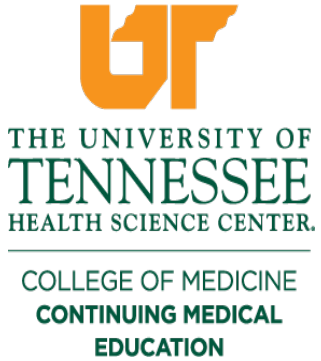
Collaborating with a Statistician Series
SESSION 3:

How can a statistician contribute to a grant application or a manuscript?

Fridtjof Thomas, PhD

1 CME/AMA credit will be provided



AMA Credit Designation: The University of Tennessee College of Medicine (UTCOM) designates this live activity for a maximum of 1 *AMA PRA Category 1 Credits™*. Physicians should claim only the credit commensurate with the extent of their participation in the activity.

Continuing Education for Non-Physicians: The UTCOM will issue Certificates of Participation to non-physicians for participating in this activity and designates it for CEUs using the national standard that 1 hour of educational instruction is awarded .1 CEU.

Accreditation: The UTCOM is accredited by the Accreditation Council for Continuing Medical Education to provide continuing medical education for physicians.

Commercial Support

No commercial support was received for this CME activity.

Planner & Speaker Disclosures

No planners & speakers have relevant financial relationships to disclose.

Objectives

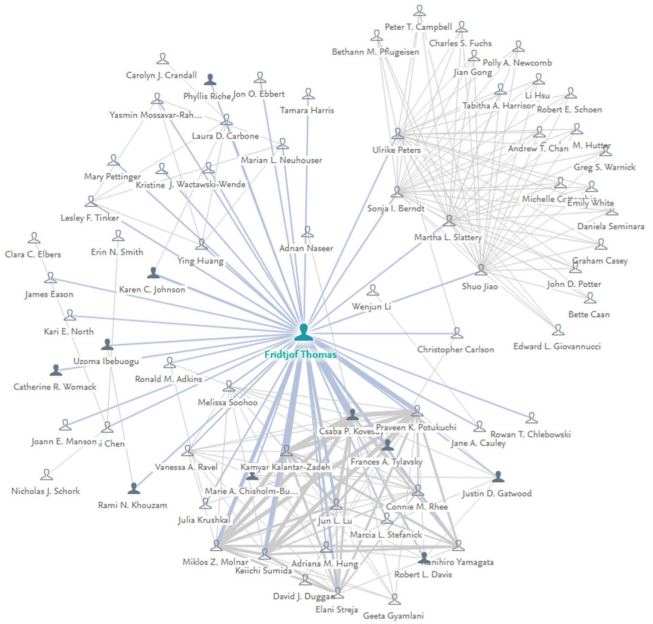
- Describe at which phase of a study a statistician should get involved/be contacted
- Find and retrieve applicable reporting standards for manuscripts
- Distinguish between consulting and collaborating with a statistician
- Define the role of authors and contributors based on the recommendations of the International Committee of Medical Journal Editors (ICMJE)

About the presenter

- Professor at the Division of Biostatistics, Dept. of Preventive Medicine
- At UTHSC since 2007
  - 6 years on the Research Subcommittee of the Faculty Senate (2014 – 2020)
- Design and Analysis Committee of the EARLY trials (2010-2016 – "Early Adult Reduction of weight through LifestYle intervention," a collection of seven randomized clinical trials funded by the National Heart, Lung, and Blood Institute (NHLBI) and the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) of the National Institutes of Health (NIH))
- Member of the Biostatistics Collaborative Core at the Southeast Regional Center of the NIH-NHLBI-funded Women's Health Initiative (WHI) study that has recruited over 160,000 women in over 40 clinical centers nationwide. (2010-2017)
- Grant review experience since 2012 from
  - Department of Defense's Congressionally Directed Medical Research Program (DoD CDMRP)
  - NIH Epidemiology of Chronic and Infectious Disease Study Section
  - NIH Neurological, Aging, and Musculoskeletal Epidemiology (NAME) Study Section
- 13+ years Associate Editor of *The Journal of Statistical Computation and Simulation* (JSCS; a Taylor & Francis print journal since 1972)

Network of collaboration (Elsevier Pure “Fingerprint” 10/09/2019)

Research output network – organizational units	
Department of Preventive Medicine	77
Department of Medicine, Division of Nephrology	23
Department of Pediatrics	6
Department of Pharmaceutical Sciences	5
Department of Clinical Pharmacy – Memphis	4
Department of Clinical Pharmacy – Nashville	4
Neuroscience Institute	4
Department of Ophthalmology	4
Department of Medicine, Division of Cardiovascular Diseases	4
Department of Medicine, Division of General Internal Medicine	3
Department of Acute and Tertiary Care	2
Department of Health Promotion and Disease Prevention	2



“If you need a statistician, it isn’t significant!”

