A general source-normalized approach to bibliometric research performance assessment

Ludo Waltman and Nees Jan van Eck
Centre for Science and Technology Studies, Leiden University

STI 2010, Leiden, The Netherlands
September 10, 2010

Universiteit Leiden

CWTS

Introduction

- Citation counts often need to be normalized for field and publication age

- Normalization is usually done based on a field classification scheme (e.g., WoS subject categories)

- Such a scheme is always somewhat arbitrary

- Source normalization does not need a classification scheme

- Proposal for a new source-normalized indicator

Universiteit Leiden

Importance of normalization for field and publication age
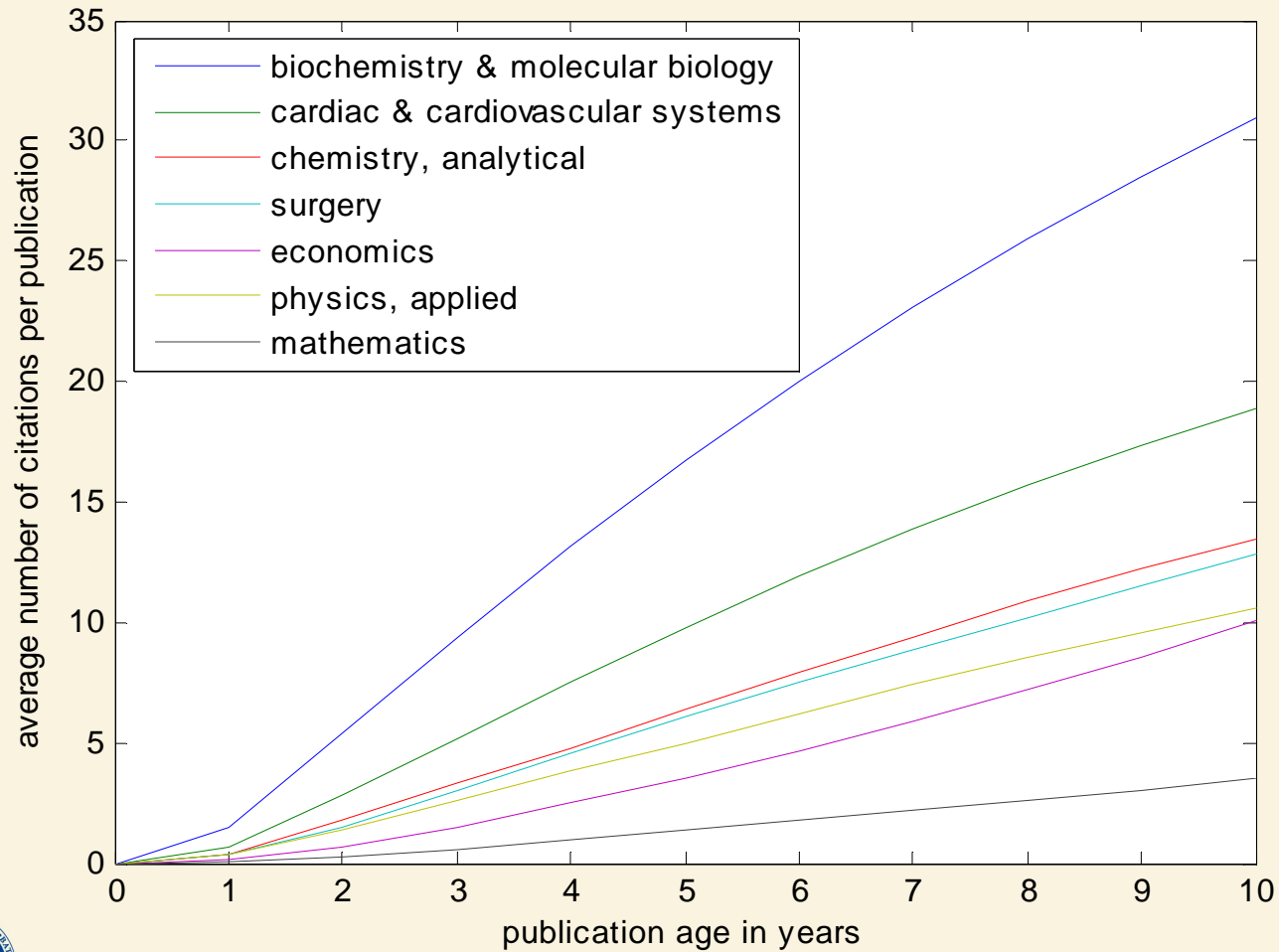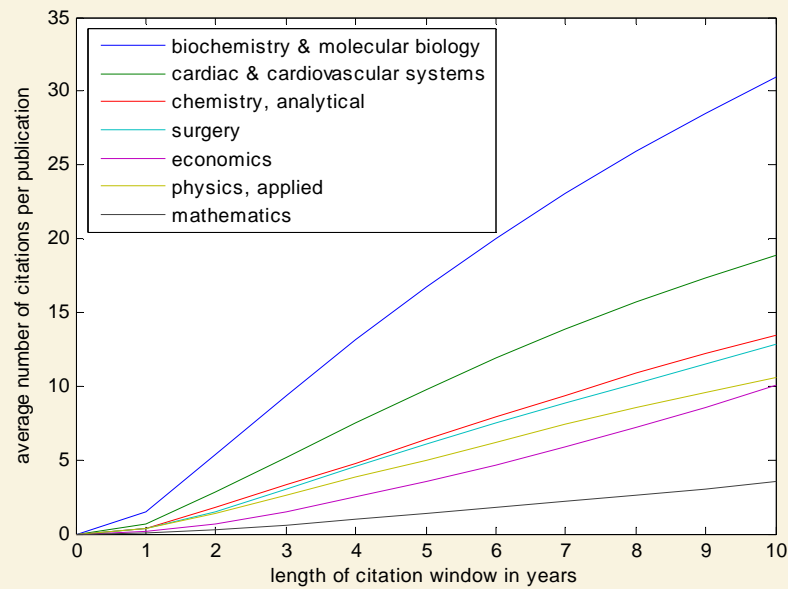
Two normalization approaches

