Basics of publishing and grant activities in science

SCIENTOMETRY AND ETHICS IN SCIENCE

prof. Zdeněk Dvořák, DSc., PhD.
Department of Cell Biology and Genetics

MEASUREMENT OF SCIENTIFIC OUTPUT

FINANCIAL ASPECT (INVESTMENT RECOVERY)

- research is financed from diverse sources (private, public etc.)
- provider of finances for research defines/requires evaluation criteria
- relatively easy in task-based research areas: Tasks were achieved or not? (e.g. „construct engine with reduced fuel consumption", „generate genetically modified corn resistant to parasites")
- difficult in „basic research" – the impact has broad societal aspects in long-term perspective, unpredictable – grey zone?
- the majority of pivotal discoveries was fully appreciated and exploited few decades after their publication

IMPACT ON RESEARCHER´S CAREER

- scientific research has important personal dimension – team composition, hierarchy in academic and research organizations, leading personalities, qualification for academic/scientific grades and titles (prof., PhD, DSc.), membership in scientific and expert boards and committees.

…to be a scientist, it is not anymore a mission, but it has became regular occupation….(→ having social, professional, economic dimensions)

THE NEED FOR EVALUATION/MEASUREMENT OF SCIENTIFIC OUTPUTS

MEASUREMENT OF SCIENTIFIC OUTPUT

- outputs/efficiency may be measured in discrete period of time
- historical aspects, boom of communication and other technologies
- can we compare scientific outputs of following researchers?

Isaac Newton × Albert Einstein × Kary Mullis

- a great challenge for scientometry is to compare outputs in different and highly diverse subjects of research (areas)

SOCIAL SCIENCES & HUMANITIES	(french linguistic)
LIFE SCIENCES	(ecotoxicology)
PHYSICAL SCIENCE	(nuclear engeneering)
HEALTH SCIENCE	(oncology)

SCIENTOMETRY

Definition:

Scientometry is a scientific discipline, which studies the evolution of science using quantitative indicators of scientific information such as number of publications in scientific journals, number of citations for articles or authors…

***Sensu stricto* – in non-scientific sense, scientometry is considered as a tool for „more objective" evaluation of scientific outputs/efficacy of individual scientists as well as „quality" of scientific research as whole**

- **Eugene Garfield – founder of scientometry (mid-20th century)**
- **He noticed that expert systems based on indexed scientific data and citations may serve as a tool for evaluation of Science evolution itself**
- **one of the fathers of ISI Web of Knowledge**

ISI Web of Knowledge: http://apps.isiknowledge.com

Comprise several scientific databases:

- Web Of Science
- Journal Citation Reports® – metrics for scientific journals
- Scientific Web Plus
- ISI HighlyCited℠ – nejcitovanější vědci
- Biology Browser ®
- ResearcherID.com
- Science Watch
- Thomspon Scientific

WEB OF SCIENCE:

CITATION DATABASES (5)

- Science Citation Index Expanded (SCI-EXPANDED)--1945-present
- Social Sciences Citation Index (SSCI)--1977-present
- Arts & Humanities Citation Index (A&HCI)--1977-present
- Conference Proceedings Citation Index- Science (CPCI-S)--1990-present
- Conference Proceedings Citation Index- Social Science & Humanities (CPCI-SSH)--1990-present

CHEMICAL DATABASES (2)