(from the basic sciences)

Tip for clear writing…

Especially in the statsection, use words with statistical meaning only in that statistical sense.

Significant: important, consequential, clinically meaningful, substantial, relevant.

If you want to create confusion, write sentences like:

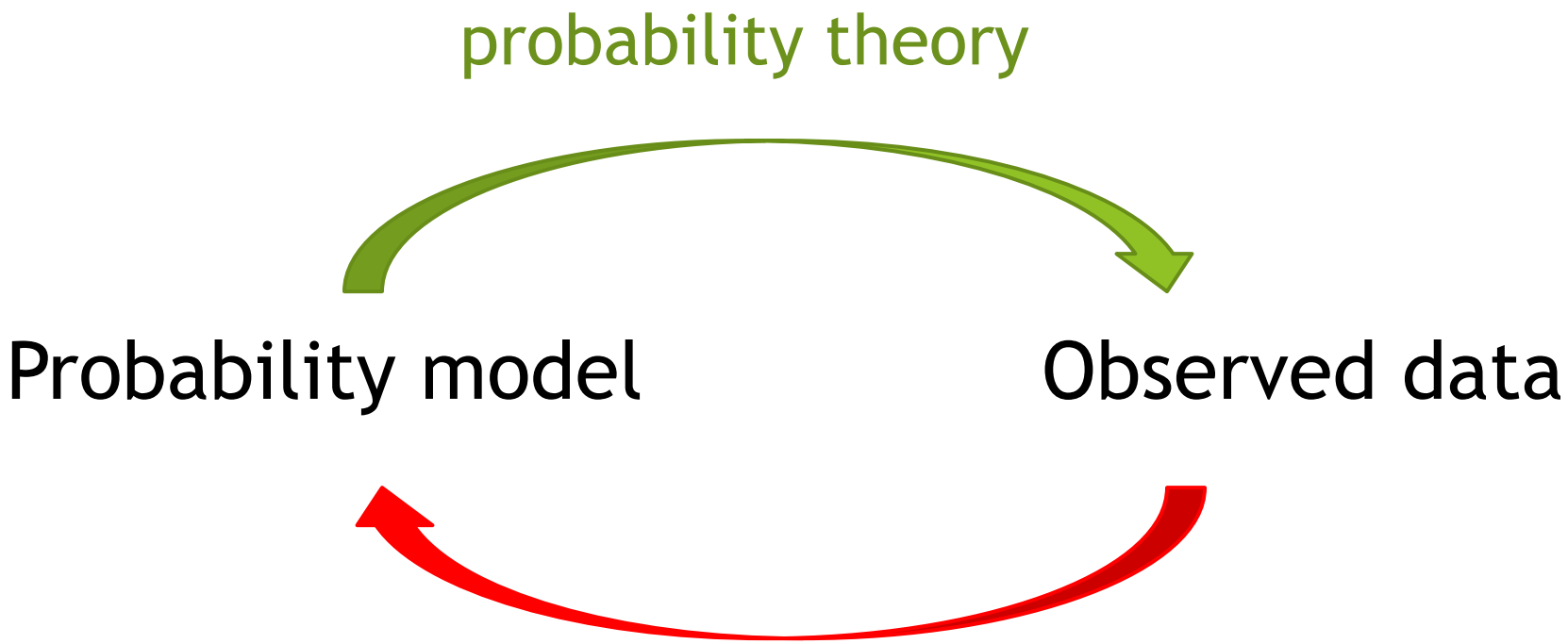
The only even prime number is 2, which makes it the oddest prime number.