- Normalization based on a field classification scheme (cited-side normalization)

- Source normalization (citing-side normalization)

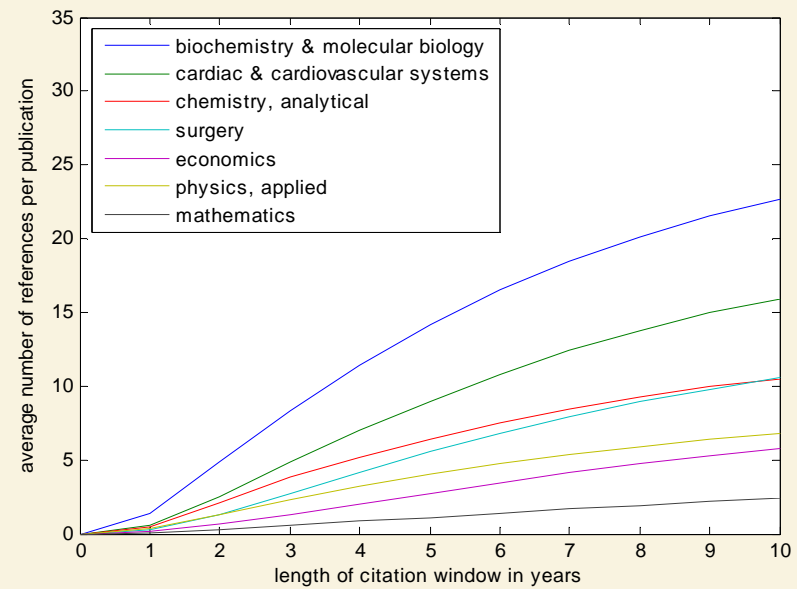| Cited-side	Citing-side
CPP/FCSm	Audience factor
MNCS	SNIP
NMCR	MSNCS
Citation z-score	

