to identify organ damage in coma patients





Ventilation/perfusion scan (gold standard)

Figure 6 Assessment of clinical diagnosis efficacy in identifying pulmonary thromboembolism

Cells *a*, *e*, *i* = matches, in this case 66.66 %

Cells *b*, *c*, f = errors overestimating the diagnosis, in this case 20 %

Cells d, g, h = errors underestimating the diagnosis, in this case 13.33 %

Ordering tests in excess, whether justified or not, generates abnormal results even in normal people, which in turn triggers a cascade of more expensive and riskier tests, in addition with anxiety for the patient.

Common Errors When Elaborating a Diagnostic Test

We already explained how to estimate the efficacy of a diagnostic test and how to make use of it; however, we should watch out for possible causes of systematic errors, with two of them standing out in particular:

- 1. Inadequate selection of patients.
- 2. Inadequate interpretation of both the test under study and the gold standard.

The selection of an inadequate spectrum of patients may happen from the clinical or the pathological point of view. For example: the efficacy of a sputum cytology study is not the same for the detection of lung cancer in a patient with a history of heavy and prolonged smoking, weight loss, cough with hemoptysis and dyspnea, than in a patient who only has a cough and whitish expectoration, nor is the efficacy of carcinoembryonic antigen measurement equal for the detection of colon cancer in a patient with Dukes' stage A, compared with a patient with stage D. It is essential for every diagnostic test to be performed with the participation of patients that cover the entire spectrum of the disease, and, in addition, that the proportion of patients in each stratum is reported, so that its usefulness in other populations can be determined. On the other hand, concomitant diseases and used therapies that may alter the efficacy of the test under study should be considered. The control group must have been selected with the same criteria than the problem group, i.e., using the same entrance door, in order for the comparison to have clinico-methodological significance.

With regard to the most common mistakes during the elaboration of a diagnostic test, it is common that when assessing the test under study, the result for the gold standard is already known; this generates an interpretation bias because the assessor is expecting a certain result. Occasionally, the performance and the assessment of the test under study precede the gold standard and influence on the selection of patients undergoing the latter, or on its interpretation when it has a subjective component and, not infrequently, the test under study is part of the gold standard with which it is compared. All these deviations overestimate the usefulness of the test.

These two large errors can be avoided during the execution of a diagnostic test if the sensitivity and specifity values are considered only when:

a) The spectrum of the disease in the population where it is to be applied is equal to the spectrum of the disease with which the study was developed.

b) The assessment of the test under study and the gold standard has been performed in a blinded and independent manner in all patients.

Finally, it should be emphasized that if the quality of a diagnostic test depends partially on mathematical strategies, the clinical judgment that it derives from is

Bibliography

- Altman DG, Bland JM. Diagnostic tests 1: sensitivity and specificity. BMJ. 1994;308:1552.
- Altman DG, Bland JM. Diagnostic tests 2: predictive values. BMJ. 1994;309:102.
- Fagan TJ. Nomogram for Bayes's theorem. N Engl J Med. 1975;293:257.
- Feinstein AR. Clinical epidemiology. The architecture of clinical research. Philadelphia: W. B. Saunders Company; 1985.
- Grund B, Sabin C. Analysis of biomarker data: logs, odds ratios, and receiver operating characteristic curves. Curr Opin HIV AIDS. 2010;5(6):473-9.
- Jaeschke R, Guyatt G, Lijmer J. Diagnostic tests. En: Guyatt G, Rennie D, editors. Users' guides to the medical literature. Chicago: AMA Press; 2002: p. 121-140.

more relevant. And although the sensitivity and specificity estimation starts with the presence or not of the disease, in clinical practice, the study of the patient occurs with the presence or absence of the symptom or sign (clinical or para-clinical).

Additionally, in all cases, the reproducibility of the test should be assessed, provided that the groups under study are comparable; this means that, in addition to the selection of both populations under the same criteria, the distribution of subjects within the different degrees of disease severity must be similar. It should be remembered that, in everyday practice, patients are treated one at a time and that, therefore, it is essential to have a full knowledge of the severity of the disease in the group under study for its subsequent application, so that the patient can be assessed and treated according to the severity of his/her condition and not according to the average severity of the disease in the group in which the diagnostic test or treatment were assessed.

- Sackett DL, Straus S, Richardson WS, Rosenberg W, Haynes RB. Evidence-based medicine. How to practice and teach EBM. Second edition. Edinburgh: Churchill Living-stone; 2000. p. 67-93.
- Sackett DL, Haynes RB. The architecture of diagnostic research. BMJ. 2002:324;7336-56.
- Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Clinical epidemiology. A basic science for clinical medicine. Third edition. US: Little Brown; 2009.
- Soreide K, Korner H, Soreide JA. Diagnostic accuracy and receiver-operating characteristics curve analysis in surgical research and decision making. Ann Surg. 2011; 253(1):27-34.
- Tripepi G, Jager KJ, Dekker FW, Zoccali C. Diagnostic methods 2: receiver. operating characteristic (ROC) curves. Kidney Int. 2009;76(3):252-6.

III. Causality Studies

Juan O. Talavera, Niels H. Wacher-Rodarte, Rodolfo Rivas-Ruiz

Although the need of solving a clinical problem leads to the establishment of a starting point for approaching it (risk, prognosis or treatment study), in all cases, there is an attempt to attribute causality. Clinical reasoning, analyzed in detail in the book Clinical Epidemiology. The architecture of clinical research offers a simple guideline for undestanding this phenomenon and uses three components: baseline state, maneuver and outcome. In this model, different systematic errors are described (biases), which can occur when features of these basic components are overlooked. Omisions of characteristics at the baseline state produce an inadequate assembly of the population and the susceptibility bias; in the application or assessment of the maneuver, the execution bias; and in the assessment of the outcome, the detection bias and the transference bias. Thus, it is important to emphasize that if this form of reasoning facilitates the comprehension of the causal phenomenon, variables to be selected in studies where causality will be attributed or not to them require additional clinical reasonings assessing their relevance.

Key words research causality

This article was originally published in Rev Med Inst Med Seguro Soc 2011; 49 (3): 289-294 and it has been reviewed for this issue.

Introduction

When trying to predict a future event, the physician has to differentiate two processes: one that occurs before the onset of the disease and other that develops once the disease is present. The first is known as risk and it is characterized by the association between a series of factors present in the healthy subject (known as risk factors) and the development of the disease; the second is known as prognosis and it is characterized by the association between a series of features present at the beginning of the disease (known as prognostic indicators) and its outcome.

Multiple interventions, either preventive or therapeutic, add up to these two events; the former are intended to prevent the onset of the disease and the latter, to revert or reduce the damage caused by it. The event whereby a baseline condition (health or disease) is modified by a maneuver (risk factors, prognostic indicators or treatment), and which in turn produces a new condition known as outcome (prevention or onset of the disease and progression or resolution of harm), corresponds to a causative event. That is, in these three cases –whether our objective consists in identifying risk factors, an etiologic agent, prognostic indicators or assessing a treatment– attribution of causality is intended.

Although the need to solve a clinical problem leads us to establish a starting point to address it –risk, prognosis or treatment study–, in the real world there is a strong association between its components. For this reason, when assessing any of them, it is essential for the relevance of the other two to be considered within the assessment. This action is often carried out under the term *control of confounding factors*.

Thus, the study of causality for assessing a treatment is not only limited to the evaluation of therapy, but it obliges to estimate the contribution of all prognostic indicators existing at baseline state that participate in the disease of interest.

Likewise, when trying to prevent the onset of a disease with some maneuver, we must assess the different risk factors specifically associated with this disease. This requirement of measuring the impact of the different risk factors and prognostic indicators when assessing a therapy is consistent with the requirement of assessing the different therapeutic procedures when what we are trying to evaluate are the risk factors or prognostic indicators.

Clinical Reasoning in Causality Studies

Clinical reasoning, which is analyzed in detail in the book *Clinical Epidemiology*. *The architecture of clinical research* offers a simple approach for understanding the phenomenon of causality. Figure 1 shows the basic model comprising the baseline state, the maneuver and the outcome. This model describes different systematic errors (biases) that may contribute to the omission of some characteristics of the three basic components.

Errors at the Baseline State

The first two errors are related with omissions of baseline state characteristics and these are improper assembly and susceptibility bias.

Improper assembly refers to the selection of a population not susceptible to experience the outcome of interest with a proposed maneuver; for example, it is rather impractical to test a vaccine in a population with low incidence of the disease we are trying to prevent, since the size of the sample would have to be enormous; it is also inconvenient to assess the kidney-protecting effect of an ACE in a population of newly-diagnosed diabetic patients, since the follow-up would have to be very long.

Susceptibility bias refers to the pre-maneuver likelihood that the subject has of experiencing a certain outcome; for example, the presence of overweight or obesity increases the likelihood of an infarction in a diabetic patient, regardless of the poor metabolic control he may have.

The characteristics that must describe the baseline state to avoid these errors are shown in Figures 2a and 2b, i.e., the method used to select the population, the diagnostic demarcation and the prognostic stratification.

Within the prognostic stratification, anatomo-histology has been used as the main indicator, especially in oncology, followed by the functional aspect. In clinical practice, it is common to use multiple prognostic indicators in order to stage the disease according to the patient's condition. The following stratification groupings are the most common:

Primary

Stratification by status: it includes the performance, nutritional and mental status of the patient. Performance status has ben assessed with scales such as Karnovsky or ECOG, based on the patient's ability to perform his/her daily activities, in such a way that a patient who is not self-sufficient is more affected than that who can perform his/her tasks. Nutritional status impacts on the immune response and the hemodynamic stability. Patients with low albumin levels have been observed to show an important increase in mortality compared with those with higher levels. Other forms to assess nutritional status could be the body mass index and the waist-hip ratio when trying to assess the impact of overweight or body fat distribution; additionally, two of the most important features for assessing the mental status are the presence of depression and anxiety, among many other conditions.

- Morphologic stratification: it refers to the distinct location and damage of the pathology. An example is the histologic lineage of tumors and cytogenetic or immunophenotypical markers (for example, two tumors with the same extent of disease may have different prognosis according the histologic lineage, the presence of tumor markers or karyotype alterations; also, a patient with heart failure may have different prognosis according to the degree and type of valvular damage).
- Clinical stratification: it considers the severity of the disease, for example, the patient with grade IV heart failure (acute pulmonary edema) does not have the same probability of death than the patient with grade II (dyspnea with moderate exertion), even when the anatomical condition in both cases may be a mitral stenosis with the same valvular opening diameter.
- Chronometric stratification: it considers two components, the patients' age and the length of the disease. Regarding the first one, many diseases have



Figure 1 Basic model of the causality phenomenon

greater impact at both extremes of life and are associated with higher susceptibility to a poor outcome; additionally, older individuals have lower life expectancy. Regarding the length of the disease, if two patients suffer the same harm, but in one of them the disease is of recent onset while in the other it is of long evolution, the prognosis will be better in the latter since those patients with less aggressive disease have already been selected.

Stratification by comorbidity: it refers to the coexistence of any other pathological process that may alter the result of interest. Different conditions exert different impact on the outcome, and even in a same condition, the impact is generally related with the

degree of illness; for example, in a patient with acute myocardial infarction, the prognosis is better when the comorbidity is rheumatoid arthritis than when it is diabetes mellitus.

. Stratification by previous maneuver: two items can be identified here: the first and most widely used is the early response to a preventive or therapeutic maneuver, i.e., a better prognosis is expected upon an early favorable response. The second refers to the adverse impact of a maneuver. Practically every maneuver is known to entail a risk; however, not in all of them it has the same magnitude. Thus, safety should be considered as a prognostic indicator for any therapy.

b) To prevent inadequate assembly and susceptibility bias

a) Patient with heart failure due to valve disease Population selection mode



- to identify when there is response differential
- at different stages of disease. Overall response

is the same, but not by stage

Figure 2 Features to be considered at the baseline state



Figure 3 Features to consider in the maneuver

• Stratification by inheritance: the impact of genetic makeup has been identified as a risk factor for several diseases and with an increased aggressiveness thereof or higher risk of harm to target organs, as in diabetes.

Secondary

Social, economic and cultural conditions, as well as the ways of coping with disease, often have a lower impact than the biological components within the prognosis; however, sometimes they are crucial, such as having access to health care services in emergency events, or the change in lifestyle in some chronic diseases.

A distinctive strategy of clinical trials to avoid susceptibility bias is the random allocation of subjects to the treatment arm, seeking, among other things, that known and unknown factors potentially related with the outcome are evenly distributed between the groups to be compared. Other benefit is to prevent that those in charge the allocation are tempted to include a subject with better prognosis in a particular arm, since randomization facilitates the blinding of treatments and seeks to homogeneously distribute the subjects with different likelihood of treatment adherence and different likelihood of study dropout. It should remain clear that although random allocation seeks that the groups to be compared are homogeneously distributed at their baseline state, it does not show the effect of the maneuvers on the different strata (Figure 2c).

Errors in the Maneuver

The third systematic error, known as performance bias, is related with omissions in the application or assessment of the maneuver, and it refers to the differences generated by quality differences between the maneuvers to be compared or by an uneven use of additional maneuvers between groups (also known as peripheral maneuvers); for example, a surgery is not the same when performed by a recently graduated surgeon than when performed by a physician with extensive experience, nor are comparable two surgeries when in one of them the patients are well nourished or brought to hemoglobin normal values, while in the other group they are not. Features that have to be considered in the maneuvers in order to prevent these errors are shown in Figures 3a and 3b, which consist in adequate application of the maneuvers.

In clinical trials, there is a strategy intended to handle errors generated by an inadequate application of the maneuver, which is the way of analyzing the information, either by means of an intention-totreat analysis or a per-protocol analysis. The *intentto treat analysis* consists in analyzing the subjects in the group they were allocated to at the beginning of the study, regardless if they were compliant with the therapeutic protocol or not. The *per-protocol analysis* consists in analyzing only those subjects who were compliant with the therapeutic protocol. In observational studies, since there is no randomization to the maneuver, this is graded within the groups, thus enabling the comparison of the different degrees of quality in the maneuver application.

Errors in the Outcome

Detection bias occurs during the assessment of the outcome, which relates to an uneven detection of the outcome between groups and it occurs mainly for two reasons:



Figure 4 Main features to consider when assessing the outcome

- A higher number of assessments in some group, mainly due to more side effects, continuous dose adjustments or comparison of populations with different healthcare accessibility.
- Presence of diagnostic suspicion.

In the assessment of the outcome it is important to identify whether it is a final outcome or an intermediate regulation; for example, in the diabetic patient, the final outcome is to prevent damage in target organs; however, an intermediate regulation is glucose control; the latter may be considered a final outcome if symptomatology is trying to be reduced in the uncontrolled patient.

Another important aspect in outcome assessment is the identification and differentiation between the primary and the secondary outcome. This point is relevant since the selection criteria and the prognostic stratification, as well as the maneuver and the sample size estimation are carried out on the primary outcome and not on the secondary. Therefore, the results obtained in most studies are only exploratory for secondary outcomes (Figures 4a and 4b).

The last bias is also related with the outcome; it is generated by the loss of subjects under study and it is known as transfer bias (Figure 4c). Although in prospective studies the sample size is increased by 20 % in order to account for potential withdrawals, it is important to emphasize that this increase does not solve the transfer bias, but it rather maintains the stability of the data.

Final Considerations

In longitudinal studies, it is easy to apply these guidelines to study the phenomenon of causality; in the transversal ones they continue to be applicable, but this is a major challenge that translates into the creation of an artificial model regarding the temporary establishment of its components. Taking into account the elements described herein is recommended, not only for the reading of a causality study, but also for the creation of a research proposal.

It is important to emphasize that if this form of reasoning facilitates the understanding of the causative phenomenon, the appropriate thing to do for selecting those variables to which causality will be attributed to or not, is taking into account additional clinical considerations assessing their relevance. The basic principles were described in 1965 by Sir Austin Bradford Hill and were updated in 2000 by Kaufman and Poole; surely, over time, the number of factors to consider when judging a potential causal relationship will increase.

We hope that the causality approach herein described, which breaks down the basis of clinical practice, will facilitate the interpretation of medical literature and serve as guidance for the planning of research proposals and to increase the quality of medical care.

Bibliography

- Charlson ME, Frederic LS. The therapeutic efficacy of critical care units from two perspectives; a traditional cohort approach vs. a new case control methodology. J Chron Dis. 1987;40:31-39.
- Feinstein AR. Clinical epidemiology. The architecture of clinical research. Philadelphia: WB Saunders; 1985.
- Feinstein AR. Directionality and scientific Inference. J Clin Epidemiol. 1989;42:829-833.
- 4. Fletcher R, Fletcher S, Wagner E. Clinical

epidemiolgy: the essentials. 2nd. ed. Baltimore: Williams & Wilkins; 1988.