(Reserve this for the conference dinner conversation…)

References on writing:

- Williams JM, Bizup J. Style: *Lessons in Clarity and Grace*. Twelfth ed. Boston: Pearson; 2017.
- Higham NJ. *Handbook of Writing for the Mathematical Sciences*. Philadelphia: Society for Industrial and Applied Mathematics (SIAM); 1993.
- Murray R. *Skillful writing of an awful research paper*. Analytical Chemistry. 2011;83(3):633-. doi: 10.1021/ac2000169.

Statistics: The basic problem



Grant applications:
why you should care about your statsection

- A good statsection will not safe a proposal based on a stupid idea
- A bad statsection can kill an otherwise fantastic proposal
- If you are in the “possibly fundable” group, the following will happen:
  Any reviewer that does not like your proposal will turn to the statsection to come up with 3-5 “objective” reasons why your proposal is not so good

Top 10 errors in grant proposals

Dr. Israel Goldberg, UTHSC grant consultant (from my notes 03/25/2008)

1. Proposing to do too much
2. No Hypotheses or predictions ("bean counting" and/or "fishing expedition")
3. Silly Hypothesis
4. Disconnect between Specific Aims and Research Design & Methods
5. Expertise missing
6. Non-modular budget
7. Tilting at other people's windmills
8. Sloppiness (typos, poor grammar, inconsistent information)
9. Unexplained hiatus in productivity
10. Amended proposals: Don't argue with the reviewers

Don't give reviewers targets to throw darts at.

What statistical methods do you need?

In a well-designed randomized study with a clear and simple outcome, and essentially complete follow-up, a simple and straight-forward analysis might be all that is needed.

In all other cases, gradually more sophisticated models need to be employed.

What statistical methods do you need?

Do you have any grouping structure in your data? (e.g., batch effects, same healthcare provider, patients repeatedly included in your encounter-based data)

How do you address sex/gender and race?

NIH Data Management and Sharing Plan? (Jan 25, 2023)
https://sharing.nih.gov/

NIH Policy on Sex as a Biological Variable

"NIH expects that sex as a biological variable will be factored into research designs, analyses, and reporting in vertebrate animal and human studies. Strong justification from the scientific literature, preliminary data, or other relevant considerations must be provided for applications proposing to study only one sex."

DMP

Applications for Receipt Dates ON/AFTER Jan 25 2023

ON THIS PAGE:

- Writing a Data Management and Sharing Plan
- Submitting Data Management and Sharing Plans
- Data Management and Sharing Plan Format
- Elements to Include in a Data Management and Sharing Plan
- Assessment of Data Management and Sharing Plans
- Revising Data Management and Sharing Plans
- Additional Considerations

https://sharing.nih.gov

Levels of evidence

1a	Systematic review of high quality RCTs with similar results and effect sizes for many different RCTs.
1b	Individual high quality RCT with high precision (narrow confidence interval)
1c	All or none
2a	Systematic review of cohort studies with similar results and effect sizes.
2b	Individual cohort study or low quality RCT (e.g., <80% follow-up)
2c	"Outcomes Research" and ecological studies (based on average exposures etc. of populations of geographical or temporal units)
3a	Systematic review of case-control studies
3b	Individual case-control study
4	Case-series and poor-quality cohort and case-control studies
5	Expert opinion (unless critically appraised or based on "first principles")

Source: Oxford Centre for Evidence-based Medicine https://www.cebm.net/2009/06/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/

All or none: Example “Bubble Boy” disease

- Babies born without functional immune system.
- SCID-X1: 1 in 50,000-100,000 affected; caused by a mutation in a gene (IL2RG)
- Most die within first year of life. (Only about 20% have access to a suitable sibling for a bone-marrow transplant as the existing cure.)

St. Jude announced April 18, 2019: Gene therapy cure for babies with X-linked severe combined immunodeficiency

“The gene therapy, produced in the Children’s GMP, LLC, manufacturing facility on the St. Jude campus, involved use of a virus to transport and insert a correct copy of a gene into the genome of patients’ blood stem cells. Following the treatment, the children began producing functioning immune cells for the first time, according to St. Jude, and most patients were discharged from the hospital within one month.” [All 8 babies started to produce complete sets of immune cells.]

https://www.stjude.org/inspire/news/bubble-boy-scid-x1-cure.html
https://www.stjude.org/research/news-publications/research-highlights/2019-research-highlights/st-jude-gene-therapy-holds-promise-for-treating-several-diseases.html