Universiteit Leiden

CWTS

Source normalization: Origin of field differences

Citations

References

Source normalization: Currently existing indicators

- Source-normalized indicators:
  - Audience factor (AF; Zitt & Small, 2008; Zitt, 2010)
  - Source-normalized impact per paper (SNIP; Moed, 2010)
  - 'Fractional citation counting' (Leydesdorff & Opthof, in press)

- Advantage:
  - No dependence on a field classification scheme

- Limitations:
  - No normalization for publication age
  - Specific focus on measuring citation impact of journals (AF and SNIP)

Universiteit Leiden

Mean source-normalized citation score (MSNCS)

- General-purpose indicator, not only for journals but also for research groups, departments, etc

- Counts all citations received by a publication, not just the citations received in one particular year

- Normalization not only for field but also for publication age

Universiteit Leiden

MNCS vs MSNCS

$$MNCS = \frac{1}{p} \sum_{i=1}^{p} NCS_i$$

p: Number of publications

NCS_i: Normalized citation score of publication i

$$MSNCS = \frac{1}{p} \sum_{i=1}^{p} SNCS_i$$

$SNCS_i$: Source-normalized citation score of publication i

Universiteit Leiden

NCS vs SNCS

$$\mathbf{NCS}_{ii} = \frac{c_{ii}}{\mathbf{MCF}_{ii}} = \sum_{j=1}^{c_i} \frac{1}{\mathrm{MCF}_i}$$

c_i: Number of citations of publication i

a_i: Age of publication i

MCF_i: Mean number of citations per publication in the field of publication i, using an a_i-year citation window

$$\mathrm{SNCS}_i = \sum_{j=1}^{c_i} \frac{1}{\mathrm{MRJ}_j}$$

MRJ_j: Mean number of references per publication in the journal of publication j, using an a_i-year citation window

Universiteit Leiden

Justification of bibliometric indicators

- Indicators can be justified in different ways:
  - Informal, intuitive arguments
  - Empirical arguments
  - Theoretical arguments

- What exactly do we mean by field normalization?
  - Normalization for reference list length
  - Normalization for inter-field citation traffic
  - Normalization for growth rate

Universiteit Leiden

# Theoretical justification of MSNCS

- Assumptions
  - No inter-field citation traffic
  - No growth of fields (or equal growth of all fields)
  - Fields in 'steady state'