- Rothman KJ, Greenland S, Lash TL. Modern epidemiology. Third edition. Baltimore: Williams & Wilkins; 2008.
- Sackett D, Haynes R, Tugwell P. Epidemiología clínica una ciencia básica para la medicina clínica. Madrid: Ediciones Díaz de Santos; 1989.
- Talavera JO. Pronóstico. En: Ramiro M, Lifshitz A, Halabe J, Frati A, editores. El internista. Medicina interna para internistas. 3a. ed. México: Nieto Editores; 2008. p. 1893-1898.

IV. Appropriateness of the Statistical Test

Juan O. Talavera, Rodolfo Rivas-Ruiz

When we observe the difference between two therapies or the association of a risk factor or prognostic indicator with its outcome, we have to assess the certainty of the result. This assessment is based on a judgement that uses information related with the design of the study and the statistical handling of the information. In this article, the relevance of the selected statistical test is specifically mentioned. Statistical tests are chosen based on two features: the objective of the study and the type of variables. The objective can be divided in three groups of tests: a) those in which showing differences between groups or in a same group before and after a maneuver is wanted; b) those in which showing a relationship between variables is wanted; c) those in which predicting an outcome is pretended. As for the types of variables, we have two: quantitative (continuous and discontinuous) and gualitative (ordinal and dichotomous). For example, if we want to demonstrate age differences (quantitative variable) between patients with systemic lupus erythematosus, with and without neurological involvement (two groups), the adequate test is Student's *t*-test for independent samples; but if what is being compared in those same groups is the frequency of females (binomial variable), then the relevant statistical test is the chi-square test (χ^2).

Key words

biomedical research research projects statistics and quantitative data

This article was originally published in Rev Med Inst Med Seguro Soc 2011; 49 (3): 289-294 and it has been reviewed for this issue.

Introduction

When we observe the difference between two therapies or the association of a risk factor or a prognostic indicator with its outcome, a question arises: Is the result real? Deciding if it is real requires two complementary judgments:

- 1. The planning and development of the process that document such difference or association are free of errors, or at least these are of a minor magnitude, which does not modify the sense of the difference or association (i.e., appropriate design and adequate execution).
- 2. The size of the sample is sufficient to maintain the stability of data and the statistical test is suitable for the objective.

The planning and development of the process have been mentioned in the three previous chapters of this series. On the other hand, data stability will be discussed in detail in a subsequent article when the size of the sample and the *p*-value are addressed.

In this article, we will discuss the relevance of the selected statistical test. Undoubtedly, this knowledge will allow for us to understand more precisely the results obtained in clinical research studies and, of course, it will increase our ability to make an adequate use of them.

Study Objective and Type of Variable

Statistical tests are selected based on two features: the objective of the study and the type of variables. Within the study objectives we can identify three:

- 1. Demonstrating differences between groups or differences in a same group before and after a maneuver (e.g., treatment with drug *A* reduces high blood pressure in a greater proportion than treatment with drug *B*).
- 2. Showing relationships (correlation) between variables (e.g., serum creatinine rises as renal function decreases).
- 3. Predicting an outcome (e.g., the likelihood for the subject with sedentary life and overweight of developing type 2 diabetes mellitus).

Frequently, the models overlap, and thus, models initially identified to predict an outcome are sometimes used to demonstrate differences between two groups. This happens especially when the principal maneuver has to be adjusted (drug A versus drug B) for multiple factors (age, sex, body mass index, etc.). But the oppo-

site phenomenon also happens when looking to predict an event that will occur in the future but there are only one or two predictors available; in this case, a test to demonstrate differences is used.

It is important to clarify that the correlation basically is useful for seeing the magnitude of the association between variables, although it should remain clear that it does not establish causality. As a matter of fact, no statistical test can. This requires covering a number of principles described by Sir Austin Bradford Hill.

Defining the *type of variable* is relevant because it is the axis in the selection of the appropriate test depending on the desired objective. Within the types of variables there are two groups:

- 1. Quantitative: continuous and discontinuous or discrete. The former are characterized because they can take any value throughout a continuum (for example, 1.75 m height). On the other hand, discontinuous or discrete variables use exclusively whole numbers (parity, 1, 2, 3...). In both instances, the distance between one unit and another throughout its scale is equidistant.
- Qualitative: these include the ordinal and the dichotomous variables. The ordinal variable allows for the characteristic under study to be ordered and, unlike what happens in quantitative

Table I Weight of subjects studied under two therapeutic regimens									
	Group A	ł		Group B					
77			65						
78			69						
80			77						
82			78						
85	83.5	Average	85	83.5	Average				
85	85.0	Median	85	85.0	Median				
85	85.0	Mode	89	85.0	Mode				
86			93						
88			96						
89			98						

Central tendency measurements are equal, but the dispersion of data is different

variables, the distance between two categories is not equidistant (e.g., heart failure grades I to IV). Dichotomous variables, as their name indicates, are those with only two categories, which can be binomial (one option or another, e.g., male or female) or nominal (it refers to the presence or absence of the feature, e.g., alive at six months, yes or no).



Mean: 59.79

Standard deviation: 13.882. Two standard deviations at either side of the mean reflect 95 % of the population Average: 59.79, 95 % CI = 32.03-87.55

Figure 1 Histogram

It is important to mention the handling that the type of variable will suffer during the analytical process, starting with the collection of "crude" data, which means that this is only a collection of information from a group of subjects. In order for these data to have a useful meaning, they have to be organized and summarized. The simplest organization method is the frequency distribution tables; however, sometimes it is easier to understand their graphic representation through a histogram or frequency polygon. Regardless of the usefulness of this information, collected data are required to provide quantitative information, i.e., numerical indices reflecting different probability distributions are required, whose primary function is to model the behavior of a large variety of biological phenomena. These numerical indices include the measures of central tendency and the measures of dispersion.

- 1. Measures of central tendency (Table I and Figure 1).
 - *a) Mean:* it is the sum of a set of data divided by its total number. The symbol to represent the mean of a population is the Greek letter mu (μ), and the mean of a sample is represented by \overline{x} . It is the most widely used summary measure for quantitative variables.
 - b) Median: it is the value located exactly in the middle of the entire set of data. The median divides a distribution of data ordered exactly in two equal parts. The advantage of the median as a measure of central tendency is that it is not affected by the value of extreme data, a phenomenon that does occur with the mean. It is the type of summary measure most widely used for quantitative variables not following a normal distribution and for ordinal variables.



Figure 2 Normal distribution curve

- *c) Mode:* it refers to the most repeated value in a distribution. This measure is hardly used in medicine.
- 2. Most common measures of dispersion.
 - *a) Standard deviation:* it reflects the variation between the whole data set and it is used when these follow a normal distribution.
 - *b) Percentile*: it describes the position of a value of the distribution. It is used for quantitative variables not following a normal distribution and for ordinal variables.
 - *c) Range:* it is the difference between the highest and the lowest value of the distribution.
 - *d) Interquartile ranges:* these are referred to the values of the first and third quartile.

In clinical research, as in many other real-life phenomena, the most commonly analyzed data are quantitative, which in most cases show a Gaussian distribution, also known as normal distribution, which is characterized for having a bell-like shape, for being symmetric with regard to its mean, for having decreasing frequency values as they move away from the mean, and for never reaching zero (asymptotic). The mode and the median are equal to the mean; about 68 % of data are within ± 1 standard deviation from their mean and 95 % within ± 2 standard deviations (Figure 2). Thus, if the set of data is quantitative with a normal distribution, its summary measure will be the mean, and its dispersion measure, the standard deviation. However, if its distribution is not Gaussian, same as it is for an ordinal-type variable, its summary measure will be the median, and its dispersion measure, the percentile or rank. Generally, these variables do not have dispersion measures and when they are used, 95 % confidence intervals are preferred.

Appropriateness of the Statistical Test

Once we know our objective and the characteristics of our data (type of variable), we can consider the appropriateness of the statistical test (Table II). However, there are two more considerations when the objective is to demonstrate difference:

- 1. If it is a study in which the value of a data item is compared before and after a maneuver, either observational or experimental, it is known as related samples test, but if it involves the comparison of data between different groups, it is called unrelated samples test.
- 2. If it consists in a comparison between different groups, it is necessary to establish if it is going to be between two or more.

Table II Selection of the s	tatistical test acco	rding to the objective ar	nd type of variable			
		To demonstrat	te difference	To show relationship ^{&}	To predict 1 variable [‡]	
Type of variable	Type of sample	Two groups	Two groups Three groups		Outcome variable	
Quantitative	NR	Student's t*	1 factor ANOVA	Pearson	Linear regression	
(normal distribution)	R	Student's t **	1 factor ANOVA			
Qualitative ordinal	NR	Mann-Whitney U	Kruskal-Wallis	Spearman		
(free distribution)	R	Wilcoxon	Friedman			
Qualitative dichotomous	NR	χ² (or Fisher exact test)	χ ² (of linear ten- dency)	Phi coefficient	Logistic regression	
R	McNemar			Survival curves		

NR = not related; R = related; R = measure of the variable in the same subject at two different time-points

Student's t for independent samples

** Student's *t* for related samples

& For the correlation between 2 variables, the test of that at the lower scale is used (actually, no scale is lower; however, variables

have been ordered from quantitative continuous to dichotomous, by way of quantitative discontinuous and ordinal variables).

the predictor can be quantitative, dichotomous or ordinal (with these last transformed into *dummy*-like variables)

With the information already complete, with Table II we can verify if the selection of the statistical test was appropriate according to the variable and the objective. For example, if age is compared (quantitative variable with normal distribution in this case) between patients with systemic lupus erythematous, with and without neurological involvement (two groups), the appropriate test is Student's t-test for independent samples. But if that what is being compared between these same patients is the frequency of females (binomial variable), then the approppriate statistical test is the chi-square (χ^2) test. If that what is being compared between both groups is their degree of lupus-like activity (ordinal scale), the appropriate statistical test is the Mann-Whitney U-test. On the other hand, if that what we are shown is the magnitude of association (relationship) between age (quantitative variable with normal distribution) and the degree of lupus-like activity (ordinal variable), the relevant test is Spearman's r. Finally, if that what is sought to be predicted is the weight of a child (quantitative variable) based on age (quantitative variable), type of nutrition (ordinal variable: good, fair or poor) and sex (dichotomous), the appropriate test is the linear regression. But if that what is wanted to be predicted is the probability of infarction (dichotomous nominal) over the next 10 years based on age (quantitative), atherogenic risk (ordinal, low, moderate and high) and sex (dichotomous binomial) the relevant test is the multiple logistic regression.

Finally, we hope this article allows for the reason of the selection of the most widely used statistical tests in health research to be understood and, at the same time, to serve as a guideline to those who are taking their first steps in statistics. It is not sufficient for establishing if the obtained results are real; it will be necessary to take into consideration the design and execution of the study and the stability of the information, but this last issue deserves to be discussed in another section. The next chapters of this series will further address Student's *t*, Mann-Whitney *U* (with which we will address how to select the type of distribution of quantitative variables) and chi-square tests.

Bibliography

- Armitage P, Berry G, Matthews JNS. Statistical methods in medical research. 4th ed. Massachusetts, MA: Blackwell Publishing; 2002.
- Bland M. Introduction to medical statistics. 3rd ed. Oxford: Oxford University Press; 2003.
- Feinstein AR. Clinical epidemiology. The architecture of clinical research. Philadelphia, PA: W.B. Saunders; 1985.
- Feinstein AR. Multivariable analysis: an introduction. New Haven, CT: Yale University Press; 1996.
- Feinstein AR. Principles of medical statistics. New York, NY: Chapman and Hall/CRC; 2002.
- Le Chap T. Introductory biostatistics. Hoboken, NJ: New Jersey: John Wiley and Sons; 2003.
- Peat J, Barton B. Medical statistics. A guide to data analysis and critical appraisal. Malden, MA: Blackwell Publishing; 2005.
- 8. Portney LG, Watkins MP. Foundations of clinical research: applications to practice. 3rd ed. Saddle River, NJ: Pearson/Prentice Hall; 2009.

All the information available online







In the electronic version, you can:

- Read the web article
- Consult previous issues
- Download contents
- Suscribe online

http://revistamedica.imss.gob.mx/
V. Sample Size

Juan O. Talavera, Rodolfo Rivas-Ruiz, Laura Paola Bernal-Rosales, Lino Palacios-Cruz

In clinical research it is impossible and inefficient to study all patients with a specific pathology; therefore, it is necessary to focus on a sample. Estimating the size of a sample warrants the stability of the results and allows for feasibility of the study to be foreseen, depending on cost and patient availability. The basic structure for estimating the sample size is based on the premise that tries to demonstrate ---among other things--that the difference between two or more maneuvers in the subsequent state is real. For this, it is necessary to know the value of the expected difference (δ) and the dispersion measure of the data that gave rise to it (standard deviation), which usually are obtained from previous studies. Afterwards, other components are considered: α , which is percentage of type I error accepted in the claim that the difference between means is real, generally of 5 %; and β , which is the percentage of type II error accepted in the claim that the non-difference between means is real, generally from 15 to 20 %. These values are substituted in the formula or in some sample size estimation electronic program. Although summary and dispersion measures may vary according to the outcome measure and, consequently, the formula, the principle is the same.

Key words sample size confidence interval

This article was originally published in Rev Med Inst Med Seguro Soc 2011; 49 (3): 289-294 and it has been reviewed for this issue.

Introduction

In clinical research, it is impossible and inefficient to study all subjects affected by a specific pathology; therefore, when we read an article, the results it shows correspond to a portion of the entire population. The number of subjects included in a study is determined by a series of features that will be addressed later, but whose primary objective is to answer a question with the certainty that the obtained result is real. In addition to this, estimation of the sample size before starting a study allows for its feasibility to be considered depending on patient availability and cost. The lack of calculation in the sample size may cause an unnecessary expenditure of both financial and human resources. It is possible for study expenses to be unnecessarily increased due to a surplus number of subjects included in it, or for the investment made to turn out being fruitless when including an insufficient number of subjects to answer the research question.

The basic structure of the sample size estimation is based on the premise that tries to demonstrate that the observed difference between measurements made before and after the maneuver, or between two maneuvers in the subsequent state, is real and not due to random effects. This structure is the same regardless of the type of variables necessary to answer the research question. In other cases, the purpose is not demonstrating the veracity of a difference but rather to obtain the average value of a particular feature within a population, with a precision indicated by the upper and lower limits of the confidence interval (CI), which in most cases is requested to be 95 or 99 %.

Estimation for Two Groups

This purpose is exemplified when we try to demonstrate that blood pressure values are different with a certain drug versus another and that this difference is not due to casuality. To estimate the sample size, the first thing that is required in this exercise is the average (\bar{x}) of the diastolic blood pressure (DBP) values of the patients that took one drug (group *A*) or another (group *B*): assuming that the average DBP in group *A* is 90 mm Hg and in group *B* 85 mm Hg, then the difference between means will be 5 mm Hg, a value that represents the first component, which is identified as delta (δ).

Afterwards, it will be necessary to have some measurement of the variation of values within each group, since there will be patients with much lower and much higher pressures than the average; for example, from 60 to 112 mm Hg. This value will allow for the variation within each group to be observed and, at the same time,



Figura 1 Total group of hypertensive patients under pharmacological treatment

to know if values between groups overlap excessively in relationship with the average difference. In a quantitative variable, as in the described model, the measure of dispersion is known as standard deviation (SD).

As is shown in Figure 1, the DBP average for the entire population is 87 mm Hg, with a standard deviation of 9 mm Hg, whereas in Figure 2a, DBP average in group A is 90 ± 9 mm Hg ($\bar{x} \pm$ SD) and DBP average (Figure 2b) in group B is 85 ± 8 mm Hg ($\bar{x} \pm$ SD). This means that the general population has an average of 87 mm Hg, but that its values in regards to two standard deviations range from 69 to 105 mm Hg ($\bar{x} \pm$ 2 SD). In group A, with an average of 90 mm Hg, their values range from 72 to 108 mm Hg ($\bar{x} \pm$ 2 SD), and in group B, with an average of 85 mm Hg, their values range from 69 to 101 mm Hg ($\bar{x} \pm$ 2 SD). Average and variable of interest dispersion values are usually obtained from existing information in already published previous or preliminary studies.

Once we have a summary measure (average) and its corresponding measure of dispersion (DE), we have to consider:

- To what degree of certainty do we want to demonstrate that the DBP difference between groups is real? When this point is not taken into account, we may incur in what is known as type I error: accepting that the difference is real without it being so.
- 2. To what degree of certainty do we want to demonstrate that the non-difference is real? When this point is not taken into account we may fall into what is known as type II error: accepting that the non-difference is real.

The certainty with which a difference is usually accepted to be real is at 95 % and this corresponds to an alpha value (α) of 0.05, indicating that once we establish that there is a difference in DBP values between groups, there is a 95 % of certainty that such difference is real and only a 5 % of error is accepted.