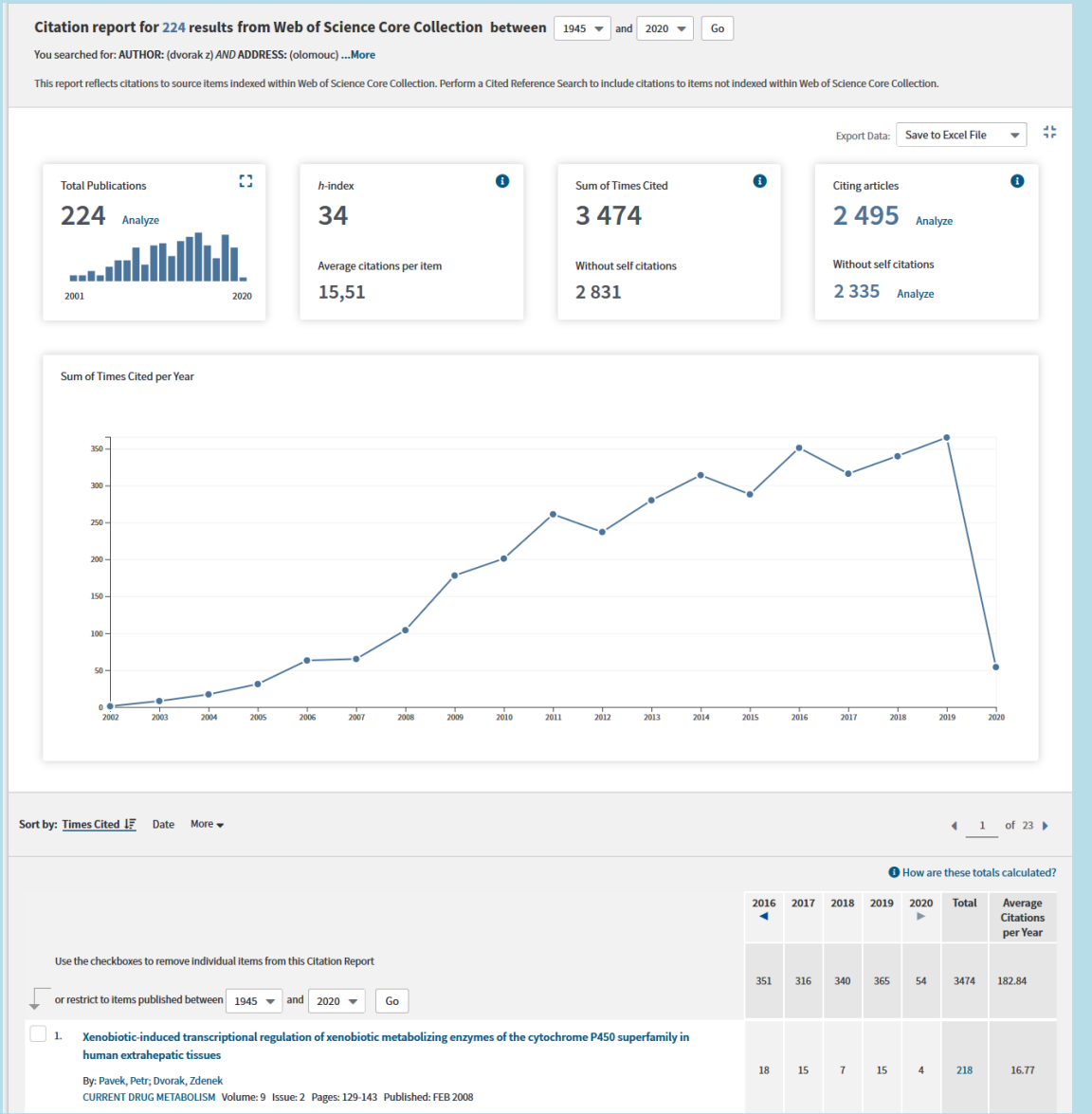
- Index Chemicus (IC)--1993-present
- Current Chemical Reactions (CCR-EXPANDED)--1986-present

Web Of Science

- searching criteria - author, laboratory, institute, journal….
- number of publications – type (article, review, letter, proceedings, meeting abstract)
- list of publications
- analysis of results

Citation report

- a number of citations in individual years
- total number of citations
- exclusion of autocitations



SCI – Science Citation Index

- number of citations in certain period of time – 1 year, 5 years, life-time
- highly cited paper is probably highly desired, read and interesting
- SCI is therefore a measure of scientific efficacy

HOWEVER: Summary SCI does not express several aspects appropriately:

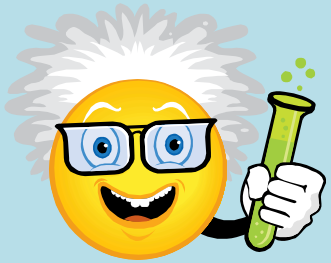
1. Highly cited articles may be cited due to their incorrectness of errors they contain; therefore, high SCI is not a measure of quality!!!