- Result
  - For any publication window and any citation window, the MSNCS of all publications in a field always equals 1
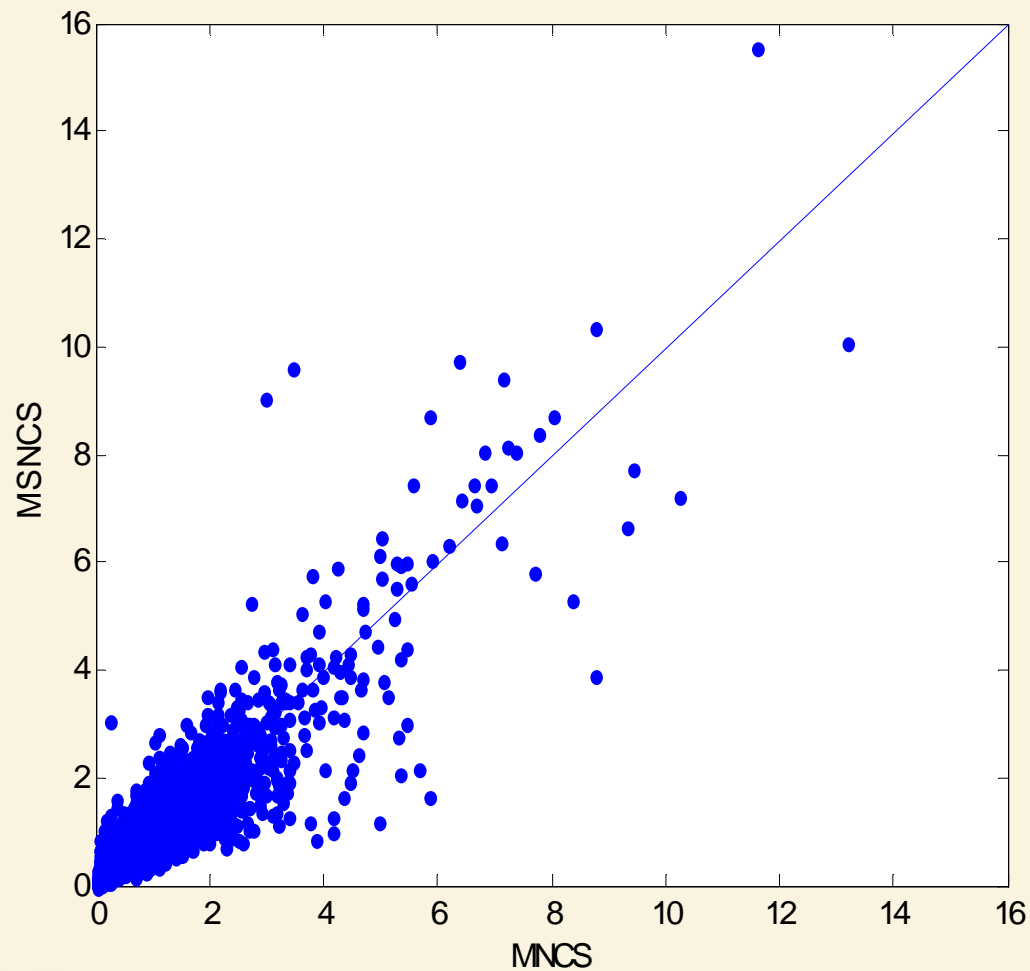
Universiteit Leiden