To accept that the non-difference found is real, we must have an initial pre-established capability to find significance when there is a difference, which is known as power and it is represented by the difference of 1 – beta (β). The accepted power value may vary from 80 to 95 %, which corresponds to a β -value of 20 to 5 % respectively.

At this point, all the components necessary for estimating the size of the sample are already available:

- δ: difference between the summary measures (in the example, it is the difference between the means).
- SD: measure of dispersion, which in the example is the standard deviation.
- Type I or α: error accepted in the claim that the difference between the means is real, usually of 5 % (0.05).
- Type II or β : error accepted in the claim that the non-difference between the means is real, generally ranging from 5 to 20 %.

Ignoring these different components usually causes that, at the end of the study, the size of the sample is insufficient and, thus, even if there is a clinically significant difference (≥ 10 %), no statistical difference is found (p < 0.05), which means insufficient power (< 80 %) and, therefore, a type II error.





Mean = 84.970Standard deviation = 08.369n = 30

Figure 2 Hypertensive patients under treatment A and B, respectively

Mean Differences

With the above components, sample size is estimated using the formula of mean differences:

$$n = 2 \left[\frac{(Z_{\alpha} - Z_{\beta}) DE}{\mu_1 - \mu_2} \right]^2$$

Where:

- Z_{α} = value of *z* related to $\alpha = 0.05$ (extracted from reference tables)
- $Z_{_{\beta}}~=~value~of~z$ related to $\beta=0.20$ (80% power).
- SD = standard deviation
- $\mu_1 = \operatorname{group} A \operatorname{mean}$
- μ_2 = group *B* mean

According to the example, the substitution of values would be as follows:

And substituting in the formula:

$$n = 2 \left[\frac{(1.96 - (-0.84))9}{90 - 85} \right]^2 50.80 \approx 51$$

Therefore, it is necessary to include 51 patients in each group if obtaining 80 % of probabilities (80% power) is desired for the detection of a mean difference of 5 mm Hg or more between the two treatment groups.

Difference of Proportions

It is used when the outcome of interest is expressed in terms of proportions. Example: comparison of two groups of patients with overweight. The first group of patients receives medication and the second, dietary advice. If the outcome event is assessed after six months and measured as the proportion of patients who manage to normalize their weight (body mass index under 25), what is it required?

$$\beta = 0.10$$

- $\pi_1 \pi_2$ = (difference of proportions) group 1 proportion minus group 2 proportion, which is clinically significant
- SD = the formula for its determination is 1 – group proportion, which remains included within the global formula

The formula for the determination of the sample size for proportions difference is:

$$n = \left[\frac{Z_{\alpha} \sqrt{2\pi_1(l-\pi_1)} - Z_{\beta} \sqrt{\pi_1(l-\pi_1) + \pi_2(l-\pi_2)}}{\pi_1 - \pi_2}\right]$$

Where:

$$Z_{\alpha} = (\alpha = 0.05) \ 1.96$$

$$Z_{\beta} = (\beta \approx 0.10 - 0.20) \approx -1.645, -0.84$$

$$\pi_{1} = \text{group 1 proportion}$$

$$\pi_{2} = \text{group 2 proportion}$$

$$\pi_{1} - \pi_{2} = \text{difference between group 1 proportion} - \text{group 2 proportion, which is clinically significant}$$

Assuming that for the study problem it would be expected that at six months, the group receiving drug therapy would succeed in 70 % of cases, whereas the group with dietary advice would succeed in 50 % of cases, the values would be replaced in the formula as follows:

$$= \left[\frac{1.96\sqrt{2 \times 0.70 \times 0.30} - (-1.645)\sqrt{(0.70 \times 0.30) + (0.50 \times 0.50)}}{0.70 - 0.50}\right]^2$$

$$n = \left[\frac{2.435}{0.20}\right]^2 = 12.18^2 = 148.35 \text{ subjects for each group}$$

This result must be rounded to the upper digit. Thus, the sample must include 149 subjects in each study group if 90 % of possibility (90 % power) is wanted for the detection of at least a difference of 20 % in the percentage of success in weight loss between the two treatment groups used as example.

Estimation for a Group

On the other hand, when the objective is to obtain the average value of a particular feature within a population, the sample size estimation requires the average value (proportion or mean) and its upper and lower limits indicated by the CI, which in most cases is requested at 95 or 99 %.

For a Proportion

To estimate the sample size for the prevalence or proportion of an event or feature, different components must be identified, starting with the summary measure (p_0) , which corresponds to the expected proportion, and its precision (*d*), which is equivalent to half the amplitude of the CI. If we understand this section, we can solve the sample size formula based on the precision formula, which in turn comes from the estimation of the standard deviation of a proportion:

$$d = Z_{\alpha} \sqrt{\frac{p_0 \times q_0}{n}}$$

Solving for *n* yields:

$$n = \frac{Z_{\alpha}^{2} \times p_{0} \times q_{0}}{d^{2}}$$

In this case, $q_0 = (1 - p_0)$; therefore if we want to look for a prevalence (p_0) of 20 %, the q_0 value would be 1 - 0.2 = 0.8. Therefore, to make the calculation of the sample size for a proportion, the following must be considered:

- Precision (*d*, equal to ½ the amplitude of the CI), whose value is conferred by the investigator and corresponds to the degree of error that might be tolerated at each side of the mean; for example, for an error of 8 % based on the mean, its d2 would be 0.0064 (0.082 = 0.0064).
- Confidence, also known as Z_{α} corresponds to 1α .
- The p_0 value intended to be estimated.

Example: How many preterm infants will it be necessary to study in order to verify if the estimated prevalence of metabolic bone disease in a neonatal intensive care unit population is 20 %, considering an accuracy of 8 % and an α of 0.05 %? With a confidence level of 95 % ($\alpha = 0.05$; $Z_{\alpha} = 1.96$), $Z_{\alpha}^2 = 3.8416$, which when solving:

$$N = (3.8416 \times 0.2 \times 0.8)/0.0064$$
$$N = 96.04$$

Therefore, the required sample size will be 97 children for an expected prevalence of 20 % with a CI ranging from 12 to 28 %.

As we can observe, the size of the sample will depend on the expected accuracy of the error based on the mean, so that for a narrower CI, a lower d is required; 0.08 and 0.04 values are generally used, with the latter being the most accurate (or the one with less error); therefore, a larger sample size will be required. Similarly, if a confidence level change from 95 to 99 % is desired, as requested in studies of genetic determinants, the sample size will increase. Table I shows some variation examples according to these parameters.

For a Mean

If the above is understood, it will be easy to understand the components for estimating the sample

Table I Different sample sizes according to different values of confidence level (α), prevalence (*p*) and precision (*d*)

a (Ζ _α)	р	d	n
0.05(1.960)	0.2	0.08	97
0.05(1.960)	0.2	0.04	385
0.01(2.576)	0.2	0.08	166
0.01(2.576)	0.2	0.04	664

size for a mean. Similarly, the basis is the formula for the CI of the mean:

$$IC \ de \ 95 \ \% = \overline{\mathbf{x}} \pm Z_{\alpha} = \frac{DE}{\sqrt{n}}$$

In this case, precision (d) is calculated as follows:

$$d = Z_{\alpha} \frac{DE}{\sqrt{n}}$$

Therefore, the formula for the calculation of the sample size for estimating a mean is:

$$n = \frac{Z^2 \times DE^2}{d^2}$$

This formula requires the knowledge of Z α , SD and the desired *d*. Thus, the sample size for an expected mean depends on Z α (1.96 for $\alpha = 0.05$), on the standard deviation that has been observed in previous studies, as well as on the desired precision.

Final Considerations

It should be clear that the assumptions above are not the only ones for estimating the size of a sample, so that if we want to estimate it in order to demonstrate differences in cumulative incidence rates (Hazard risk ratio) or in units obtained in models such as Cox proportional hazards survival curves, the estimation is more complex since it considers the outcome over time; nevertheless, the basic concept is the same.

On the other hand, if the intention is controlling for multiple confounders or exploring multiple risk factors using a multiple logistic regression model, then it will be necessary using a number of events per variable, for which 10 to 20 subjects for each will be required in the smallest of the outcome groups (so that if mortality is 30 %, this is the smallest of the groups, since the remaining 70 % will survive).

Recommended readings

Cohen J. Statistical power analysis for the behavioural sciences. Second edition. New Jersey: Lawrence Earlbaum; 1988.

- Dawson B, Trapp R. Basic and clinical biostatistics. Fourth edition. Lange Medical Books-McGraw-Hill; New York, USA; 2004.
- Hulley SP, Cummings SR. Diseño de la investigación clínica. Un enfoque epidemiológico. Barcelona: Doyma; 1993.

Portney LG, Watkins MP. Foundations of clinical research. Applications to practice. Appleton and Lange; Chicago, IL: 1993. Feinstein AR. Principles of medical statistics. London, UK: Chapman and Hall-CRC; 2002.

For the calculation of sample size

- Brixton Health. [Website]. Epicalc 2000. Available at http:// www.brixtonhealth. com/epicalc.html
- Epilnfo 2000. Available at http://huespedes.cica.es/huespedes/epiinfo/ espanol.htm
- Department of Biostatistics, Vanderbilt University. [Website]. PS: Power and Sample Size Calculation. Available at http://biostat.mc.vanderbilt.edu/ wiki/Main/PowerSampleSize

VI. Clinical Relevance

Juan O. Talavera, Rodolfo Rivas-Ruiz, Marcela Pérez-Rodríguez

In clinical practice, the maneuver that is usually selected is the one that achieves an outcome with at least 10 % of direct superiority or when the number needed to treat is ≈10. Although these parameters serve for estimating the magnitude of an association, we are forced to differenciate the measures of impact (attributable risk, preventable fraction), association (relative risk, odds ratio, risk ratio) and frequency (incidence and prevalence), which are applicable when the outcome is nominal. We also have to identify the way for measuring the strength of association and the magnitude of association when the outcome variable is quantitative. Not unfrequently, association measures are interpreted as if they were impact measures, v.gr., for a relative risk of 0.68, a 32 % of outcome reduction is assumed without considering that this is a relative reduction that can be generated by a ratio of 0.4/0.6, 0.04/0.06 or 0.00004/0.00006 as well; however, the direct reduction is 20 % (60-40 %), 2 % and 2 per 100 000, respectively. Therefore, in order to estimate the impact of a maneuver, it is important that the direct difference or the number needed to treat is available.

Key words

association measures exposure risk or outcome relative risk number needed to treat

This article was originally published in Rev Med Inst Med Seguro Soc 2011; 49 (6): 631-635 and it has been reviewed for this issue.

Introduction

Even with a well-designed trial, with an adequate statistical analysis and sample size, in which statistical significance in the association between a maneuver and an outcome is shown (whether it is the association between a risk factor or preventive maneuver and the occurrence of a disease, or between a prognostic factor or therapeutic maneuver and the course of the disease), the clinician needs to identify the magnitude of this association -- impact of the maneuver-- in order to consider its usefulness in common clinical practice, in which most of the time, the benefit of a therapeutic maneuver is considered and it is usually selected that which achieves a favorable outcome with at least 10 % of direct superiority over others. This means that, for example, if the outcome is survival and the selected maneuver is A, it is expected for it to be 10 % superior than standard maneuver B (70 % two-year survival for maneuver A versus 60 % for maneuver B), or if the outcome is the level of glucose, then a reduction of at least 10 % is expected (from 140 to 126 mg/dL). And if the outcome is heart failure, a reduction of at least 10 % is expected in the degree of heart failure (overall, at least 10 % more of patients improving their heart failure grade). It should be noted that the substraction of a proportion from another was made directly, whereas for quantitative data, 10 % is estimated based on the reference value.

In public health or preventive medicine, direct differences lower than 10 %, and even as low as 4 to 7 %, are highly relevant, since susceptible populations may include millions of subjects. The same happens in clinical care, where the rate of unwanted outcomes is around 10 %, for which any expected reduction will be lower than this and its relevance will depend on the severity and cost of the disruption. On the other hand, in case of adverse events, differences even lower than 10 % are significant, especially depending on the severity of the event. Nevertheless, in most clinical situations, a minimum gain of 10 % is considered desirable.

While for clinicians it is common and understandable a percentage difference to estimate the impact of an association, in literature there is a series of calculations known as *impact measures* that, in spite of being discretely more elaborated, turn out to be an association between proportions. In the process of obtaining the impact measures, *association measures* are estimated (indicators that assess the strength at which a variable or feature is associated with another), which would be meaningless if they would not be accompanied by the certainty that such association is real and not due to chance, and for this purpose, *statistical significance* is estimated (an association is real when the

Table I Double input table for measures of re	elative frequency (exampl	le), association and impact	
	Outcome +	Outcome –	Total
Exposed (treated)	a 5	b 95	a + b = 100
Non-exposed (placebo)	c 15	d 85	c + d = 100
Total	a + c = 20	b + d = 180	
Clinical trial and cohort	Formula	Example	Interpretation
Exposed incidence (Ei)	le = a/a + b	5/100 = 0.05	5 new cases in 100 subjects or 5 %
Incidence of observed or non-exposed (Io)	lo = c/c + d	15/100 = 0.15	15 new cases in 100 subjects or 15 %
Relative risk (RR)	RR = Io – Ie	0.05/0.15 = 0.33	A protection exists. Relative or risk reduction. The risk is below the unit
Absolute risk reduction (ARR) (attributable risk [AR])	RR = Io – Ie	0.15 - 0.05 = 0.1	The direct reduction of risk attribu- ted to treatment is 10 %
Number needed to treat (NNT)	NNT = 1/RAR	NNT = 1/0.1 = 10	10 people have to be exposed to observe the beneficial effect in one
Attributable fraction (AF) (for RR > 1)	le – lo/le	Since in this example RR is > 1, AF is not calculated	Interpreted as the proportion of cases exposed due to the risk factor
Relative risk reduction (RRR) (for RR < 1, preventable fraction)	RRR = 1 – RR x 100	1 – 0.33 x 100 = 67 %	67 % of cases were prevented due to the exposition factor
Case-controls, and cross-sectional surve	ÿ		
Prevalence of exposed (Pe) (only in cross- sectional survey)	Pe = a/a + b		Number of events in the exposed group (used in cross-sectional studies)
Prevalence of non-exposed (Po) (only in cross-sectional survey)	Po = c/c + d		Number of events in non-exposed group or control (used in cross- sectional studies)
Exposition factor prevalence in cases	PfrCa = a/a + c	5/20 = 0.25	25 % of cases were exposed to exposition factor
Exposition factor prevalence in controls	PfrCo = b/b + d	95/180 = 0.527	52.7 % of controls were exposed to exposition factor
Odds ratio (OR)	a x d/b x c	RM = 5 x 85/15 x 95 RM = 425/1.425 RM = 0.29	The exposed group is protected. The risk is below the unit

Incidence and prevalence are frequency measures; relative risk and odds ratio are considered association measures; and absolute risk reduction and relative risk reduction are impact measures. Another association measure is the risk ratio, obtained in the Cox proportional hazards survival analysis (Hazard risk ratio, HRR). Attributable risk and preventable fraction can also be estimated based on the OR (instead of using Ei using Pe and instead of Io, Po)

p-value is < 0.05). Before these two types of measures, during the process of data management, we have to make use of what is known as *frequency measures*, which estimate the absolute number of events. It should be emphasized that, in most cases, what we

observe in articles are *relative frequency measures*, in which the number of events is related with the total number of individuals in the population or sample under study, so that comparisons can be made at a later stage between groups with different *n* (Table I).

Table II Examples of RR and 95 %	confidence interv	/als				
	A	N	E	3		
Study examples	Events	Total	Events	Total	RR (CI 95 %)	RR (CI 95 %)
Aspirin (A) versus placebo	65	5000	95	5000	0.68 (0.50, 0.94)	-
Coffee consumption (A) versus placebo (B)	25	5003	24	5000	1.04 (0.60, 1.82)	0.50 .7 1 1.5 2.0
With dyslipidemia (A) versus healthy (B)	205	5000	115	5000	1.78 (1.42, 2.23)	Protection Risk

RR = relative risk; 95 % CI = 95 % confidence interval; RRR = relative risk reduction

Aspirins have a statistically significant RRR of 32 %; dyslipidemia has a statistically significant RR increase of 78 %. Coffee consumption has a non-statistically significant relative increase of 4 %.