Typically needed: CONSORT chart

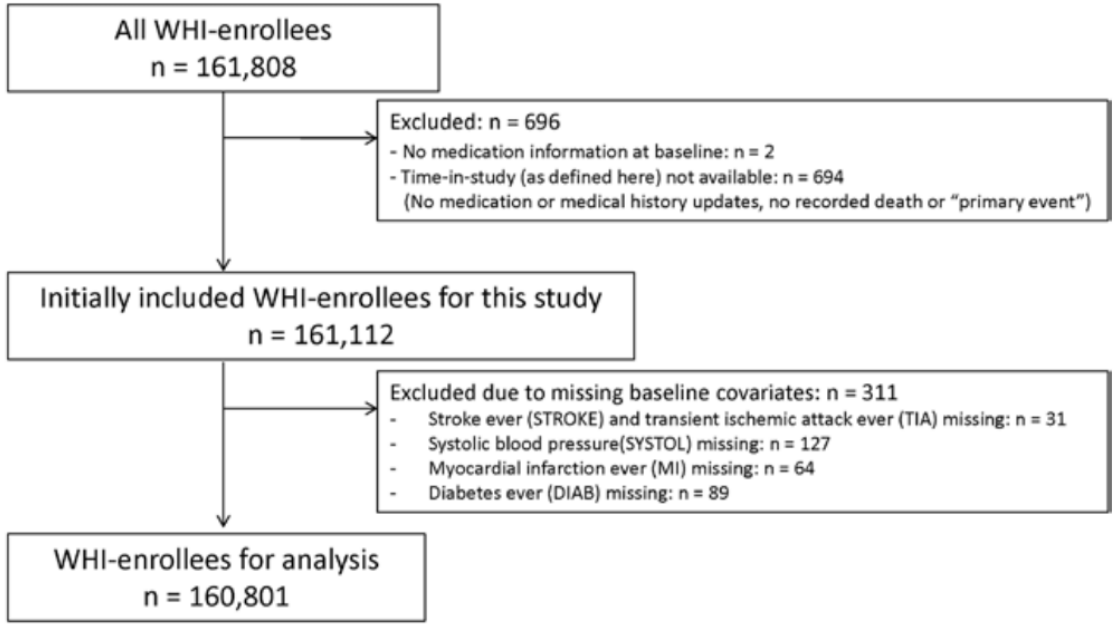


Figure 1. Of all 161 808 Women Health Initiative (WHI) enrollees, 160 801 were used in the analysis. Women were only excluded because of missing baseline covariates (n=311) if the number of women with that missing variable was small. Otherwise, a separate factor level missing was maintained in categorical variables.

Source (this and next slides): Bavry AA, Thomas F, Allison M, Johnson KC, Howard BV, Hlatky M, Manson JE, Limacher MC. Nonsteroidal Anti-Inflammatory Drugs and Cardiovascular Outcomes in Women: Results From the Women's Health Initiative. Circulation Cardiovascular quality and outcomes. 2014;7(4):603-10. PubMed PMID: 25006185.

Typically needed: Description of study group(s)

Table 1. Baseline Characteristics for Women With Regular NSAID Use and Those With No NSAID Use

Characteristic	NSAID Use at Baseline (n=31 433)	No NSAID Use at Baseline (n=129 368)	*P* Value
Age, y; mean (SD)	63.3 (7.1)	63.2 (7.2)	0.689
Age >70 y, %	21.6	22.1	0.031
Race/ethnicity, %			
White	85.2	82.1	<0.001
Black	8.8	9.1	
BMI, kg/m^2; mean (SD)	29.4 (6.4)	27.6 (5.8)	<0.001
BMI categories, %			
Normal (18.5–24.9)	26.2	35.9	<0.001
Overweight (25.0–29.9)	33.5	34.7	
Obesity I (30.0–34.9)	21.9	17.5	
Obesity II (35.0–39.9)	10.6	6.8	
Extreme obesity III (≥40)	6.6	3.3	
Systolic blood pressure, mm Hg, SD	128.7 (17.6)	127.0 (17.8)	<0.001
Systolic blood pressure, %			
<120	34.7	39.4	<0.001
120–140	42.6	40.4	
>140	22.8	20.2	
History, %			
Hypertension	38.1	32.4	<0.001
Hypercholesterolemia	14.3	13.1	<0.001
Diabetes mellitus	6.9	5.7	<0.001
Smoking status			
Never smoked	48.4	50.8	<0.001
Past smoker	43.7	41.0	
Current smoker	6.7	6.9	
Congestive heart failure	0.9	0.8	0.074
Myocardial infarction	2.3	2.3	0.479
CABG or PCI	1.7	1.7	0.970
Stroke or transient ischemic attack	3.0	3.0	0.825
Peripheral arterial disease	2.6	1.9	<0.001
Gastric or duodenal ulcer	7.1	6.3	<0.001
Bleeding problem	2.7	2.5	0.117
Rheumatoid arthritis	8.3	4.0	<0.001
Cancer*	9.6	9.1	0.010
Medications, %			
Aspirin	22.2	22.7	0.036
Acetaminophen	21.4	11.2	<0.001
Statin	8.7	7.3	<0.001
Menopausal hormones	47.4	40.8	<0.001