MNCS vs MSNCS: Chemistry research groups with > 50 publications between 1991 and 1999
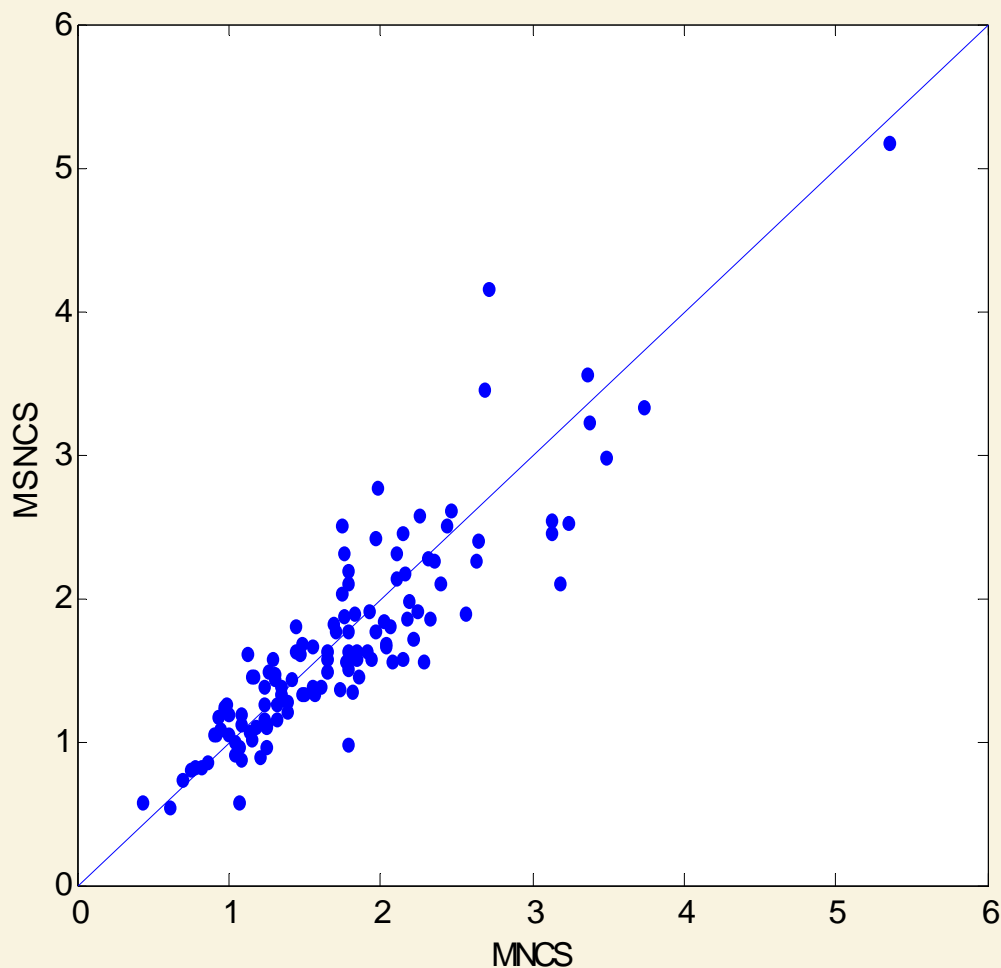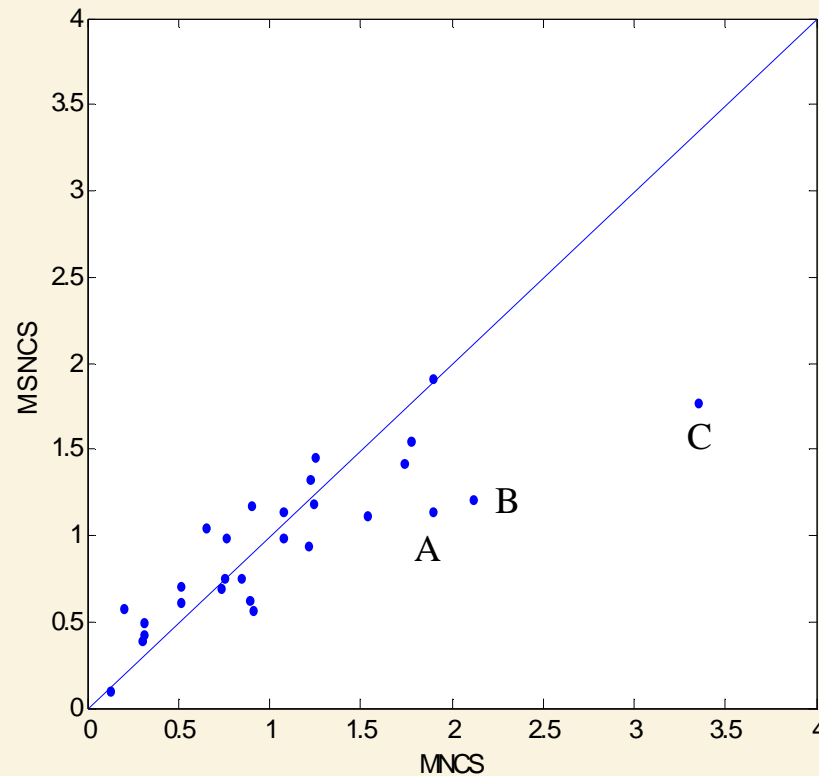
MNCS vs MSNCS: UMC departments with > 50 publications between 2003 and 2008
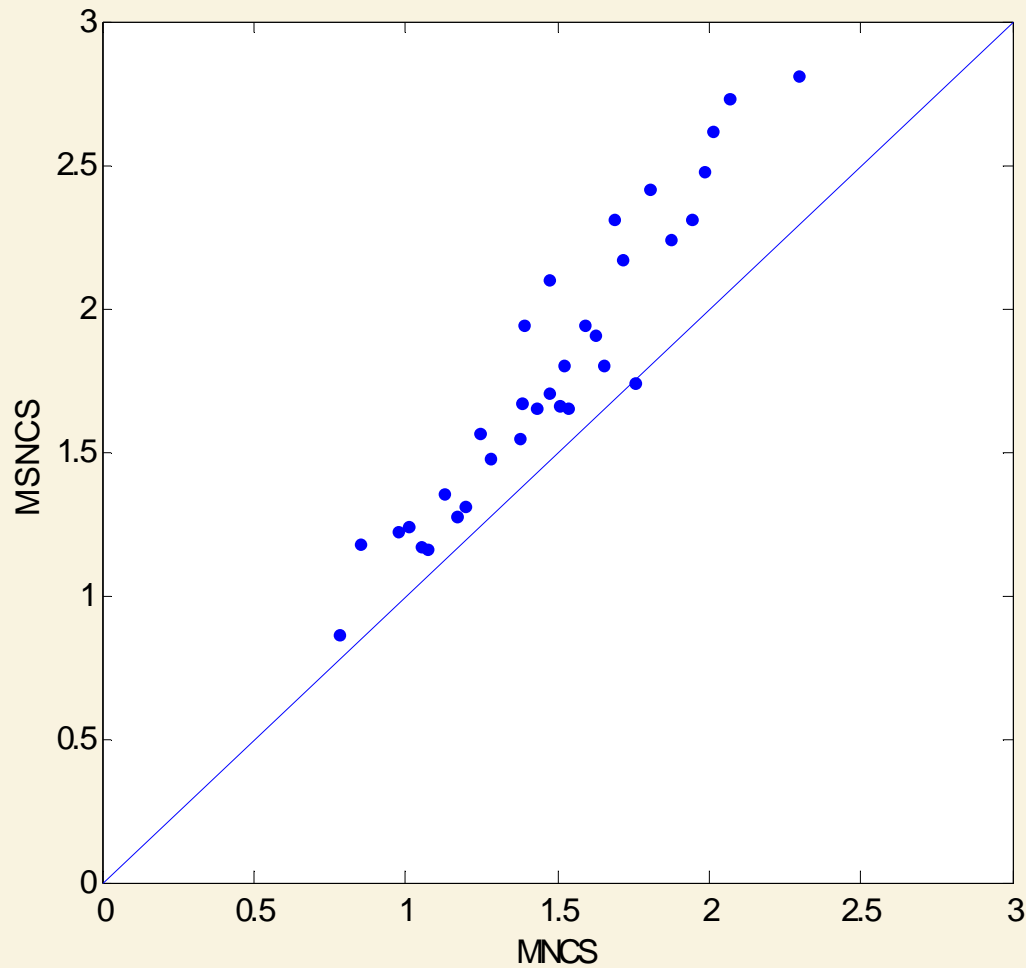
MNCS vs MSNCS: All journals with > 100 publications between 2005 and 2007