In clinical practice, measurements of the association between two variables (maneuver and outcome) by means of relative risk (RR), odds ratio (OR) and hazard ratio (Hazard risk ratio, HR) are common and are interpreted similarly; variables with a value below 1 are considered protective, whereas those with values above 1 are considered risk variables. This way, we have that common risk for the population or sample of suffering or having the event of interest without identifying any factor, either protective or of risk is 1 (which corresponds to the incidence or prevalence of the event in the entire sample or popu-

Table III Association measure	sures and equivalents for q	uantitative variables			
Qualitat	ive dependent variable (n	ominal)	Quantitative dependent variable		
Frequency measures	Association measures	Impact measures	Power of association	Magnitude of association	
Incidence • Incidence rate • Cumulative incidence	RR (cumulative inci- dence ratio)	Attributable risk (etiolo- gic fraction, ARR and NNT)	ľ2	% of difference of the means b coefficient	
	HR (Hazzard risk ratio)	RRR, AF (attributable fraction)	R ²	% of difference of the means through the regres- sion equation	
PrevalencePoint prevalencePeriod prevalence	OR (prevalence odds ratio or crossover products)			(ŷ= a + b1X1)	
		r b coefficient R ²	% of proportion differen- ces through the probability equation		

 $\hat{y} = 1/1 + e^{-(a + b1X1...)} = probability of the event$

RRR = relative risk reduction

The NNT (number needed to treat) is a relatively new way for estimating the magnitude of association

 $\hat{y} = 1/1 + e^{-(a + b1X1...)}$

lation under study). But if we identify a risk factor, we observe that the incidence in this subgroup increases and that in those without this risk factor, it decreases in relationship with the risk of the entire population or sample. For example, if we consider the use of aspirin to prevent myocardial infarction in a population where the one-year incidence is 1.6 %, the incidence in the aspirin-exposed group will be 1.3 %, while in the control group it will be 1.9 % with a relative risk of 0.68 (0.013/0.019), which means that there is a relative risk reduction of 32 %. So far, there seems to be an association between the use of aspirin and the reduction of infarction, but the confidence interval of 95 % for such relative risk will have to be examined: if the interval within its limits (lower and upper) is below the unit, it is considered to be statistically significant, but if the upper value exceeds the unit (1), then it is not statistically significant and, therefore, the possibility that the observed point value of 0.68 is due to chance can not be ruled out. Similarly, when we talk about a risk factor, the lower limit of the 95 % confidence interval is expected to be above the unit (1) in order for it to be statistically significant (Table II).

Frequency, association and impact measures are based on the presence or not of an event or outcome and, therefore, these are nominal variables, but, in clinical practice, there are numerous outcome variables that are measured through the change in the value of a quantitative variable, in which there is equal interest in knowing the strength and magnitude of the association, and thus, it is important to have an equivalent.

Table III shows the relative frequency, association and impact measures in a global context, basically described for a nominal dependent variable. Other measures also applicable that can define the power of association are added —association measures—:

- The determination ratio r^2 , which measures the percentage of explanation of one variable based on the other and which is the square of the *r* obtained in a correlation, in this case the phi coefficient.
- The beta coefficient, which is the value obtained in a regression model (in this case logistic), which corresponds to the odds ratio logarithm.
- The R^2 similar to r^2 , whose result is obtained from the regression model.

As for the magnitude of association, the estimated probability of a phenomenon occurrence can be obtained from the result of a regression model $(y = 1/1 + e^{-(a + bIXI...)})$, which in the basis of the

equation for its calculation adds the beta coefficients of the different variables, and finally, calculates its global OR. With this equation, if two treatments are compared, the difference of such probability (difference of proportions) can be estimated, even if adjusted for multiple variables of interest; similarly, the different probabilities for a phenomenon to occur by exposure to different values of a quantitative variable can be compared.

The same table III shows when the dependent variable is quantitative: the units to measure the strength of association are limited to Pearson's r^2 , coefficient *b* and R^2 , the latter two as a result of the linear regression model.

Finally, to assess the association magnitude of a quantitative variable, the mean differences are used, more specifically the mean difference ratio, either directly estimated or as a result of the regression equation (in the linear regression, the value of the dependent variable is obtained directly).

A measure for the association magnitude that has become widely accepted is the number needed to *treat* (NNT = 1/RAR), which refers to the number of subjects that have to be treated in order to obtain the benefit in one when compared with placebo; when this number is negative, it is known as number needed to harm. Therefore, to define if a maneuver is clinically significant, a direct difference of 10 % can still be used or the number needed to treat (NNT), in which although there is no pre-established parameter, a value around 10 is considered ideal, which would represent treating 10 subjects to obtain the desired benefit in one (equivalent to 10%). It is worth mentioning that, generally, placebo is rarely used as the comparative group in clinical trials; therefore, this number may be underestimated when comparing it with other active maneuver.

Comments

Proper use of measures of frequency, association or impact and their equivalents is essential to avoid common errors committed in clinical practice. It is not uncommon to interpret association measures as if they were impact measures; for example, if the OR, RR or HR of a maneuver is 0.68, a 32 % reduction of the outcome is assumed. However, it should be considered that this is a relative reduction that the same can be generated by a 0.4/0.6 ratio than from a 0.04/0.06 or 0.00004/0.00006 ratio (RR = 0.66); nevertheless, in the first case, the NNT is 5, in the second 50, and in the third 50 000. Therefore, for estimating the impact of a maneuver, it is important that the direct difference or NNT (RAR) is available.

Bibliography

For online calculation

- 1. Cordell WH. Number needed to treat (NNT). Ann Emerg Med 1999;33:433-6.
- 2. Feinstein AR. Principles of medical statistics. New York, NY: Chapman and Hall/CRC; 2002.
- Guyatt GH, Sackett DL, Cook DJ. Users guides to the medical literature. II. How to use an article about therapy or prevention. B. What were the results and will they help in caring for my patients? Evidence Based Medicine Working Group. JAMA 1994;271:59-63.
- Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. BMJ. 1995; 310:452-4.
- KT Clearing House. [Website]. Odds ratio to NNT converter. Disponible en http://ktclearinghouse.ca/cebm/ practise/ca/calculators/ortonnt
- Sociedad Española de Hipertensión/Liga Española para la Lucha contra la Hipertensión Arterial. [Sitio web]. Odds ratio, riesgo relativo y número necesario a tratar. Available at http://www.seh-lelha.org/ oddsratio.htm
- University of British Columbia. [Sitio web]. UBC clinical significance calculator. Available at http://spph. ubc.ca/sites/healthcare/files/calc/clinsig.html

Twenty Years of Combining Clinical Practice and Clinical Research



Centro de Adiestramiento en Investigación Clínica (CAIC)

(Training Center for Clinical Research)

CENAIDS (Centro Nacional de Investigación Documental en Salud) Centro Médico Nacional Siglo XXI, Av. Cuauhtémoc 330 Col. Doctores Email: imss.caic@gmail.com



VII. Systematic Research: How to Locate Articles to Answer a Clinical Question

Rodolfo Rivas-Ruiz, Juan O. Talavera

In the process of solving doubts generated in the process of medical care, the amount of articles appearing during the search is so vast, that a strategy must be considered to refine it. The present article describes the process for searching and selecting information that may help us answering to our patients' needs. Judgment of the quality and relevance of the response will depend on each reader. The search has to be done in peerreviewed sites, and for that reason, we recommend PubMed and to start the search after breaking down the PICO acronym, where P = patients, I = intervention, C = comparator and O = outcome. The PICO acronym shares components with the classical research architecture model described by doctor Alvan R. Feinstein. A good search must be involved with the answer to our question in the first 20 articles; otherwise, the search will have to be more specific by using filters.

Key words

PubMed MeSH clinical research

This article was originally published in Rev Med Inst Med Seguro Soc 2012; 50 (1): 53-58 and it has been reviewed for this issue.

Introduction

In the process of solving doubts generated in the process of medical care, the number of articles appearing during the search is so vast, that we must consider a strategy, which in short time allows us to find those answers to our needs as physicians, so that we are not overwhelmed by an ocean of information. The present article describes the process to systematically search documents that help us answering our patients' needs, although the judgment on quality and relevance will depend on each reader.

Accessibility to medical information has changed with Internet and electronic media. Worldwide, there is an estimated 20 000 journals in the area, which provide approximately 2 million papers each year. This amount of articles, which represents new knowledge, generates great difficulties in keeping updated in every aspect of medicine.

The problem is aggravated by Internet postings on medical issues without peer-review, which depend on the good will of those who edit them and sometimes do not serve scientific purposes. Unfortunately, metabrowsers such as Google or Yahoo identify them easily, which results in these materials being highly consulted by patients and some doctors.

For these reasons, the search for medical literature must be performed in sites where publications are peer-reviewed and according to a system that avoids overseeing relevant articles and inclusion of unspecific articles to solve our questions. Hence, systematic search offers a clear, reproducible and auditable protocol.

The browser we recommend is PubMed, because it is simple, free and, most importantly, the manuscripts that appear are peer-reviewed by experts. Besides, recently it has included options to perform searches on mobile devices. This system is responsible for spreading the Medline database archives of the United States National Library of Medicine, which has over 21 million articles¹ (in areas such as genetics, medicine, nursing, psychology, veterinary medicine, among others), 90 % with an abstract in English; some magazines have links to the full-text article from this page. This medical library claims to be the largest in the world and has started integrating full-length articles, although free-access journals are still few.

Now, the first step in solving a question is to structure it properly based on the three items of the architectural approach outlined in previous chapters: *baseline state, maneuver* and *outcome*.² For an electronic search, an adaptation of Dr. Alvan R. Feinstein's architectural model has been proposed, in which the acronym PICO is formed, where P is

patients, with specification of the disease, if applies; I, the intervention or maneuver, treatment, risk factor, prognostic indicator and even a diagnostic procedure; C is the comparator, which may be a placebo group, another treatment or an observational maneuver; and, finally, the O for outcome corresponds to the result or outcome³ —this acronym may have some variations such as PEO (patients, exposure, outcome) or PICOST, where S and T represent the type of study and follow-up time—.⁴ Let's translate this into an example where a clinician wants to know if the use of albumin reduces mortality in patients with hypovolemic shock, compared with the use of

saline. With this proposal, the following acronym would be formed:

- P = patients with hypovolemia
- I = treatment with albumin
- C = saline
- O = mortality
 - With this acronym, the question would be:

Will the use of albumin (when comparing it with saline) reduce the mortality in patients with hypovolemia?

1

A tool that complements this method is the MeSH (Medical Subject Headings) acronym, a United States National Library of Medicine controlled vocabulary by means of which articles are indexed and organized in PubMed. These words enable having the definition of the subject that is being searched. Its catalog can be accessed from the PubMed main screen by selecting three options: the type of catalog (MeSH) (1), the word to be searched (2) and the Search button (3), as is shown in Figure 1.



Figure 1 Options in PubMed to search in the MeSH words catalog

Variants in the process

For novel terms, not recorded in the MeSH catalog or if the nomenclature under which a concept is recorded is unknown, text words or free words can be used, which will be identified anywhere within the articles: title, abstract or body of the article. The advantage is a wide search, with the risk or inconvenience that it may yield articles not directly related with the topic. Other drawback is that *text words* must be written directly in the search box together with the Boolean operator.



In our example, saline (saline solution) is not recorded as MeSH word; we used it for considering it to be widely used. It was entered as a text word (manually, together with its Boolean operator)





With the first PICO acronym term entered (2) (in our example hypovolemia), as shown in Figure 2, it will be necessary activating the check box (4) and pressing the *Add to search builder* option (5) to enter the term in the text box (6). Steps 2 to 6 must be repeated for entering other PICO acronym words, which will be linked to each other with ligands, which correspond to Boolean operators (7):

- *and* to link one or more criteria, which allows for more specific searches to be performed;
- *or* takes care of including one term or another, making the search broader.
- *not*, which is used to make total exclusion of the term that follows.

Let's see how our acronym words would combine if we added Boolean operators:

Will the use of albumin, compared with (AND) saline, reduce (AND) mortality in (AND) patients with hypovolemia?

As shown in the same figure 2, the PICO acronym words, the Boolean operators and, automatically, brackets, will be added in the search box (8), so that when we finish to include all terms in the system, the search will be recorded as follows:

(("Hypovolemia"[Mesh]) AND "Albumins"[Mesh] AND saline solution) AND ("Mortality" [Mesh] OR "mortality" [Subheading] OR "Hospital Mortality" [Mesh])

This is because when the terms are combined, the PubMed system includes brackets to perform the search following a similar logic to that of algebra notation, i.e., it solves first the inner parentheses and their results are combined with the external ones.

Once all the PICO acronym terms have been entered into the search box, all that is left to do is pressing the Search PubMed button (9).

The importance of previously constructing the PICO question lies in the fact that the order of terms entrance will be followed, which will allow for a search targeted to specifically find information related to our question.

A good search must succeed in finding the solution to our question in the first 20 articles (when there are studies). When no article is retrieved when searching for very rare diseases, the search must be done using only two or three terms or it must be expanded with the Boolean operator OR.

When in order to include a larger number of articles, two or more MeSH terms are selected in the same step, words are automatically linked with OR.

In our example, three *morta-lity* options were selected

5/4/3+1	Cett. Dimet		N
Dans Jetter	() former: 21 an juip	Bertha D	Fagilitati paretti bullor
	lanenne 1 gerinne is a genet personnen, a 1900		MT "Assesser" man and assesses and per att 1 Thread per att 2 Thread per att 3 "Assesses" (Assessment 3 "Assesses")
	and and we may design by Automy primes by Automotive by Automating for series processed hereby by a provide raw, and follow, buffacted, or any and an y		Mark Autom / MCC
0	netter De transming er sek enligt for and ef tinde tryft per e man til fengelationet propositionen		Salar Set 3

3

As shown in Figure 3, PubMed also has other resources for enhancing searches. One of them is Related citations (10), which generates an identification mechanism that displays the articles that most resemble the article selected in our list as the ideal, thereby extending the range of documents that we are able to consult. As we can observe, 10 articles were found in the example (11); when Related citations was used, 130 were retrieved (12). Another PubMed resource are the filters or limits (13), which can be accessed from the main browser.

Limits or filters are a useful system to limit the search to dates (14), type of article (clinical trial, cohort study, meta-analysis, clinical practice guideline) (15), species (humans and animals) (16), language (17), sex (18) and other parameters.

With these limits, more specific results are obtained, which is an essential issue when the number of identified articles is abundant (Figure 4).



Figure 3 Other search resources

notice	- B		-
Lähfe			
Palatet e tre can (Sig Sig	用 14		
C Once has 1 Electric Life 2 Mile Security 2 Partice Control	15	1 Stape 1 Faut 1 Garlage 2 Martin	17
- Seree	16	23 Male 33 Ferreir	18
11.400		La das un Direct	

Figure 4 Detail of the limits screen

If the user makes a typing error (typo) (1), the system leads to a screen where a warning is shown and terms that can replace or are related with the desired one are displayed (2). If the user activates the MeSH term (3), another screen will appear

		-
Remain Jamasa () Tarrenzo () and a Saman Saman Saman () and a Saman () and a Saman Saman ()	Public word to fee	-
	Annual and a	
Annual Shawa, Saturata Jestow A	failure liver 2	

Rivas Ruiz R et al. How to Locate Articles to Answer a Clinical Question

The new screen displays the definition of the term and related concepts in order for the user to verify if it is the desired one. He will be able to add it into the search box (4) with the Add to search builder option (5). To continue, all he has to do is entering the next PICO term (6)

(NSH) :: : :		- 6		
No. of Cases of the			(best so ??	A company of the second se
Appointents In elements, increases of the Tax elements of 1990	 44 4 di 1 ma mat 2 iu	particle and the DADA	4→	"Report Lang" (Section
the state of the second s				
Const.	integr.	T designed way	5→	int a service in the p

This same PICO acronym system can be used in meta-browsers such as Google or Yahoo as well. The words just have to be typed in English and linked with their Boolean terms, as shown in Figure 5. In Google, it is possible that more articles will be found than in PubMed and some that may be sponsored or not endorsed by peers. However, when the order of the PICO words is followed and the search is restricted to them, the result is often similar to that found in PubMed in complementary cases. In this example, we can see similar results to those obtained in PubMed, with the advantage that, in most cases, the full-text is available.

This electronic strategy shares the components of the classical research architecture model described by Dr. Alvan R. Feinstein in his book *Clinical Epidemiology*.⁶ This model was recently quoted by Julian P. T. Higgins and Sally Green in Chapter 6 of the *Cochrane Handbook for Systematic Reviews of Interventions*,⁵ and was employed by The Cochrane Collaboration for the elaboration of systematic reviews.⁷⁻⁸ This acronym has been used recently by the GRADE model as a search mechanism for the development of clinical practice guidelines.

Importantly, for more extensive searches, such as systematic reviews, other sources must be consulted in addition to PubMed, such as EMBASE, LILACS, Imbiomed, conference abstracts and even metabrowsers such as Google and Yahoo.