2. SCI, in principle, it increases in time, with the age of scientist. Incomparable for post-doc (5 years of experience) *versus* senior researcher (35 years of experience)

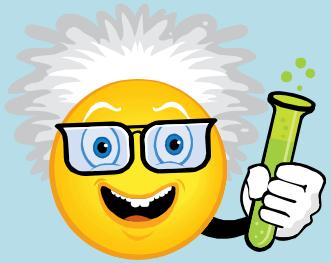
3. SCI is unable to reveal about homogenity and dynamics of citations:

a) Time aspect – at the beginning of career extensive SCI, then decreasing in time, but it is not distiguished by summary SCI
b) Age aspect – citations may be acquired even when researcher is not scientifically active or passed away…
c) Homogenity – out of 50 articles, 49 is cited 100 x a 1 is cited 900 x In average, each article was cited 20 x ?

SCI A AGE OF RESEARCHER



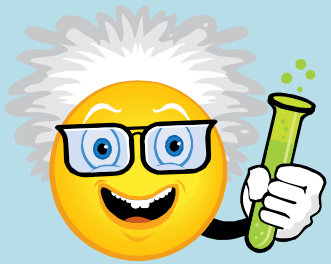
- age 55 years; 30 years of career
- N = 50 publications; SCI = 150



- age 30 years; 5 years of career
- N = 25 publications; SCI = 50

- when we compare N and SCI of both researchers, R#1 is much more efficient than R#2 (…scientific board decides that R#1 will be appointed as professor, because of high „quantity" of scientific outputs…") – but is really R#1 better than R#2?
- R#2 produced 25 publications in 5 years, i.e. with steady dynamics in the career, R#2 will have 150 publications at the age of 55, which is 3-fold of those by R#1
- R#1 has 5 citations per year per 50 publications; with steady trend, R#2 will have at the same age with 10 citations per year per 150 publications cca 1500 citations – 10 x more than R#1!!!

SCI AND FLOW OF TIME



Number of citations acquired each year

	2000	2001	2002	2003	2004	2005	2006	2007	SUM
#1	100	140	160	90	40	20	10	10	**570**
#2	10	10	40	50	80	100	140	160	**570**

- both researchers have 20 y. of scientific practice and their SCI = 570
- R#1 was highly successful at the beginning of the career, but in last few years the efficiency/impact of the publications dissipated
- R#2 systematically grows in their career and cumulative SCI also grows

HIRSCH INDEX: h-index

An index to quantify an individual's scientific research output

J. E. Hirsch*

Department of Physics, University of California at San Diego, La Jolla, CA 92093-0319

Communicated by Manuel Cardona, Max Planck Institute for Solid State Research, Stuttgart, Germany, September 1, 2005 (received for review August 15, 2005)

I propose the index h, defined as the number of papers with citation number $\geq h$, as a useful index to characterize the scientific output of a researcher.

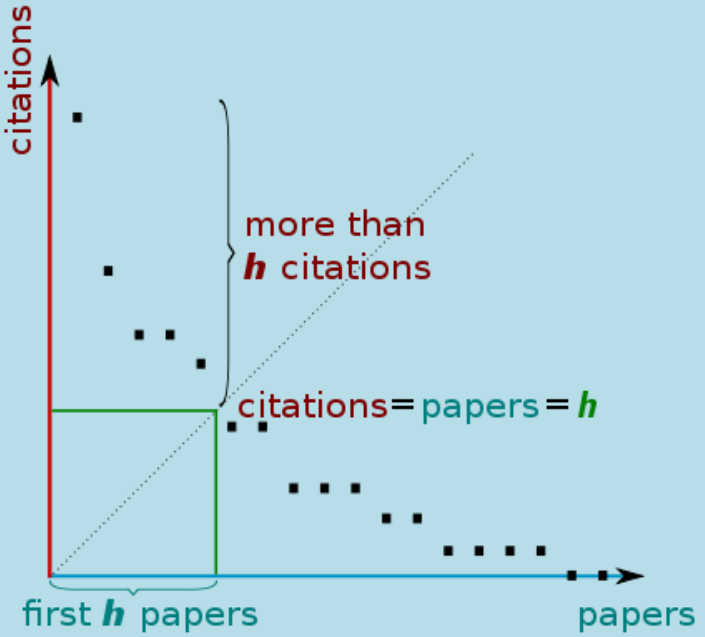
(*i*) Total number of papers (N_p). Advantage: measures productivity. Disadvantage: does not measure importance or impact of papers.