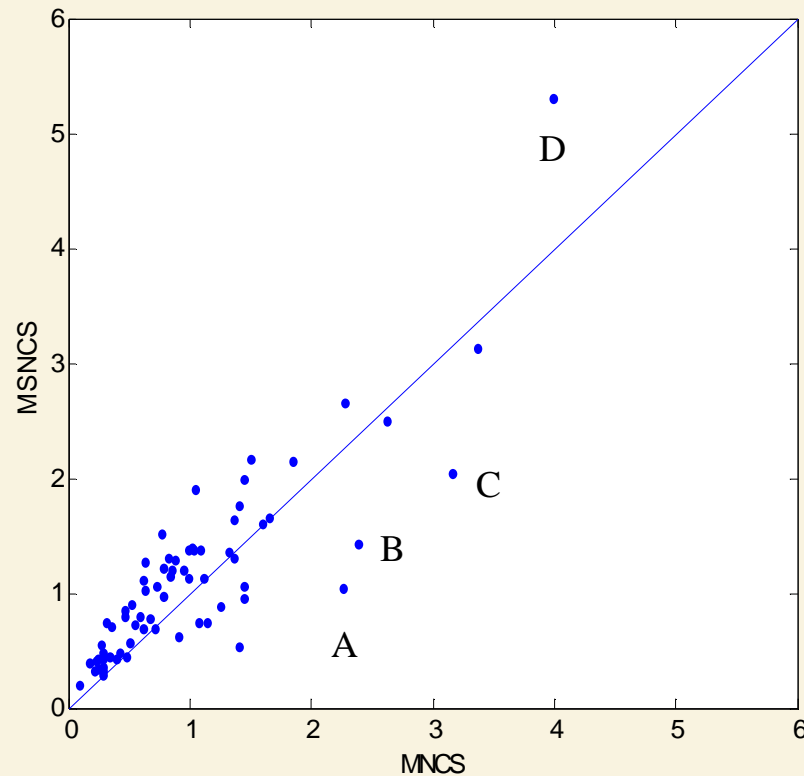# MNCS vs MSNCS: Information science & library science