Google	Hyposiama AND Albumms AND same solution AND mortality
Rusqueda	Annual Contraction of Contraction Contraction
Tem Indigenes Urbans Desaus Inter Content de Affecteur, 201 Content de	Support to the set of

Figure 5 Meta-browsers respond to the PICO acronym with the advantage of including not only PubMed articles, but other local publications as well. They have the "disadvantage" that they identify a large number of results, which sometimes precludes a full enquiry



Figure 6 Usefulness of the classical research architecture model proposed by Dr. Alvan R. Feinstein

We consider that this search and clinical questions formulation mechanism, based on the architectural model and synthesized in the PICO acronym is one of the most useful in current clinical practice, since it is highly sensitive to the available electronic search engines, even in portable devices.

The advantage of the traditional scheme (Figure 6) is that it allows for the parts of a study, potential biases, statistical analysis, feasibility of the study or clinical significance to be identified, and forms the basis of electronic search.^{2,9-12}

anisms in hospitals might help considerably in the solution of clinical questions more quickly —with practice we estimate no more than 10 minutes and in increasing the certainty in prescription, in the selection of a diagnostic test or in the issue of a prognosis, thus facilitating medical education, peerwise discussion and the clinician's general work. As a complement to adequate reading and comprehension of articles, this approach might improve health care quality.

Disseminating and promoting these search mech-

References

- US National Library of Medicine, National Institutes of Health. Pubmed. http://www.ncbi.nlm.nih. gov/pubmed/
- Talavera JO, Wacher-Rodarte NH, Rivas-Ruiz R. Clinical research III. The causality studies. Rev Med Inst Mex Seguro Soc. 2011;49(3):289-94.
- Stone PW. Popping the (PICO) question in research and evidence based practice. Appl Nurs Res. 2002;15(3):197e-198e.
- Tricco A, Tetzlaff J, Moher D. The art and science of knowledge synthesis J. Clin epidemiol. 2011;64:11-20
- Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. [Updated March, 2011]. The Cochrane Collaboration; 2011.
- 6. Feinstein AR. Clinical epidemiology. The architecture of clinical research. US: WB Saunders; 1985.

- Egger M, Smith GD, Altman D. Systematic reviews in health care: meta-analysis in context. Second edition. London: BMJ; 2001.
- Khan K, Kunz R, Kleijnen J, Antes G. Systematic reviews to support evidence-based medicine. Second edition. London: Royal Society of Medicine; 2011.
- Talavera JO. Clinical research I. The importance of research design. Rev Med Inst Mex Seguro Soc. 2011;49(1): 53-8.
- Talavera JO, Rivas-Ruiz R. Clinical research IV. Relevancy of the statistical test chosen. Rev Med Inst Mex Seguro Soc. 2011;49(4):401-5.
- Talavera JO, Rivas-Ruiz R, Bernal-Rosales LP. Clinical research V. Sample size. Rev Med Inst Mex Seguro Soc. 2011;49(5):517-22.
- Talavera JO, Rivas-Ruiz R. Clinical research VI. Clinical relevance. Rev Med Inst Mex Seguro Soc. 2011;49(6):631-5.

VIII. Structured Review of an Article

Juan O. Talavera, Rodolfo Rivas-Ruiz

Several strategies have been attempted to select an article under assumptions of relevance and good quality. They depend largely on the presence or not of a series of features and in other occasions on the judgment of those who classify the article. However, these strategies do not allow for us to know the magnitude of error. Since there is no such thing as a perfect article, it is relevant to identify the magnitude of error and its impact on the final result; hence, it is necessary to develop skills that allow for us to review an article, identify possible errors and generate an idea of their impact on the result. According to the information contained in parts I to VII of this series of articles on clinical research, we have tried to demonstrate its application in a structured review of a causality article, starting with the examination of the baseline state, the maneuver and the result, with the systematic errors (biases) generated in each item, followed by the relevance of the test, the appropriateness of the sample size and, finally, clinical relevance.

Key words

journal article causality statistics and numerical data biases sample size association, exposition, risk or outcome measures

This article was originally published in Rev Med Inst Med Seguro Soc 2012; 50 (2): 163-166 and it has been reviewed for this issue.

Introduction

Several strategies have been attempted to select an article under assumptions of relevance and good quality. They depend largely on the presence or not of a series of characteristics and, in other occasions, on the judgment of those who classify the article. This entails a classification of "adequate" or "inadequate", or in best case to a graduation of major to minor quality or relevance. However, these strategies do not really allow for us to know the magnitude of error. And since there is no such thing as a perfect article, it is important to identify the magnitude of error and the impact it may have had on the final result; hence, it becomes necessary to develop skills that allow for us to review an article in a structured way, to identify possible errors and to generate an idea of their impact on the result. That is, we cannot rely on a classification or on the judgment of others to decide what to read and what not to read, or what to consider adequate or inadequate. We will have to learn the minimum basic structure that allows for us to assess ourselves the relevance of each article, its errors and its results.

In parts I and III to VI of this series on clinical research, we have tried to show the characteristics that we consider as being basic to perform a reading and a structured review of an article on causality (risk factor or etiologic agent, prognosis or treatment), once the article has been identified by means of a systematic search (topic addressed in part VII). We started with a model comprising the baseline state, the maneuver and the result (described in article I), with the systematic errors (biases) generated when defining and operating each of these items (article III). And we continued with the appropriateness of the test (part IV), the sample size estimation (part V) and, finally, the clinical relevance (part VI).

Next, we will make an exercise on the use of said information under a structured review proposal; for that, we will use an article of our own authorship: "Reduction in the incidence of post-stroke nosocomial pneumonia by using the 'Turn-Mob' Program", published in the Journal of Stroke and Cerebrovascular Diseases 2010;19:23-28. The purpose of the study was to demonstrate the efficacy of a program of mobilization in bed named "turn-mob" in decreasing the incidence of nosocomial pneumonia in patients with ischemic stroke.

In Figure 1, we can find baseline state characteristics such as the form of test selection and the prognostic demarcation; we can observe that randomization was able to balance the groups' characteristics, with the exception of chronic obstructive pulmonary disease, slightly higher in group b (14 %

Population selection method

Patient with acute neurological deficit,

> 12 hours duration referred from





Demarcation diagnosis	Prognostic stratif	ication: group a versus b					
< 48-hour evolution	Chronometric	72 and 74 years of age					
No requirement of ventilatory support	BMI status	Normal 18 versus 17%; overweight 69.4 versus 70.5%; Obesity 12.6 versus 12.5%					
	Clinical	Motor deficit, hemiparesis 66.7 versus 75.9 %					
First vascular event		Hemiplegia 33.3 versus 24.1 %; aphasia 50.5 versus 40.2 % Sensory deficit: 56.8 versus 40.2; nauseous reflex 82 versus 7					
No clinical evidence	Glasgow score 15, 40.5 versus 32.1 %						
of upper/lower RTI		NIHSS score 2-7, 30.6 versus 32.1 %					
No psychomotor agitation		8-13, 41.4 versus 43.8 % 14-18, 16.2 versus 17.9 % 19-23, 11.7 versus 6.3 %					
Ischemic stroke tomographic diagnosis	Morphologic	Cerebrovascular disease subtype Anterior circulation partial infarction 88.3 versus 90.2 %					
Those developing RTI in the first 48 hours	Comorbidity	DM 50.5 versus 42%; HBP 83 versus 84%; COPD 7 versus 14%; CVE 39 versus 40%					
were excluded	Previous treatment	Previous treatment Corticosteroids; antibiotic					
	Socioeconomic, cultural, habits = Smoking 31 versus 35 %; alcohol 24 versus 24 %						

RTI = respiratory tract infection; BMI = body mass index; DM = diabetes mellitus; HBP = high blood pressure, COPD = chronic obstructive pulmonary disease

Figure 1 Baseline state characteristics: diagnostic demarcation (selection criteria) and prognostic stratification (demarcation) (variables that impact on the outcome regardeless of the maneuver)

versus 7 %, p = 0.088), which could have impacted on the final result. Since a stratified analysis was not performed, is not possible to observe the impact of each maneuver according to different risk factors and thus, the result can be attributable mainly to the average characteristics of the population under study.

In Figure 2, the quality of the maneuver application (turn-mob program against usual position changes) has to be considered, verifying that peripheral maneuvers are implemented similarly in both groups.

Although there were no differences in peripheral maneuvers, the application of the turn-mob program was initially standardized and verified day by day; on the other hand, the application of the usual treatment was never standardized or verified and, therefore, there is no guarantee that it was carried out; furthermore, at hospital discharge, the patient did not receive nursing support at home. This could represent more than superiority for the turn-mob program over the usual treatment: the result of application of the turnmob program against nothing. Regarding the outcome, there was no possibility of having differentially detected the presence of nosocomial pneumonia, since all patients were submitted to chest X-ray at discharge or upon the slightest clinical suspicion. Similarly, there was no problem due to patient losses (transfer bias), since only two patients were excluded out of a total of 225 due to the presence of pneumonia within the first 48 hours of admission to the hospital (Figure 3).

General Comments

As an overall comment on the methodologic design and development of the project, we could say that the population selection was adequate (adequate assembly), by considering subjects with high probability of developing nosocomial pneumonia and in whom the application of the program turn-mob was feasible. The distribution of different prognostic factors was shown to be similar between groups, which



Figure 2 Characteristics to be considered during the application of the maneuver



Figure 3 Characteristics to be considered in the outcome

partially prevented susceptibility bias, since no stratified analysis was performed that would allow for the maneuver to be assessed in different risk groups (prognostic susceptibility). As for the maneuver, the adequate execution of the usual maneuver was not properly supervised and, therefore, we cannot guarantee that there was no performance bias. The outcome measure was the same in both groups, which prevented detection bias. Finally, we did not observe losses that could have reversed the observed difference in the outcome between groups (there was no transfer bias).

Regarding the test used (topic developed in Part IV of this series on clinical research), the chi-square shows the comparison of a nominal outcome variable between two groups, such as the presence or not of nosocomial pneumonia.

On the other hand, although the absence of a difference between the presence of diverse characteristics and the treatment group was demonstrated (chi-square test), a multivariate adjustment for the effect of the turn-mob program would have been attractive, due to the multiple characteristics of the baseline state and the co-maneuvers that could have impacted on the outcome. In this case, the multiple logistic regression test would have been appropriate, since the outcome nominal.

As for the sample size (addressed in part V), the method used for its calculation is not mentioned; however, we should remember that this calculation is performed in order to obtain the required number of patients to demonstrate that an expected difference between two groups is real and not by chance. In this case the observed clini-

Conclusions

cal difference of 12.6 % versus 26.8 % was statistically significant and thus, we can assume that it is real, since the probability of it being due by chance is lower than 5 % (p < 0.05). And even when the calculations are not described, with the incidence of 2 to 23 % mentioned in the introduction, we can estimate that the highest value was used and a direct reduction of about 15 % was considered, which yields a sample size between 90 and 103 subjects per group (Fleiss-Kelsey formula) and if we add 20 % to this, we obtain a value around the 225 subjects included in the study (sample size estimation for proportions difference).

Finally, in general, direct differences greater than 10 % or an NNT \approx 10 (CI-VI) were considered clinically relevant. In this case, the difference was 14.2 and the NNT consisted of 7.04 patients (which rounded is equivalent to 8) to see the benefit in one. With these results, we can clearly conclude that it is clinically relevant.

We cannot rule out the presence of a performance bias where the usual treatment would had not been carried out, in which case the conclusion would not be that the turn-mob program is better than usual mobilization performed by nursing staff, but rather it would have to be concluded that the program turn-mob in a post-ischemic stroke patient is better than no rotation or mobilization. On the other hand, we cannot identify whether the turn-mob program retains its benefit in different severity strata, since no stratified analysis was performed and no adjustment was made through a multivariate analysis; probably, these analyses were not performed due to the sample size, since 44 nosocomial pneumonia cases are insufficient when stratifying or adjusting. As we can see, every study has errors and yet, there is valuable information; however, to weigh it, is essential to have some notion on clinical research.

Bibliography

- Talavera JO. Clinical research I. The importance of the research design. Rev Med Inst Mex Seguro Soc. 2011;49(1): 53-8.
- Talavera JO, Wacher-Rodarte NH, Rivas-Ruiz R. Clinical research II. Studying the process (the diagnosis test). Rev Med Inst Mex Seguro Soc. 2011;49(2):163-70.
- Talavera JO, Wacher-Rodarte NH, Rivas-Ruiz R. Clinical research III. The causality studies. Rev Med Inst Mex Seguro Soc. 2011;49(3):289-94.
- Talavera JO, Rivas-Ruiz R. Clinical research IV. Relevancy of the statistical test chosen. Rev Med Inst Mex Seguro Soc. 2011;49(4):401-5.
- Talavera JO, Rivas-Ruiz R. Clinical research V. Sample size. Rev Med Inst Mex Seguro Soc. 2011;49(5):517-22.
- Talavera JO, Rivas-Ruiz R. Clinical research VI. Clinical relevance. Rev Med Inst Mex Seguro Soc. 2011;49(6):631-5.

IX. From Clinical Judgment to Clinical Trial

Juan O. Talavera, Rodolfo Rivas-Ruiz

Two strategies are described, intended to understand causality and documenting it with the best evidence: the clinical judgment and clinical trial. In the first one, the baseline state, the maneuver and the outcome are identified, each one with characteristics showing the complexity of the causality phenomenon, whose control allows for systematic errors to be prevented: in the baseline state, inadequate assembly and susceptibility bias; during the application of the maneuver, the performance bias; in the outcome measurement, detection and transfer biases. In the clinical trial, the tactics that try to isolate the effect of the principal maneuver from that of other components of the causality phenomenon --previously described in the clinical judgment section- are mentioned. For that purpose, the opportunity for the maneuver to be manipulated, and the temporary nature of the causal relationship are used. Its characteristics include allocation and blinding of the maneuver, feasibility of its early interruption, the analysis according to the adherence to the maneuver, the groups to be compared, the transient nature of the comparative maneuver and the informed consent. When the physician applies this knowledge in a conscious and structured manner with his/her patient, he/she improves his/her efficiency and brings medical practice closer to clinical research.

Key words

clinical trial bias

This article was originally published in Rev Med Inst Med Seguro Soc 2012; 50 (3): 267-272 and it has been reviewed for this issue.

In "Clinical Research III" of this series, clinical reasoning (clinical judgment) was addressed as a logical model to explain the phenomenon of causality, which was previously described by Dr. Alvan Feinstein in his books *Clinical Biostatistics* and *Clinical Epidemiology. The Architecture of Clinical Research.* According to Dr. Feinstein, every sensible physician should know this reasoning. We dare saying that not only every sensible doctor knows it, but applies it during his/her clinical practice as well. However, sometimes doctors fail to do it consciously and, consequently, in a structured way. Similarly, in number I of this series, research designs were mentioned as a strategy to obtain evidence of such causality. Among them, clinical trials provide the highest quality evidence.

The present article shows these two strategies for explaining and documenting the phenomenon of causality and tries to show them in parallel, in such a way that based on one, the reason for the other is easily understood:

- Clinical judgment, or clinical reasoning/architecture of clinical research, as a phenomenological description of clinical research.
- Clinical trial, as the design that offers the highest quality of information during the clinical research process, by attempting to control or at least to document the involvement of every component within the causality phenomenon.

Clinical Judgment

In order to explain the causality phenomenon, the baseline state, the maneuver, and the result (and its characteristics) are described, as well as five sources of systematic error that can arise if they are overlooked: two in the baseline state, one during the execution or measurement of the maneuver and two in the outcome.

Sources of Error in the Baseline State (Figure 1)

- a) Inadequate assembly. Usually occurs when components of the diagnostic demarcation are omitted. It is defined by the population place of origin, the diagnostic criteria and the selection criteria.
- *b) Prognostic susceptibility bias.* Generally observed when the prognostic stratification is omitted. In it, all the factors present at the baseline state that may impact on the outcome must be considered, regardless of the effect of the maneuver.

Sources of Error in the Maneuver (Figure 2)

a) Performance bias. Usually occurs when the different components are not considered in order



Figure 1 Characteristics to be considered in the basal state to prevent an inadequate assembly and susceptibility bias

for the maneuver to have optimum power and, therefore, the quality of the maneuver turns out being deficient; it also occurs when those actions accompanying it before, during or afterwards are not considered, and which are known as comaneuvers or peripheral maneuvers. In addition, the comparability of the maneuver has to be specified (efficacy, effectiveness and efficiency), as well as the multiplicity of maneuvers and the temporary concurrence of the comparative maneuver.

Sources of Error in the Outcome (Figure 3)

- *a) Detection bias.* Uneven identification of the outcome, either by diagnostic suspicion or uneven number of outcome assessments between groups.
- b) Transfer bias. Patients lost to follow-up not due to random effects. The 20 % sample size increase does not solve the problem when losses are associated with the maneuver; it simply maintains data stability in order for the power of the test to be preserved during the statistical analysis.