Proceedings of the National Academy of Sciences of the United States of America 102 (46): 16569-16572 November 15 2005

„A scientist has index *h* if *h* of his/her N_p papers have at least *h* citations each, and the other (N_p – *h*) papers have no more than *h* citations each."

h-index:

- **combines the measure of productivity and citation index**
- **eliminates disproportion between isolated highly cited publications and publications without citations**

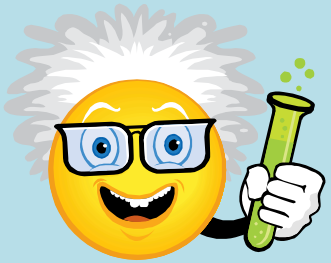


SCI A HOMOGENITY

Number of citations per article

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	N	SCI	Hi
A	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	**20**	**200**	**10**
B	153	2	2	3	3	2	2	3	3	2	2	3	3	2	2	3	3	2	2	3	**20**	**200**	**3**
C	40	40	40	40	40	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**5**	**200**	**5**
D	153	23	3	3	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**5**	**200**	**3**

Who is the best?



IMPACT FACTOR

- **Journal Citation Reports®**
- **journal impact factor (IF) is a measure reflecting the average number of citations to recent articles published in the journal**
- **journals with higher IFs are deemed to be more important than those with lower ones**
- **a qualitative measure of scientific journals and scientific efficiency**
- **value of IF is (re)calculated in each year**
- **due to unexpected fluctuations of the incoming data for IF calculation from year to year, the value of 5-years IF is calculated**

Journal Impact Factor for year 2008 – *Biochemical Pharmacology*

Citations for 2008 per articles from 2007 = 1461
2006 = 1848

# of articles 2007 = 365
2006 = 319

SUM(A) = 3309

SUM(B) = 684

Journal Impact Factor = SUM(A)/SUM(B) = 3309/684 = 4.838

IMPACT FACTOR - CLUSTERS

- absolute value of IF is a quality indicator within certain discipline, but it varies substantially between the disciplines
- major categories – life sciences, technical sciences & engineering, humanities & social sciences
- the factors influencing the absolute IF are e.g. The size of scientific community, number of journals, dynamics of research, individual characteristics of the subject
- we should take in consideration other aspects, such as median IF or occurrence of review journals and sub-disciplinary journals within the cluster
- more complex tools and measures, e.g. Article Influence Score (considers, in which journals is the article cited, and excludes journals autocitations)

CLUSTER	# of JOURNALS	HIGHEST IF	MEDIAN IF
AGRICULTURE, MULTIDISC.	56	3.20	0.65
MATHEMATICS	302	3.08	0.58
BIOCHEMISTRY & MOL. BIOL.	291	33,12	2.86
CELL BIOLOGY	185	36.46	3.33
GENETICS & HEREDITY	165	39.79	2.58
PHYSICS, MULTIDISCIPLINARY	78	42.86	1.30
ZOOLOGY	153	4.73	0.98

ISI Web of Knowledge vs SCOPUS

- **Even though ISI Thompson Scientific is massively used within scientific community, it faces a critics, since it is private company and the source data, which it processes, are not openly available**
- **Therefore, an alternative database SCOPUS (www.scopus.com) is also used, and „unofficial impact factor", citation index and analyses of publication record can be calculated**
- **Owner and provider of SCOPUS is consortium Elsevier (Amsterdam)**

Several national science evaluation systems (including in CZ) are based on ISI and/or SCOPUS data („coffee mincer"

J_{imp} = 10 + 295*FACTOR

FACTOR = (1 – N) / (1 + (N/0.057), where N = (P – 1) / (P_{max} – 1), where

P ranking of journal in WOS cluster
P_{max} total number of journals in WOS cluster

E.g.: CHEMICO-BIOLOGICAL INTERACTIONS

	P	P_{max}	J_{imp}
Biochemistry & Molecular Biology	109	275	32
Pharmacology & Pharmacy	63	219	45
Toxicology	18	75	55

ETHICS IN SCIENCE

- no universal moral and ethics codex in science
- generally, the level of ethics in science reflects the level of ethics in the society

Two main aspects of ethics in science are basically considered, discussed and applied:

- generation of the data — EXPERIMENTATION
- presentation of the data — PUBLICATION

ETHICS IN SCIENCE- EXPERIMENTATION

- ***Data Falsification*** is manipulating research materials, equipment, or processes, or changing or omitting/suppressing data or results without scientific or statistical justification, such that the research is not accurately represented in the research record. This would include the "misrepresentation of uncertainty" during statistical analysis of the data.