A: Journal of the Medical Library Association

B: Journal of Health Communication

C: Journal of the American Medical Informatics Association

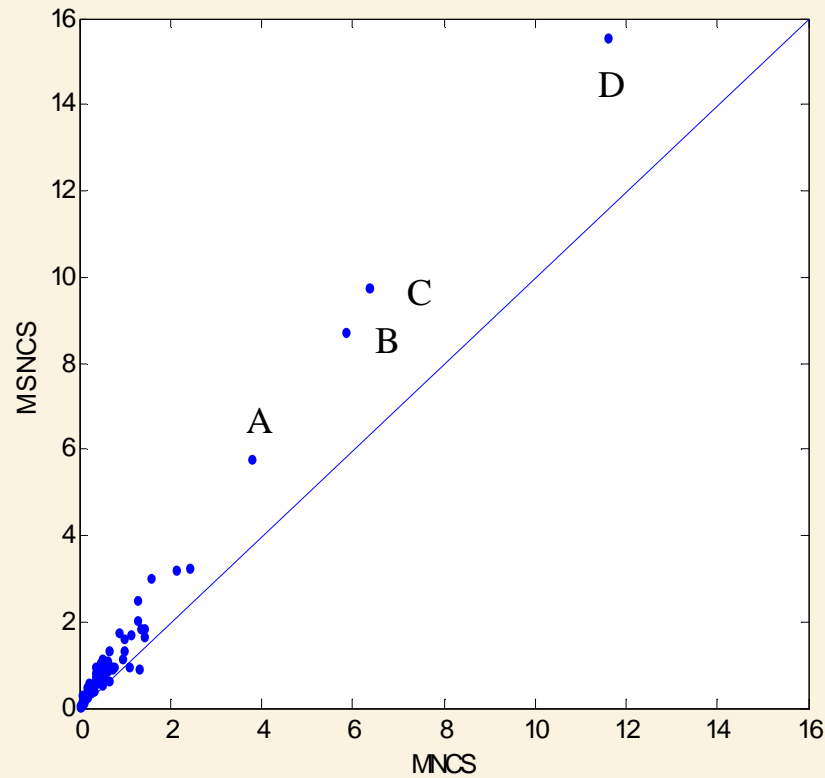# MNCS vs MSNCS: Statistics & probability