Figure 2 Characteristics to be considered in the maneuver to prevent performance bias



Figure 3 Characteristics that have to be considered in the outcome to prevent diagnostic detection and transfer bias

Clinical Trial

Clinical trials allow for information to be obtained with such quality that it attempts to isolate the result provoqued by the principal maneuver on the baseline state and controls for components that may participate in the outcome or provoque a biased assessment of it. Clinical trials, unlike observational studies, allow for the maneuver to be manipulated, which confers distinctive characteristics to it.

Among the characteristics accompanying the maneuver, either in an immediately previous period, during or in a subsequent period, the following are exclusive of clinical trials (Figure 4):



- Maneuver assignment: is the distinctive characteristic between the clinical trial and other designs, since only the clinical trial offers the opportunity for the maneuver to be assigned. Random assignment of the maneuver attempts to generate groups with similar baseline conditions between the different maneuvers (to avoid prognostic susceptibility bias), thereby preventing discrepancies that might later be the cause of outcome differences. Even though this is a highly popular strategy, it does not prevent the presence of the "trans-stratification" phenomenon, nor does it specify the impact of the maneuver on different prognosis strata (see "Clinical Research III"). This phenomenon can be prevented if a randomization by strata is performed, provided the analysis of the results is carried out within each stratum and not just globally. Similarly, randomization has other functions such as compliance with the ethical principle of offering each individual the same opportunity of receiving the experimental maneuver, and the possibility that subjects with similar maneuver adherence probabilities are assigned to each treatment arm (to avoid performance bias) and similar probability of dropping out from the study (which reduces the transfer bias). Finally, it is worth mentioning that randomization facilitates the blinding of the maneuver. This is how the maneuver reduces the probability of biases that are distinctive of the baseline state, the maneuver and the outcome.
- Blinding of the maneuver: this strategy seeks primarily to prevent the involvement of subjectivity in the assessment of the outcome (in order to avoid detection bias). It is subdivided in three categories depending on who does not know the treatment maneuver within the research process:
- a) Single-blind: this is considered when the patient ignores which treatment he/she is receiving, i.e., doesn't know to which maneuver he/she was assigned.
- *b) Double-blind:* when the patient and the investigator do not know the treatment arm.
- *c) Triple-blind:* when the patient, the investigator and the one who analyzes the data do not know the treatment arm.

In addition to this, when the form of delivering a drug is different (e.g., drug a is administered twicedaily and drug b thrice-daily; or drug a is orally administered and drug b intramuscularly), or when the physical appearence of the drug is different (drug a, blue pill; drug b, yellow) a double simulation is used (double-dummy); for example, if the patient receives drug a only twice a day and drug b three times a day, three drug b placebos will have to be added to drug a, which have to be taken the same way three times daily and two drug a placebos will have to be added to drug b, which have to be taken twice daily.

- Early interruption: clinical trials may be interrupted for two inherent reasons to the treatment: early difference between groups in the primary outcome, provided there is no probability of such differences to be lost once the sample or the follow-up are completed; and due to the presence of adverse events, above the upper limit of the 95 % confidence interval, estimated according to the corresponding sample size or follow-up period.
- Analysis according to adherence to the maneuver: hardly a clinical trial with a follow-up period exceeding a few days ends with an adherence of all participants to the maneuver of at least 80 % (e.g., taking the drug at 80 % of the doses). In general, non-adherent patients are expected to be similar in number and characteristics —at baseline and in peripheral maneuvers— between treatment groups; similarly, subjects lacking adherence are expected to have similar characteristics to those reaching the end of the study with adequate adherence. Thus, assuming a random lack of adherence between groups, data are analyzed using two strategies:
- a) Intention to treat (ITT) analysis, which is characterized for including in the outcome assessment both those subjects who complied with an adequate adherence (≥ 80 %) and those who did not (< 80 % adherence).
- b) Per-protocol analysis, when the decision consists in including in the analysis only data from subjects with $a \ge 80$ % adherence.

In the intent-to-treat analysis, a decrease in differences between treatment groups is usually observed, whereas in the per-protocol analysis, that what could be the real difference between the maneuvers is usually preserved, provided losses have been random; otherwise, one of the groups might end up being favoured (let's imagine that those subjects with more adverse events are not adherent and that these are differentials between the maneuvers, or that the subjects with better or worse response to the treatment are not adherent and that the response was also differential between the groups; if this occured, performance bias would be present).

Other non-exclusive characteristics to clinical trials, since they can also be considered in observational studies, include the following (Figure 4):



Figure 5 Clinical trial characteristics in parallel to clinical reasoning

- Groups to be compared. It is important assessing which is the comparator of the principal maneuver, since depending on this, clinical trials have been classified in efficacy, effectiveness and efficiency studies:
- *a) Efficacy*: when the active maneuver is compared against placebo or against nothing. This comparison tries to demonstrate that the active maneuver works better than doing nothing or just giving a placebo
- b) Effectiveness: represents the comparison of the active maneuver with a standard treatment; therefore, it tries to demonstrate the superiority of a maneuver against another. This study must be weighed carefully, since not finding any significant differences does not mean that the maneuvers are equal or equivalent. If that what is sought is to demonstrate equivalence, the sample size will have to be estimated for a maximum difference of about 3 %. If that what is looked for is non-inferiority, a maximum difference of 9 % will have to be considered.
- *c) Efficiency:* it refers not to a comparison, but to the impact of the maneuver once it is applied in the community.
- Transient nature of the comparative maneuver. In most cases, clinical trials comparing two or more maneuvers have the virtue of doing so within a time schedule and, consequently, with simultaneous (in parallel) application of the maneuver. Other different comparison modality are the crossover studies, where the maneuvers to be compared are carried out on successive periods and alternately in each one of the subjects under study; the big advantage is that the subjects to be compared are the same and, therefore, the remaining variables outside the principal maneuver are identical; however, these studies have some problems, such as: 1) the carry-over effect, in which when introducing the second maneuver, the subject's basal conditions have changed by the action of the first, or 2) when the disease has changed by itself during the period of time of application between the first and the second maneuver. On the other hand, this type of design is typical in stable pathologies with minimum changes expected in the scheduled study period (where after removing the first maneuver the subject actually returns to the previous baseline state) and in cyclical pathologies (whose behavior is practically the same at each cycle).

When comparing different maneuvers at the same time or at very close periods, diagnostic conditions of the pathology under study are expected to be similar, and the possibility of accessing to peripheral maneuvers to be alike; in this way, the possibility that the differences between therapies are not due to differences in diagnosis (susceptibility bias) or in accessibility to peripheral maneuvers (performance bias), or in diagnostic criteria (inadequate assembly), or in outcome assessment criteria (detection bias) is avoided. Finally, we should mention that in a clinical trial, the baseline conditions and followup time of subjects included and randomized to one therapy or another is the same.

 Informed consent. Since in all cases the maneuver will be assigned, even if it entails a minimal risk, ethical principles of research in human beings must be protected. (Therefore, the principles that must be considered to safeguard the rights and wellbeing of patients participating in research projects will be highlighted.)

Conclusions

Identifying and mentally organizing the details of the causality phenomenon during the clinical course of a disease, and knowing the reasons of the distinctive characteristics of a clinical trial, allows for the bond of clinical practice with clinical research to be understood and, consequently, it facilitates a reasoned and structured bidirectional exploitation of both for the benefit of patients. It is important to note that, as mentioned by Dr. A. Feinstein, the people more used to the handling of causality is the clinician, since everytime he assigns a maneuver to a patient he/she is applying this knowledge and skills, and that doing it in a conscious and structured way, undoubtedly will improve his efficiency and will bring medical practice closer to clinical research (Figure 5).

Recommended readings

Feinstein AR. Clinical Biostatics. Saint Louis: The CU Mosby Co; 1977.

- Feinstein AR. Clinical epidemiology. The architecture of clinical research. Philadelphia: WB Saunders; 1985.
- Feinstein AR. Directionality and scientific inference. J Clin Epidemiol. 1989;42(9):829-33.
- Portney LG, Watkins MP. Foundations of clinical research: applications to practice. Third edition. New Jersey: Pearson-Prentice Hall; 2009.

Rothman KJ, Greenland S, Lash TL. Modern epidemiology. Baltimore: Williams & Wilkins; 2008.

Sackett D, Haynes R, Tugwell P. Epidemiología clínica una ciencia básica para la medicina clínica. Madrid: Ediciones Díaz de Santos; 1989.

X. From Clinical Judgment to Cohort Design

Juan O. Talavera, Rodolfo Rivas-Ruiz

After the clinical trial, the second research design with the best quality of information is the cohort. Although the possibility of randomization of the maneuver is not available, there is the opportunity of having the subjects followed over time. Any research that tries to explain the causality fenomenon is at risk of incurring biases; however, the cohort studies distinctive features try to avoid them. Its main characteristics are: 1 being observational, situation where the investigator only measures the presence of the maneuver, which is a characteristic that divides the subjects into exposed and non-exposed; 2 being longitudinal, which offers the opportunity to follow the subject over time, documenting the time-sequence of appearance of the causality phenomenon components; 3 measurements have directionality, which generates the existence of prolective, retrolective and retro-prolective cohorts (the first are the ones with the highest quality, since they perform a real-time measurement of the variables of interest; 4 being comparative.

Key words

cohort studies follow-up studies longitudinal studies prospective studies retrospective studies

This article was originally published in Rev Med Inst Med Seguro Soc 2012; 50 (4): 383-388 and it has been reviewed for this issue.

The cohort study is characterized for the followup of a group of subjects with similar characteristics over time. After the clinical trial, this is the second research design with the highest quality in the collection of information. Although there is no assignment of the maneuver that characterizes the clinical trial, there is the opportunity of having the subjects followed over time and, consequently, with the consistency of having the maneuver measured before the onset of the outcome (observational maneuver, since it is not assigned by the investigator —also known as "measuring the exposure"—).

It is important to mention that any research study that attempts to explain the phenomenon of causality is at risk of generating biases, either when defining the baseline state (by inadequate assembly and susceptibility bias), during the maneuver (performance bias) or when measuring the outcome (detection bias and transfer bias), as shown in Figures 1a, 1b and 1c, previously described in "Clinical Research III" and "Clinical Research IX" from this same series. However, the characteristics of the cohort studies try to avoid them.

Main Characteristics (Table I)

Exposure to the Maneuver

This is an observational study and, hence, the researcher is able only to measure the exposure to the maneuver, unlike the clinical trial, where the investigator assigns it. It should be mentioned that, although the clinical trial is the ideal design for assessing a therapeutic maneuver, its assessment by means of observational studies such as cohort studies is currently accepted (the effect of a drug prescribed by someone other than the investigator can be assessed, for example, phase IV trials). It even happens to be the ideal model when trying to assess a maneuver that cannot be assigned by the investigator due to ethical issues.

It is important to mention that the maneuver divides the cohort into the groups to be compared; at their baseline state, the subjects comprise the cohort as a single group sharing similar characteristics and, with the principal maneuver, they are distributed into *exposed* and *unexposed*. The effect of the main variable on the baseline state to generate the outcome shall be estimated, always adjusting for confounders that may be present at the baseline state (inadequate assembly and susceptibility bias) or during the action of the principal maneuver (performance bias). In a clinical trial, random assignment of the maneuver



Figure 1a Characteristics that have to be considered in order to prevent an inadequate assembly and susceptibility bias



Figure 1b Characteristics that have to be considered in order to prevent performance bias

tries to control the confounding variables, a possibility that does not exist in the cohort design; hence, possible confounding variables should be thoroughly measured.

Subject Follow-up

The second and most important feature of this design is its *longitudinal* nature, i.e., there is a follow-up of the subject under study, with the variable(s) of interest being measured over time, so that change (e.g., glucose values) or the appearance of the variable of interest (e.g., infarction, death, adverse event) can be documented. During the follow-up of the cohort, there is the possibility of including subjects in a similar moment within the clinical course of their condition —generally at the beginning, which is known as an *inception cohort* and homogeneously following them during a previously established period, either until the end of the follow-up period or until the outcome. In these cases, the study is known as a *closed cohort* study, characterized by having similar follow-up periods (Figure 2a). In contrast, there is the *open* or *dynamic cohort*, when the inclusion and exit of study subjects at different points during the clinical course of the disease is accepted, with follow-up periods being heterogeneous in this case (Figure 2b).



Figure 1c Characteristics that have to be to considered in order to prevent detection and transfer bias

Due to the follow-up of the study subjects, there is a possibility for execution bias to occur if the maneuver is not homogeneous and constant within each group and upon heterogeneous peripheral maneuvers between groups. Moreover, being a design that involves following subjects over time, the possibility of losing them is elevated, which provokes a transfer bias. Finally, it should be mentioned that particularly in dynamic cohorts, inadequate assembly or susceptibility bias can be induced when including subjects with less or more likelihood of suffering the outcome; for example, when only survivors are included in periods subsequent to the baseline (*survivor cohort*).

Directionality in Measurements

The third characteristic of cohort design is the directionality in the measurement of information, which results in what we know as prolective cohort study (prospective), historical cohort or retrolective cohort (retrospective) and the ambispective or retro-prolective cohort (retro-prospective) (Figure 3).

The prospective or prolective cohort is characterized by the measurement of baseline, follow-up where the maneuver is included and outcome characteristics in real-time and under previously established standards, which provides high quality to the collection of such information and, therefore, the assessment of the impact of the principal maneuver on the baseline state in order to generate an outcome is highly accurate.

In the measurement of the main maneuver and other variables involved in the phenomenon of causality (confounding variables), there are multiple possibilities likely to be generated, such as measurement using criteria as specific as desired or measuring the degree of exposure to it, either at baseline state or during the follow-up —simulating adherence in

Table I Characteristics of the cohort design							
Design	Exp/Obs	Long/Trans	Prol/Retrol	Comp/Desc	Measure		
Cohort	Observational	Longitudinal	Prol/Retrol/Rp	Comparative	Incidence		

The methodological approach considers four features: 1. Imposition or not of the maneuver for investigational purposes: experimental (Exp) or observational (Obs) study, respectively. 2. Patient follow-up (Long) or not (Trans) over time. 3. Directionality in the collection of information: prolective (Prol), retrolective (Retrol) and retro-prolective (RP). 4. Search or not of association between two or more variables: comparative (Comp) and descriptive (Desc), respectively. Measurement of outcome occurrence (Measure), either through incidence, prevalence, or simply the case-control ratio



Figure 2b Open or dynamic cohort design

the case of the principal maneuver (which prevents performance bias)—. Prediction and measurement of possible maneuvers that may lead to confusion allow for adjustments to be made, either at the baseline state (thereby avoiding susceptibility bias) or during the execution of peripheral maneuvers (in order to avoid performance bias). Finally, objective, specific and homogeneous measurement of the absence of the outcome at the baseline state and the occurrence thereof during follow-up or at study termination prevents an inadequate assembly at the beginning (when the outcome was already present in an early form at the beginning of the study) and subsequently, the detection bias.

In order to simulate the blinding of the maneuver, typical only of clinical trials, in the cohort study the measurement of variables at the baseline state is expected to have been performed by staff that is independent to those who assess the exposure to the maneuver and, in turn, that both these are independent from those assessing the outcome. The advantages offered by early planning of events within the causality phenomenon are only characteristic of prolective cohort studies and clinical trials. Thus, among observational studies, the prolective cohort is the model with the highest quality in the collection of ideal data for assessing causality.

The historical or retrolective cohort does not allow for the maneuver impact to be measured with the same accuracy as the prolective cohort, since no variable is measured in real-time in any of the components described in the architectural design -reasoning or clinical judgment-. In the historical cohort, the population selected to be assessed has already been exposed to the variable of interest and has already suffered or not the outcome, with the followup period having concluded. However, although no component can be measured in real-time, there must be specific criteria for each variable to be measured, but own and expectable in a routine clinical record. During the planning of the study, the researchers must have specified criteria for each variable to be measured and strategies to improve the quality of the information. One of these consists in fragmenting the clinical record into three sections: one that corresponds to the baseline state, other to the exposure



Directionality in the collection of information

a = Prolective cohort: all variables, either from the baseline state, from exposure to the maneuver or from the outcome, are measured in realtime. b = Retrolective cohort: measurement is performed when the follow-up period is over and the outcome has happened; consequently, exposure to the maneuver, baseline conditions and outcome are not measured in real-time. c = Retro-prolective cohort, is a combination: basal conditions already occurred, exposure to the maneuver has occurred entirely or for a partial period, but the outcome has not yet occurred and, therefore, it is measured in real-time

Figure 3 Type of cohort according to directionality of variables

to the maneuver and other to the measurement of the outcome, so that each block of information can be reviewed independently (similar to that described in the prospective cohort). Although this strategy has the great disadvantage that some of the information may not be found in the clinical record or its quality may be questionable, the historical cohort shows what happens in real practice; therefore, when assessing a therapeutic maneuver, the result is closer to that what will happen once it is applied in the population, unlike to what happens with the clinical trial or the prolective cohort, without the effect of surveilance and thoroughness in measurements or follow-up of the subject.