- ***Data Fabrication*** is the intentional act of making up data or results and recording or reporting them.

- ***Data Manipulation*** – combines data falsification and data fabrication. Manipulating research data with the intention of giving a false impression. This includes manipulating images (e.g. micrographs, gels, radiological images), removing outliers or 'inconvenient' results, changing data points, etc...

- ***Data Misinterpretation*** – concerns usually statistics, e.g. not including data outliers

- ***Scientific Misconduct or Scientific Fraud:*** are rather general terms; it is the intentional violation of the standard codes of scholarly conduct and ethical behavior in professional scientific research

ETHICS IN SCIENCE- EXPERIMENTATION

• Scientific journals require declaration from authors that the data were obtained in accordance with ethical rules.
• The misconduct can be detected by editors or reviewers in the process of publications, or later due to the inability to reproduce the data by others. Due to the latter, scientific misconduct is unveiled rather in top journals such as Nature or Science, than in sub-average journals, where the data decay without any impact on society

The motivators for scientists to commit misconduct:

Career pressure: Science is still a very strongly career-driven discipline. Scientists depend on a good reputation to receive ongoing support and funding, and a good reputation relies largely on the publication of high-profile scientific papers. Hence, there is a strong imperative to *„publish or perish*". Clearly, this may motivate desperate (or fame-hungry) scientists to fabricate results.

Ease of fabrication: (*„opportunity makes the thief*") In many scientific fields, results are often difficult to reproduce accurately, being obscured by noise, artifacts, and other extraneous data. That means that even if a scientist does falsify data, they can expect to get away with it – or at least claim innocence if their results conflict with others in the same field. There are no "scientific police" who are trained to fight scientific crimes; all investigations are made by experts in science but amateurs in dealing with criminals.

SCHÖN AFFAIR

• Jan Hendrik Schon – worked at Bells laboratories - physics of condensed materials and nanotechnology – received many scientific awards - high scientific output - in 2001 – in average 1 paper every 8 days (Nature, Science…..)
• in 2001 Schon published in Nature, the construction of transistor at molecular scale, using organic dyes. It would be breaking discovery, since it would be the end of silicon age in electronics, and the invention of molecular semiconductors.

• soon after that „discovery", the doubts from scientific community aroused, concerning anomalies in the paper; prof. Sohn pointed at the identical level of background noise in two experiments carried out at different temperatures
• Nature editors questioned Schon, who answered that he submitted 2 x the same graph by mistake
• prof. McEuen found that even 3 graphs have identical noise; many other scientists detected duplicate data in Schons´ papers
• Lucent Technologies (owner of Bell) opened detailed investigation – Schon was asked to provide original data, but he *„did not write laboratory protocols"*, *„erased data from his PC"*, and experimental samples were *„lost or discarded"*
• investigative committee described 24 cases of „misconduct", e.g. where Schon did not constructed plots from experimental data but instead of he fabricated them using mathematic functions

ETHICS IN SCIENCE- PLAGIARISM

- The use of ideas, concepts, words, or structures without appropriately acknowledging the source to benefit in a setting where originality is expected.

- publications, books, theses (bachelors, diploma, PhD)
- pay attention to properly cite the source
- the most serious violation that can ever be done in academic sphere
- may result in annulation of the product, retraction of publications or abrogation of the academic degree

ETHICS IN SCIENCE- AUTHORSHIP

Emergency Medicine (2003) 15, 263–270

Emergency Medicine

SHORT REPORT

Unethical practices in authorship of scientific papers

Dianne M Bennett and David McD Taylor
Emergency Department, Royal Melbourne Hospital, Parkville, Victoria, Australia

BENEFITS OF AUTHORSHIP

- Contribution to the progress of science
- Personal sense of achievement
- Evidence of an individual's intellectual efforts
- Contribution to an individual's professional reputation
- Creation of currency for: academic appointment, promotion and research funding entry to professional bodies

RESPONSIBILITIES OF AUTHORSHIP

In accepting this responsibility, authors are certifying the integrity of their work. Responsibility for publications includes the **requirement and willingness to publicly defend their content if challenged by readers**. Authors also have a responsibility to honor the trust of readers who, perforce, place trust in not being able to check the work that was undertaken for the writing of the paper. This honor is fulfilled by ensuring that the contents of publications are an accurate representation of the work undertaken.