A: Chemometrics and Intelligent Laboratory Systems

B: Fuzzy Sets and Systems

C: Biostatistics

D: Econometrica

MNCS vs MSNCS: Medicine, general & internal

A: Annals of Internal Medicine
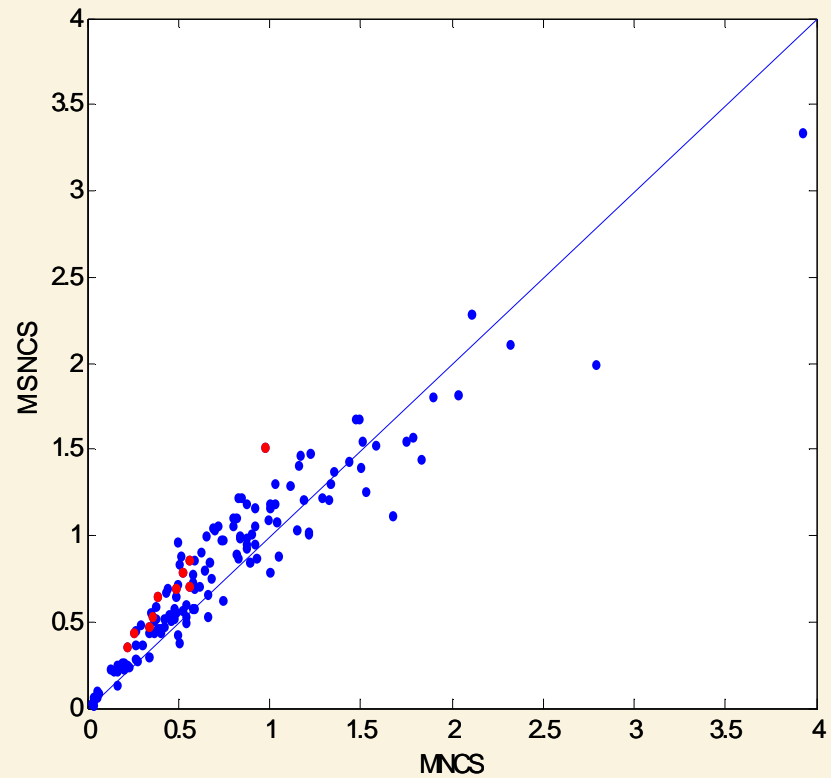
B: Lancet

C: JAMA – Journal of the American Medical Association

D: New England Journal of Medicine

Universiteit Leiden

MNCS vs MSNCS: Surgery

Plastic and reconstructive surgery journals are shown in red.

The 'trade journal problem'

- Problematic journals for source normalization:
  - Trade journals
  - National or regional journals

- These journals have very few references per publication

- Consequently, citations from these journals have much more weight than citations from ordinary scholarly journals

- For example, weight of a citation from
  - *Journal of the American Society for Information Science and Technology*: 0.26
  - *African Journal of Library Archives and Information Science*: 1.78

Universiteit Leiden

Conclusions

- There is no perfect way of normalizing citation counts

- Cited-side and citing-side normalization can best be seen as complementary to each other

- Limitations of the two types of normalization:

Citing-side	Cited-side
Dependence on a field classification scheme	No normalization for growth of fields
	No normalization for inter-field citation traffic
	Trade journal problem

Universiteit Leiden

Thank you for your attention!

Universiteit Leiden

References

Leydesdorff, L., & Opthof, T. (in press). Normalization at the field level: Fractional counting of citations. *Journal of Informetrics*.

Moed, H.F. (2010). Measuring contextual citation impact of scientific journals. *Journal of Informetrics*, *4*(3), 265–277.

Waltman, L., Van Eck, N.J., Van Leeuwen, T.N., Visser, M.S., & Van Raan, A.F.J. (in press). Towards a new crown indicator: Some theoretical considerations. *Journal of Informetrics*.

Zitt, M. (2010). Citing-side normalization of journal impact: A robust variant of the Audience Factor. *Journal of Informetrics*, *4*(3), 392–406.

Zitt, M., & Small, H. (2008). Modifying the journal impact factor by fractional citation weighting: The audience factor. *Journal of the American Society for Information Science and Technology*, *59*(11), 1856–1860.

Universiteit Leiden