Search for Association

The fourth characteristic of the cohort design is the search for association. Actually, at present few descriptive studies are performed; however, every study describes the characteristics of its population in the first paragraph of the results. The cohort is a comparative study, either because it compares the study subjects' exposure with different maneuvers or with the change or appearence of some characteristic over time.

Comments

It is important to emphasize at what moment the assembly of the population occurs in cohort design, since it is one of the characteristics that clearly differentiates this study from other observational designs. In the cohort, the population enters at the baseline state, regardless of the directionality of measurements. For instance, if we are dealing with a prospective cohort of patients with type 2 diabetes mellitus and we want to follow them for 10 years, every newly-diagnosed patient with the disease in a specific population who meets the selection criteria will be able to enter and will be followed for 10 years, with variables being measured in real-time. But if we have a ret-

rolective cohort (historical), every patient belonging to the population of interest that 10 years ago or more was diagnosed with type 2 diabetes mellitus, and that at that time fulfilled the selection criteria, will be able to enter and will be followed in his/her records from that time until the follow-up time is covered or until the onset of the outcome; clearly, in that case variables will not be measured in real-time.

Recommended readings

Feinstein AR. Clinical biostatics. Saint Louis: Mosby; 1977.

- Feinstein AR. Clinical epidemiology. The architecture of clinical research. Philadelphia: WB Sanders; 1985.
- Feinstein AR. Directionality and scientific inference. J Clin Epidemiol. 1989;42:829-33.
- Portney LG, Watkins MP. Foundations of clinical research: applications to practice. Third edition. New Jersey: Pearson-Prentice Hall; 2009.
- Rothman KJ, Greenland S, Lash TL. Modern epidemiology. Third edition. Baltimore: Williams & Wilkins; 2008.
- Talavera JO, Wacher-Rodarte NH, Rivas-Ruiz R. Clinical research III. The causality studies. Rev Med Inst Mex Seguro Soc. 2011;49(3):289-94.
- Talavera JO. Clinical research I. The importance of research design. Rev Med Inst Mex Seguro Soc. 2011;49 (1):53-8.

XI. From Clinical Judgment to Case-control Design

Juan O. Talavera, Rodolfo Rivas-Ruiz

The case-control design, just as the historic cohort, is loaded with a series of potential biases resulting from reconstructing the facts once the outcome has occurred, in addition to biases generated by the selection of the control group. It is characterized by having a series of cases for which a comparative group (controls) is identified. That is, it goes from the outcome to the cause and, consequently, facts must be reconstructed in the opposite sense as to the way the causality phenomenon occurs. Nevertheless, architectural design will have to be borne in mind and in each section —baseline state, maneuver and outcome— those features necessary to demonstrate the effect of the maneuver will have to be considered, thus preventing an inadequate assembly and the susceptibility, performance and detection biases. Transfer bias can only be controlled by having a defined population, either based on general population or nested in a cohort. When a defined population is not available, this design is recommended only for rare diseases.

Key words case-control studies

clinical trial

This article was originally published in Rev Med Inst Med Seguro Soc 2012; 50 (5): 505-510 and it has been reviewed for this issue.

Ithough the case-control study is apparently a simple design for solving questions, it is without any doubt the most complex. Like the historical cohort, it is loaded with a series of potential biases resulting from the reconstruction of the events preceding the outcome, in addition to the biases in the selection of the control group. Therefore, this design should be considered only in cases where answering the clinical question through a clinical trial or a cohort study is not possible.

The collection of the information required to document the causality phenomenon —described under the concept of research architecture or clinical judgment (Figures 1, 2 and 3)— is carried out, in ideal conditions, by means of a clinical trial, whose most important characteristic is the assignment of the maneuver (experimental). When this design is not possible, the cohort is used, which preserves the opportunity of following the study population over time, with the possibility for the maneuver to be documented before the outcome occurs (longitudinal). However, the case-control design will have to be considered if the uncommonness of the phenomenon being analyzed, the difficulty to complete the sample size or the relevant use of resources, force to do so.

This design is characterized by having a series of cases for which a control group (comparative group) is identified. Unlike the clinical trial and the cohort study —where the maneuver is assigned (experimental) or identified before the outcome (observational) and a follow-up is conducted until its assessment (longitudi-nal)—, the case-control study tries to reconstruct the effect of the maneuver once the outcome has occurred (for the cases) or its absence documented (control group) (Figure 4). That is, it starts from the outcome and the information is reconstructed in the direction of the probable cause (figure 5); this design requires for the facts to be reconstructed in the opposite sense as to the way the phenomenon of causality occurs.

Main Characteristics

Case-control design has limits in documenting information, which are similar to those in historical cohort studies (Table 1) and, as a consequence, biases are similar.

Exposure to the Maneuver

This is an observational study that only measures the exposure to maneuver. Unlike cohort studies, the maneuver here does not divide the subjects in two groups (in the cohort, exposed and unexposed), but identification of exposure is part of the fact of being a



Figure 1 Characteristics that have to be considered in order to prevent an inadequate assembly and susceptibility bias

case or a control, which causes that within each one of these groups (cases or controls) a subgroup is generated of exposed and unexposed subjects (Figure 5). Documenting the effect of the principal maneuver in case-controls studies —conversely to what happens in clinical trials, where baseline conditions and comaneuvers are controlled and the principal maneuver is randomly assigned— implies recording all possible confounding variables present at the baseline state (susceptibility bias) or how do co-maneuvers participate (performance bias).

Subject Follow-up

Some authors consider case-control studies to be longitudinal when records exist prior to the outcome, both for cases and for controls. However, it is difficult for this to happen, except for vaccine records, which are kept in the entire population, or when the study is performed in a cohort; in these situations, the quality of evidence will be higher, since exposure measurements will be known before the outcome appears.

In most cases, the reconstruction is made using interviews, whereby the record of what happened with the exposition and the outcome is simultaneous (transversal). This way of getting information is common when the control group members are related to the cases or when they agree to participate in the trial by telephone or Internet; this can even happen with hospital controls, although in these, information can occasionally be reconstructed longitudinally if previous records are available. Obtaining information in a



Figure 2 Characteristics that have to be considered in order to prevent performance bias


Figure 3 Characteristics that have to be to considered in order to prevent detection and transfer bias

cross-sectional form may produce biases due to poor data quality in all components of the causality phenomenon (baseline state, maneuver, outcome), commonly due to differential recall between the cases group and the control group members.

Directionality in Measurements

The case-control design is retrolective (retrospective). Unlike historical cohort —which is also retrolective, but whose population assembly is made based on the baseline state—, population assembly is made on the basis of the outcome (either case or control). That is, at best, the quality of information depends not only on its previous collection with purposes other than the objective of interest (e.g. the vaccination record was not designed thinking on further evaluating its association with any pathology and, similarly, a lot of confounding variables were ommited), but also transfer biases in a cohort of survivors (in a population defined according to the baseline state, it is possible to include both alive and dead cases and alive and dead controls).

Search for Association

The search for a control group for a series of cases is always carried out attempting to establish associations.





Figure 4 Case-control studies. Case identification and control selection



The presence or not of exposition to the factor of interest is documented. Starting from the outcome, probable cause is tried to be identified

Figure 5 Case-control study. Exposure documentation

Selection of the Control Group

Selecting the control group is the most difficult process in this type of design, and it can induce bias in all sections of the causality phenomenon, especially transfer bias.

Usually, the members of the cases group are selected among patients that in spite of being cared for in the same medical unit, they come from different geographical areas. They are pre-selected patients: in theory, they looked for medical care for different reasons; then, they had to be assessed by at least one doctor before reaching the hospital; in addition, they have to agree to participate or not in the trial and meet a series of selection criteria. Thus, it is difficult to define which population they come from or whom they represent.

Defined Population

If the population where the cases come from is known and, in turn, it is clearly defined, the biggest difficulty

Table I Main characteristics of the case-control studies										
Design	Observational/ Experimental	Longitudinal/ Transversal	Prolective/ Retrolective/ Retro-prolective	Comparative/ Descriptive	Measure					
Cohort	Observational	Longitudinal	Prol/Retrol/Rp	Comparative	Incidence					
Case-control	Observational	Long/Cross	Retrolective	Comparative	Case/control ratio					

The methodological approach considers four features: 1. Imposition or not of the maneuver for investigational purposes: experimental or observational study. 2. Patient follow-up (longitudinal) or not (cross-sectional) over time. 3. Directionality in the collection of information: prolective, retrolective and retro-prolective. 4. Search or not for association of two or more variables: comparation or description. Measurement of outcome occurrence is determined by incidence, prevalence or case-control ratio

of the study design is solved. This happens when the case-control study is population-based or when it analyzes a group nested in a cohort. In both situations, the total population where the cases come from is available and, evidently, this is where the controls will be selected from. It is even possible to determine which group the deaths (if any) correspond to. When the number of subjects in the population exceeds the size calculated for the sample, it is also possible to make a random selection of cases, as well as of controls.

Given that generally in cohort studies the information of the population under analysis is documented ---which was measured before the ocurrence of the outcome that will be examined in the case-control study-, errors are avoided in the documentation of such information. Cohortnested case-control studies have additional characteristics: they usually are restricted to the analysis of elements of interest obtained during the initial assessment of the cohort (which would correspond to the baseline state from the case-control study), instead of addressing elements of the total cohort. This way, only the subjects who have developed the outcome and a control group are examined. This allows for resources to be optimized and to preserve the elements under study in the rest of subjects in the cohort (blood samples, tissues, etc.).

Undefined Population or from a Secondary Source

Since it is common for a defined population not to be available, there are different strategies to obtain control subjects likely to belong to the same population of the cases. The most usual is to include neighbors or friends of the cases, individuals invited by telephone or Internet (previously identified as coming from the same geographic region as the cases) and, in other occasions, hospital-based controls. Whichever the situation, usually there is a sub- or over-representation of the exposure that will alter the results.

Phenomenological Reconstruction of the Facts

Facts must be reconstructed according to the causality phenomenon, regardless of their own limitations on how the population is assembled (from outcome to exposure) and how the data are collected (retrolectively and transversally). For this, a series of recommendations exist:

• To clearly establish the criteria for integrating the population to be studied, applicable both to cases

and controls (Figure 1). The questioning or search for information on records has to be transferred to the period that for each case or control would correspond to the baseline state, and the following should be attempted for the entire population:

- *a*) Restrict as much as possible the scope of the research only to subjects belonging to the same region.
- *b)* Define the diagnostic criteria, i.e., the population to be analyzed.
- c) Define the selection criteria, i.e., requirements to be met by subjects in which the outcome has not occurred or, if the interest is to assess its progression rather than its manifestation, in those in which it still is incipient. Although this might sound obvious, care should be taken to avoid that these criteria do not include subjects with indication or contraindication for the maneuver, but do include those in which the outcome is likely to occur. It is important to remember that the baseline state, even in the group of cases, must be free of the outcome. In fact, criteria are equal for both.
- Document all baseline state variables that are likely to modify the effect of the maneuver on the outcome, or that regardless of the maneuver contribute to the onset of the outcome (Figure 1).
- Clearly define the exposure and, if possible, graduate it for magnitude and time, as well as for all possible co-maneuvers (Figure 2).
- Specify the criteria defining the case and the control.
- Try to select recently diagnosed cases, in order to ensure that the exposition to the maneuver has not been modified after the diagnosis.
- Determine which will be the documentation sources to obtain data for the cases. These must be the same as for controls (figure 3).
- Standardize the way to reconstruct the information for both cases and controls, whether based on previously obtained data or by means of questioning. It would be erroneus obtaining the information for the cases from the record and for controls by means of questioning.
- Assign the tasks of facts reconstruction to different people. Ideally, those who obtain the baseline state information should have no contact with those documenting the exposure to the maneuver and, in turn, both should be different of those who document the outcome.
- Obtain the information in the order at which the causality phenomenon occurs (baseline state, maneuver and outcome).

Comments

Without a doubt, in addition to the mentioned errors, the reconstruction of events based on the outcome entails transfer biases, since in cases and controls only survivors are usually assessed.

It is advisable to avoid the case-control design as a strategy to document the causality phenomenon when the answer can be obtained by means of a clinical trial or a cohort. What this design has in common with the other research designs is that it is only a tool to document the causality phenomenon; therefore, the most important suggestion is to always maintain the mental structure of clinical judgment, by means of which three well-known elements are conceptualized: a baseline state where the distinctive characteristics of a group of subjects lead to their distribution in sub-groups according to their likelihood to suffer the outcome even before the exposure to any maneuver (prognostic demarcation); a principal maneuver with characteristics of its own, accompanied by a series of actions around it (co-maneuvers); and measurement of the changes in the baseline condition or the onset of new characteristics, known as the outcome.

That phenomenological structure, usual for clinicians —clinical judgment/research architecture— is universal and is not modified by the way the information is obtained, either in a clinical trial or an observational study. When performing a structured evaluation of an article or when trying to answer a question by means of a research study, the causality phenomenon should always be thought of from the clinical point of view.

Recommended readings

Feinstein AR. Clinical biostatics. Saint Louis: Mosby; 1977.

- Feinstein AR. Clinical epidemiology. The architecture of clinical research. Philadelphia: WB Sanders; 1985.
- Feinstein AR. Directionality and scientific inference. J Clin Epidemiol. 1989;42:829-33.
- Rothman KJ, Greenland S, Lash TL. Modern epidemiology. Third edition. Baltimore: Williams & Wilkins; 2008.
- Portney LG, Watkins MP. Foundations of clinical research: applications to practice. Third edition. New Jersey: Pearson-Prentice Hall; 2009.
- Talavera JO, Wacher-Rodarte NH, Rivas-Ruiz R. Clinical research III. The causality studies. Rev Med Inst Mex Seguro Soc. 2011;49(3):289-94.
- Talavera JO. Clinical research I. The importance of research design. Rev Med Inst Mex Seguro Soc. 2011;49 (1):53-8.
- Cruz-Anguiano V, Talavera JO, Vázquez L, Antonio A, Castellanos A, Lezana MA, *et al.* The importance of quality of care in perinatal mortality: a casecontrol study in Chiapas, Mexico. Arch Med Res. 2004;35(6):554-62.

XII. From Clinical Judgment to Cross-sectional Survey

Juan O. Talavera, Rodolfo Rivas-Ruiz

A longitudinal study, whether it is a clinical trial or a cohort study, has the virtue of following the logical sequence in which the components of the causality phenomenon occur. However, in a cross-sectional study, this logical sequence does not exist and it is consistent with the measurement of the three components (baseline state, maneuver and result). "Clinical judgment" helps us to artificially reconstruct these components in the time sequence in which they occurred. However, the way in which the population is assembled and how the information is obtained —cross-sectionalwise and retrolectively— entails the risk of producing biases. Although the use of cross-sectional survey in order to associate a maneuver with a probable outcome is difficult and often generates errors (especially when pathological phenomena are studied), it is extraordinary to show the development of a healthy subject simulating a longitudinal study, as is happens when height and weight are estimated according to age; this type of design has been named "longitudinal cross-sectional study".

Key words

cross-sectional studies cohort studies case-control studies clinical trial

This article was originally published in Rev Med Inst Med Seguro Soc 2012; 50 (6): 641-644 and it has been reviewed for this issue.

longitudinal study, whether it is a clinical trial or a cohort study, has the virtue of following the logical sequence in which a phenomenon occurs (at a baseline state, the effect of a maneuver to generate an outcome is observed). In contrast, in a crosssectional study, this logical sequence does not exist, since at the moment of measurement the three components coincide: baseline state, maneuver and result.

Architectural design (clinical judgment) helps us to artificially reconstruct the components in the timesequence they occurred. This way, in cross-sectional designs we can even make causality assessments, knowing full well the limitations and risks (Figures 1 to 3). Cross-sectional designs include the case-control study and the cross-sectional survey.

The cross-sectional survey is probably the most widely used design in medical research. In general, except for the analysis of therapeutic maneuvers (in which the clinical trial design is generally used), most causality studies use the cross-sectional survey and only sometimes the cohort design, which is complex and costly due to the large population that must be followed during extended periods.

Cross-sectional survey is characterized for studying a specific population or a sample of such population with data being collected at the same time. That is, the information on the baseline state, the maneuver and the outcome is obtained retrospectively; when the analysis begins, the outcome and the exposure to the maneuver have already happened. Thus, it is not possible to observe the study subject's baseline conditions and their change over time. However, according to the phenomenon of causality logical sequence, it is assumed that the outcome did not exist before the maneuver was applied. So, the intensity and length of exposure to the maneuver can also be reconstructed in order to establish the magnitude of its association with the outcome. Although all the components of the causality phenomenon are measured at one time, the reconstruction of facts should be made following the logical time-sequence (Figure 4).