ETHICS IN SCIENCE- AUTHORSHIP

- large increase of MULTI-AUTHORED papers in last decades
- in past times - usually 1or 2 authors, now often 10-20 authors (or more)
- career pressure – „...publish or perish…"
- NON- ETICAL PRACTICES concerning authorship

International Committee of Medical Journals Editors (ICMJE)

- **„*Vancouver group*" – released ICMJE authorship guidelines**
- **All three of the following criteria must be met for individuals to qualify for authorship:**

- **Substantial contribution to the conception and design of a study; or acquisition of data; or interpretation of data**
- **Drafting the study manuscript or critically revising it for important intellectual content**
- **Giving final approval of the version to be published**

AUTHORSHIP AND PUBLICATION IRREGULARITIES

- GUEST AUTHORSHIP
- PRESSURED AUTHORSHIP
- GHOST AUTHORSHIP
- DIVIDED PUBLICATION
- DUPLICATE PUBLICATION

GUEST AUTHORSHIP

- **Inclusion of an individual in the by-line who does not meet authorship criteria also known as 'gift, honorary or unjustified' authorship**
- Guest authors do not help write the paper and may not have seen the final version submitted to the journal. Therefore, they are incapable of defending its contents.

The most common reasons for guest authorship are:

- The pressure to publish, which sometimes provokes junior researchers to add a senior colleague, whose name carries kudos within the scientific community, to the by-line in the hope that this will increase the chances of their work being published
- Repaying favors; for example, referral of a patient, motivating a team and encouraging collaboration, or including laboratory technicians whose contribution was nothing more than their routine work that would have been done regardless

- **Guest authorship is seen as intellectually dishonest, deceptive, unethical and causes dilution of credit for scientific work and the validity of a paper**
- In many journals „**authors contribution**" is imperatively disclosed

PRESSURED AUTHORSHIP

- A variation of guest authorship, may be subtle or overt in nature.
- A person who uses their position of authority to apply pressure upon staff more junior to them to include them as an author, even though they do not qualify.

GHOST AUTHORSHIP

- **The failure to name an individual as an author when they have contributed substantially to the research or writing of the article.**

- Ghost authorship is almost the reverse of guest authorship and may exist in several forms. In all cases, the individual is not listed as an author but could have either made contributions worthy of authorship or participated in the writing.
- An example is when a corporation hires a professional writer to write a review article on topics related to a new product in order to promote it. Notable clinicians are then invited to submit the article to a journal for publication, with their names as authors, in exchange for honorariums and without revealing any conflict of interest.
- Paradoxical situations of ghost authorship have been known to occur. In one instance, a researcher intentionally excluded his name from a manuscript that reported the poor performance of a cholesterol analyser because the negative conclusion may have been perceived as being unfriendly to industry and therefore have the potential to jeopardise future funding.

DIVIDED PUBLICATIONS

- **The fragmentation of data and findings from a single study: the researcher(s) publish their work in several short papers when it probably could have been published as one single, longer paper.**

- This trend for papers of shorter length has precipitated divided publishing, otherwise known as the '**Least Publishable Unit (LPU)**', '**salami science**' or '**slicing the salami**'.
- Divided publication is difficult for journal editors to police because they are usually not informed that other papers have been derived from the work they have accepted, or that these other papers have been simultaneously submitted to other journals.
- One problem with divided publication is that it swells the amount of literature published, and not necessarily for the better.

DUPLICATE PUBLICATIONS

- **The same content being republished in successive papers.**

- Example - two papers that have been published by the same author, in separate journals but in the same year, on communication failure among medical personnel. The papers are close to identical, except for the order in which examples of cases of failed communication have been discussed.
- Duplicate publication can also manifest when a case is reported for a second time, only by a different author. This type of duplication is misleading because it misrepresents the incidence of the case in electronic literature searches.
- An acceptable form of duplicate publication, however, may be for the conveying of information to separate audiences. For example, a paper being republished in a second journal because the one in which it was originally published is not available in the country of its republication. In all cases, duplicate publication requires written permission of the original journal editor, formal acknowledgement of the original journal and publication of the original article.

- Great problem in countries where the world languages are not spoken
- Often one paper published in Czech and the same content in English published elsewhere – undetectable.
- Republication at different platform – no overlap of readers - undetectable