Exposure to the Maneuver

In cross-sectional survey only the exposure to the maneuver is measured, unlike the clinical trial, where the investigator assigns the maneuver. And unlike the historical cohort, where exposure to the maneuver has already been measured, even though with purposes other than research, in the cross-sectional survey, as in the case-control study, the quality of the maneuver measurement is low. The status of the patient, at the moment of measurement, influences on the accuracy of data (whereby the effect or knowledge of the



Figure 1 Characteristics that have to be considered in the baseline state in order to prevent an inadequate assembly and susceptibility bias

outcome has some impact) and its distance from the components of the causality phenomenon (the longer the time since the exposure to the maneuver, the less accurate the information). The same happens with the measurement of variables that may confound the effect of the maneuver —conditions previous to the maneuver (baseline state) and conditions accompanying the maneuver in its time (peripheral maneuver)— (Figure 2).

Subject Follow-Up

When the observation of the causality phenomenon components agrees with their time sequence (baseline

state, maneuver and outcome), it allows for a series of errors to be predicted and prevented; however, this only happens in clinical trials and the cohort design. In the cross-sectional survey, the assessment of all components is simultaneous —which characterizes it as a cross-sectional study— and the time sequence is artificially reconstructed, but at the risk of placing the maneuver ahead of the outcome or measuring an assumed maneuver that in reality is a consequence of the outcome or a characteristic acompanying the outcome (in a diabetic patient, for example, attributing hypertriglyceridemia to uncontrolled glycemia, when both can be a consequence of other factors).

Although associating an outcome to a probable cause is difficult and errors are frequently generated,



Figure 2 Characteristics that have to be considered during the maneuver in order to avoid performance bias



Figure 3 Characteristics that have to be considered during the outcome measurement in order to prevent detection and transfer bias



Figure 4 Artificial reconstruction of the causality phenomenon in the crosssectional survey

cross-sectional survey design is extraordinary for knowing the development of a healthy subject. The height and weight charts for children according to age and sex are an example. These charts were made with cross-sectional measurements of children of each gender and different ages; subsequently, a cohort was simulated where the boy or girl's size and weight changed according to life-years. This design is known as longitudinal cross-sectional study and is suitable for showing the development of the healthy subject, but does not allow for the natural history or clinical course of a disease to be known, since sicker subjects are lost over time and subsequent assessments only include survivors, which renders for false results of the disease evolution to be obtained. However, this design may be useful in diseases with low mortality, as long as the potential effect of the outcome on the measurement of preceding characteristics is controlled.

Directionality in Measurements

Measurement of all the components of the causality phenomenon at the same time is influenced by the fact that exposure to the maneuver has occurred previously on certain baseline conditions, same as the outcome; i.e., measurements directionality turns the cross-sectional survey into a retrolective (retrospective) study. Unlike the historical cohort (or retrolective cohort) —whose measurements directionality makes it also retrolective in nature—, where the record of facts was made sequentially as they went occurring longitudinally, although for reasons other than research, the reconstruction of facts in the cross-sectional survey is made at the same time, in such way that the temporary nature and magnitude of exposure to the maneuver and co-maneuvers, as well as the baseline conditions —those preceding the maneuver— will depend, most of the time, on the memory of the subject under study, which affects the accuracy of data and attributions of causality due possible biases in the baseline state, the maneuver and the outcome (Figure 4).

Search for Association

The search for causality will always imply comparing regardless of the design. Similarly, cross-sectional survey involves comparing the effect of the maneuver of interest on the baseline state, against its absence or against the effect of other maneuvers.

Phenomenological Recreation of Facts

Being a cross-sectional and retrolective study, recommendations are provided in order to reconstruct the facts as close as possible to the phenomenon of causality.

The process of gathering information should always begin with that what would correspond to the baseline state, specifically with the selection criteria, which must be the same for the entire study population. Similarly, at the baseline state, all the characteristics that might influence on the outcome should be

Recommended readings

- Feinstein AR. Clinical biostatics. Saint Louis: The CU Mosby Co; 1977.
- Feinstein AR. Clinical epidemiology. The architecture of clinical research. Philadelphia: WB Sanders; 1985.
- Feinstein AR. Directionality and scientific inference. J Clin Epidemiol. 1989;42:829-33.
- Rothman KJ, Greenland S, Lash TL. Modern epidemiology. Third edition. Baltimore: Williams & Wilkins; 2008.
- Portney LG, Watkins MP. Foundations of clinical research: applications to practice. Third edition. New Jersey: Pearson-Prentice Hall; 2009.
- Talavera JO, Wacher-Rodarte NH, Rivas-Ruiz R. Clinical research III. The causality studies. Rev Med Inst Mex Seguro Soc. 2011;49(3):289-94.
- Talavera JO. Clinical research I. The importance of research design. Rev Med Inst Mex Seguro Soc. 2011;49(1): 53-8.

documented, regardless of the maneuver or by interaction with it.

The characteristics of the maneuver and comaneuver should be defined as far as possible, as well as those of the outcome.

It is necessary to try that among the subjects in whom the outcome of interest has occurred only those recently diagnosed are included, in order for the effect of the principal maneuver to be assessed on it and to reduce the probability of the outcome modifying what the maneuver could have been.

It is essential to take care that the structure where information is obtained is always the same and not to favor any tendency, in order for the subjects' responses not to be biased.

Finally, the collection of information should be segmented, starting with the baseline conditions, continuing with the maneuver, and finishing with the outcome.

Comments

Even when cross-sectional designs (case-control and cross-sectional survey) are somewhat uncomfortable, much of the research used to solve the patients' ailments comes from studies with these designs. Although the actual structure of the phenomenon of causality and the reconstruction of its components in the cross-sectional survey are artificial, they are logical and necessary when using clinical judgment.

XIII. Research Design in the Structured Review of an Article

Juan O. Talavera, Rodolfo Rivas-Ruiz

The quality of information obtained according to the research design is integrated to the structured review in accordance with the causality model. For example, it is used in the article "Reduction in the incidence of post-stroke nosocomial pneumonia by using the 'Turn-mob' Program", whose design corresponds to a clinical trial. The aspects that have to be identified and analyzed include ethical issues, which are intended to safeguard the safety and respect for the patient; the random assignment, intended to generate groups with homogeneous baseline conditions, comprised by subjects with the same probability of receiving any of the maneuvers being compared and with the same pre-maneuver likelihood of adherence to them and the same chances of dropping out from the study for causes other than the maneuver. Other aspects include the relativity of the comparison, the blinding of the maneuver, the application in parallel of the comparative maneuver, the early termination and the analysis according to the degree of adherence. The analysis according to research design is supplementary to that performed on the basis of the causality architectural model and statistical and clinical relevance considerations

Key words

research design clinical trial causality bias

This article was originally published in Rev Med Inst Med Seguro Soc 2012; 51 (1): 68-72 and it has been reviewed for this issue.

his text integrates the structured review of an article (Figures 1 to 3 from part VIII of this series), the characteristics of the research design and the resulting quality of the obtained information (parts IX and XII, also from this series).

We will use again the article "Reduction in the incidence of post-stroke nosocomial pneumonia by using the 'Turn-mob' Program" (published in J Stroke Cerebrovasc Dis. 2010;19:23-8), which aimed to demonstrate the efficacy of a mobilization program in bed in order to decrease the incidence of nosocomial pneumonia in patients with ischemic stroke. The research design used was the clinical trial; therefore, we will analyze its characteristics (Figure 4) and integrate them to the example based on the causality architectural approach described by doctor Alvan R. Feinstein.

Design Characteristics. Clinical Trial

Ethical Aspect

Although the first aspect that has to be analyzed is the ethical one, in view of its extension and distinct nature, it will be discussed in other article.

Randon Assignment

An element that defines the clinical trial is the *random assignment*. This is intended to generate groups with homogeneous baseline conditions in order to avoid susceptibility bias; to integrate in the groups subjects with the same probability of receiving any of the maneuvers being compared, and with the same pre-maneuver likelihood of adherence to them in order to avoid performance bias; to facilitate the blinding in the assessment of the outcome and, consequently, to reduce the diagnostic detection bias. Randomization also distributes the subjects between the groups with the same probability of dropping out from the study for causes other than the maneuver, thereby reducing transfer bias.

As for the Turn-mob program, it was randomly assigned and achieved balanced groups at the baseline state, except for chronic pulmonary obstructive disease, which could have favored the experimental maneuver. Thanks to randomization, groups were generated with the same likelihood of adherence to the maneuver, although in this study, adherence to the standard maneuver was never verified, whereby it is possible that it was total absence of mobility of the patient. As for the assessment of the outcome, it is not specified if it was performed by a second assessor without any knowledge of the group the patient



Figure 1 Characteristics that have to be considered at baseline state: diagnostic demarcation (scope of research, stroke definition, selection criteria) and prognostic stratification (variables that impact on the outcome regardeless of the maneuver). In the Turn-mob program, although randomization was able to balance groups characteristics, except for chronic obstructive pulmonary disease (COPD) —discretely higher in group *b* (14 versus 7 %, p = 0.088) and may impact on the final result—, it is not possible to observe the effect of each one of the maneuvers depending on different risk factors and, thus, the result must be attributed mainly to average characteristics of the population

belonged to. Finally, no losses are observed that might have caused transfer bias.

Relativity of the Comparison

Although the Turn-mob program was planned as an effectiveness study by comparing the new against the standard maneuver, it could have turned out to be an efficacy analysis since the possibility exists for the comparative maneuver to be precisely not applying any action.

Blinding

Blinding of the maneuver was impossible in the Turnmob program and, although a second assessor of the outcome could have been promoted, this is not mentioned. Therefore, there was the likelihood of diagnostic detection bias.

Parallel Comparative Maneuver

The requirement of performing a comparative maneuver in parallel (during the same calendar days) was covered and was met by preventing differences in the diagnostic or stratification demarcations (in order to avoid inadequate assembly and prognostic susceptibility biases), differences in accessibility to peripheral maneuvers (to avoid performance bias) and differences in outcome diagnosis criteria (which reduces the possibility of detection bias).

Early Termination

There was no presence of adverse events due to the maneuvers. Nor were there early differences in the outcome. Should events or differences have been present, these might have stopped the Turn-mob program.



Figure 2 Characteristics that have to be considered during the application of the maneuver: quality of application of the principal maneuver (Turn-mob compared with usual position changes) and verifying that peripheral maneuvers are applied similarly in both groups. Although there was no difference in peripheral maneuvers, the application of the Turn-mob program was inicially standardized and verified day by day. Conversely, usual treatment was never standardized or its aplication verified on a daily basis; therefore, there is no guarantee that it was carried out; furthermore, when the patient was discharged to home, nursing support ceased to exist. This could represent more the result of applying the program against no action than superiority of the Turn-mob program over the usual treatment

Analysis According to Adherence

Final Comments

The last aspect is the analysis according to adherence, which shows clearly that the Turn-mob program was carried out in the intent-to-treat modality, since all patients were assessed in each one of the groups they were assigned, regardless of whether in the group with the standard maneuver they received it or not, as it could have been the case, with the consequent performance bias. As we can observe, the analysis of a research article or work according to the design used is complementary to the analysis made on the basis of the causality architectural model; on the other hand, statistical and clinical relevance considerations will have to be taken into account. Without any doubt, the performance of a structured analysis requires time and knowledge and with no doubt it is more enrichening than just



Figure 3 Characteristics that have to be considered in the outcome: there is no possibility of having diferentially detected nosocomial pneumonia, since all patients underwent chest X-rays at discharge or upon the slightest clinical suspicion. Similarly, there is no problen due to patient losses; only 2 cases were excluded out of a total of 225 and due to the presence of pneumonia within the first 48 hours of hospital admission



Figure 4 Clinical trial characteristics in paralell to clinical reasoning

accepting a foreign and superficial quality judgment, as it is pretended in the classification by level of evidence. On the other hand, keep in mind that although every article specifically tries to answer one question, it happens to contain a large amount of useful information for the clinician, such as epidemiological and clinical aspects of the pathology under study.

References

Talavera JO.	Clinical	research I	The	importance	of the	research	design.	Rev
Med Ins	t Mex S	eguro Soc.	2011	;49(1):53-8.				

- Talavera JO, Wacher-Rodarte NH, Rivas-Ruiz R. Clinical research III. The causality studies. Rev Med Inst Mex Seguro Soc. 2011;49(3):289-94.
- Feinstein AR. Clinical epidemiology. The architecture of clinical research. Philadelphia: WB Saunders; 1985.
- Feinstein AR. Clinical biostatistics. Washington: C.V.Mosby; 1977.
- Rothman KJ, Greenland S, Lash TL. Modern Epidemiology. Third ed. Baltimore: Williams & Wilkins; 2008.
- Portney LG, Watkins MP. Foundations of clinical research: applications to practice. Third edition. Pearson/Prentice Hall; 2009.
- Talavera JO, Rivas-Ruiz R. Clinical research VIII.Structured review of an article. Rev Med Inst Mex Seguro Soc. 2012;50(2):163-6.
- Talavera JO, Rivas-Ruiz R. Clinical research IX.From the clinical judgment to the clinical trial. Rev Med Inst Mex Seguro Soc. 2012;50(3):267-72.









Internal medicine specialist, dedicated to teaching and to clinical research. He was born in Mexico City in 1965. Since 2010, he is part of the Centro de Adiestramiento en Investigación Clínica (CAIC) (Training Center for Clinical Research), which is located at the Centro Médico Nacional Siglo XXI (CMN SXXI), and belongs to the Coordinación de Investigación en Salud (Health Research Coordination) from the Instituto Mexicano del Seguro Social (IMSS).

jotalaverap@uaemex.mx

Rodolfo Rivas Ruiz

Pediatric neonatologist in the Hospital de Pediatría, CMN SXXI. IMSS. He holds a master in **Clinical Research** from the Universidad Autónoma del Estado de México (UAEM), and is candidate for a doctorate in Clinical Epidemiology at the Universidad Nacional Autónoma de México (UNAM). Member of the Sistema Nacional de Investigadores (SNI). Professor in the School of Medicine at UNAM. and in the master in Clinical Research IMSS-Instituto Politécnico Nacional (IPN). He also belongs to the CAIC.

rivasrodolfo@gmail.com

Niels H. Wacher Rodarte

He is an internal medicine specialist, and holds a master in Medical Sciences (both, from UNAM). He is the Head of the Clinical **Epidemiology Research** Unit at the Hospital de Especialidades (CMN SXXI). Professor of Clinical Epidemiology, and Evidence-based Medicine (UNAM). He also teaches the master and doctorate programs in Clinical Epidemiology, and the course Clinimetría (Clinimetry) in the Master in Health Sciences (UNAM). He is a member of the SNI.

nwacher@hotmail.com



Carla Martínez Castuera Gómez

She has a bachelor's degree in International Relations, and a master's degree in North-American Studies both from Universidad de las Américas-Puebla. Currently. she is the Head of the Área de Vinculación Internacional (Department of International Relations), which belongs to the Coordinación de Investigación en Salud (Health Research Coordination), IMSS. She is responsible for the support programs and the management for international cooperation. She has experience in the public sector, and collaborates with the CAIC.

cmartinez.imss@gmail.com



Nutriologist from the Escuela de Dietética y Nutrición (School of Dietetics and Nutrition) of the Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado (ISSSTE). She has a master in Clinical Epidemiology from UNAM, and is currently studying the Doctorate in Clinical Epidemiology (also at UNAM). She teaches Classical Epidemiology in the master in Clinical Research IMSS-IPN. and belongs to the CAIC.

marxelapr@gmail.com

Lino Palacios Cruz

General and adolescent psychiatrist from the Instituto Nacional de Psiguiatría Ramón de la Fuente Muñiz (INPRFM)-UNAM. He has the degree of master in Medical Sciences from INPRFM-UNAM. Currently, he is studying the Doctorate in Health Sciences-Clinical Epidemiology (UNAM). He is professor of several curses of Psychiatry (INPRFM-UNAM), of the Master in Clinical Epidemiology (UNAM), and of the master in Clinical Research (IMSS-IPN). He belongs to the SNI, and collaborates with the CAIC team.

palacioslino@gmail.com



She holds a bachelor's degree in Nutrition from the Universidad Iberoamericana. She has a master's and a doctor's degree both in Health Sciences from UNAM. She has worked in the Instituto Nacional de Ciencias Médicas v Nutrición Salvador Zubirán (INCMNSZ) (Salvador Zubirán National Center of Medical Sciences and Nutrition); in the CMN SXXI, IMSS; and in several private institutions. Currently, she has her own practice, and is the Director of the Licenciatura en Nutrición (bachelor's in Nutrition) at the Universidad Tecnológica de México (UNITEC). She collaborates with the CAIC. lau.pao.br@gmail.com