BMI indicates body mass index; CABG, coronary artery bypass grafting; NSAID, nonsteroidal anti-inflammatory drug; and PCI, percutaneous coronary intervention.

*Excluding nonmelanoma skin cancer.

t-tests, Wilcoxon rank-sum tests
Chi-square tests, Fishers exact test

Typically needed: Results

Table 3. Association of NSAIDs for the Primary Outcome (Cardiovascular Death, Nonfatal Myocardial Infarction, or Nonfatal Stroke)

Exposure	Person-Time, y	No. of Events (Cases)	Incidence rate (Per 100 Person-Years)	Hazard Rate (Covariate-Adjusted Cox Regression)*	
				HR (95% CI)	*P* Value
Group 1 NSAID (cox-2 selective agents)				1.13 (1.04–1.23)	0.004
Celecoxib only	29 344	317	1.08	1.13 (1.01–1.27)	0.031
Rofecoxib only	23 835	240	1.01	1.14 (1.00–1.29)	0.055
Group 2 NSAID (nonselective agents with cox-2>cox-1 inhibition)				1.17 (1.10–1.24)	<0.001
Naproxen only	58 623	530	0.90	1.22 (1.12–1.34)	<0.001
Nabumetone only	16 580	148	0.89	1.14 (0.97–1.34)	0.118
Diclofenac only	18 226	165	0.91	1.15 (0.99–1.35)	0.070
Etodolac only	7591	61	0.80	1.01 (0.78–1.30)	0.963

(long list…)

Typically needed: Results

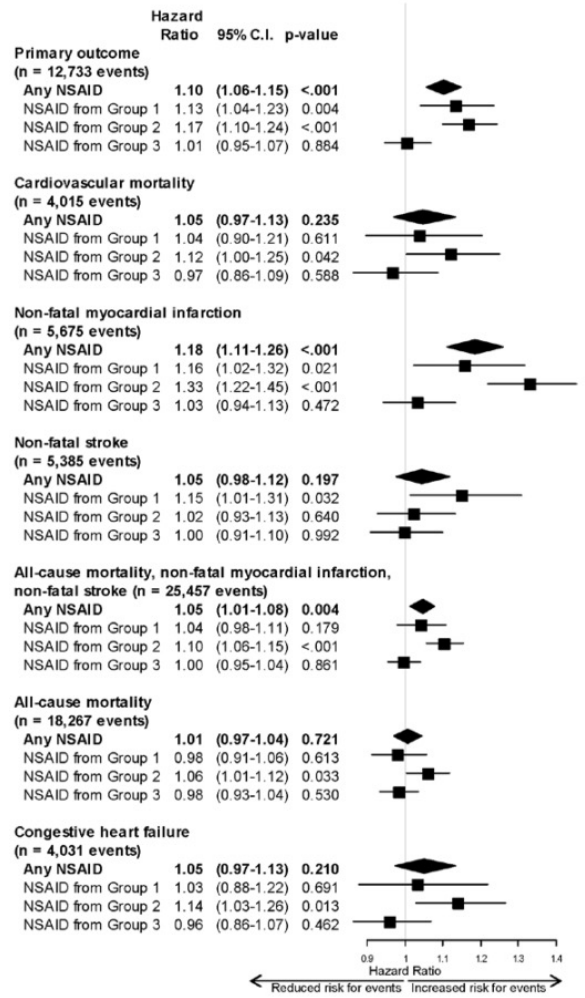


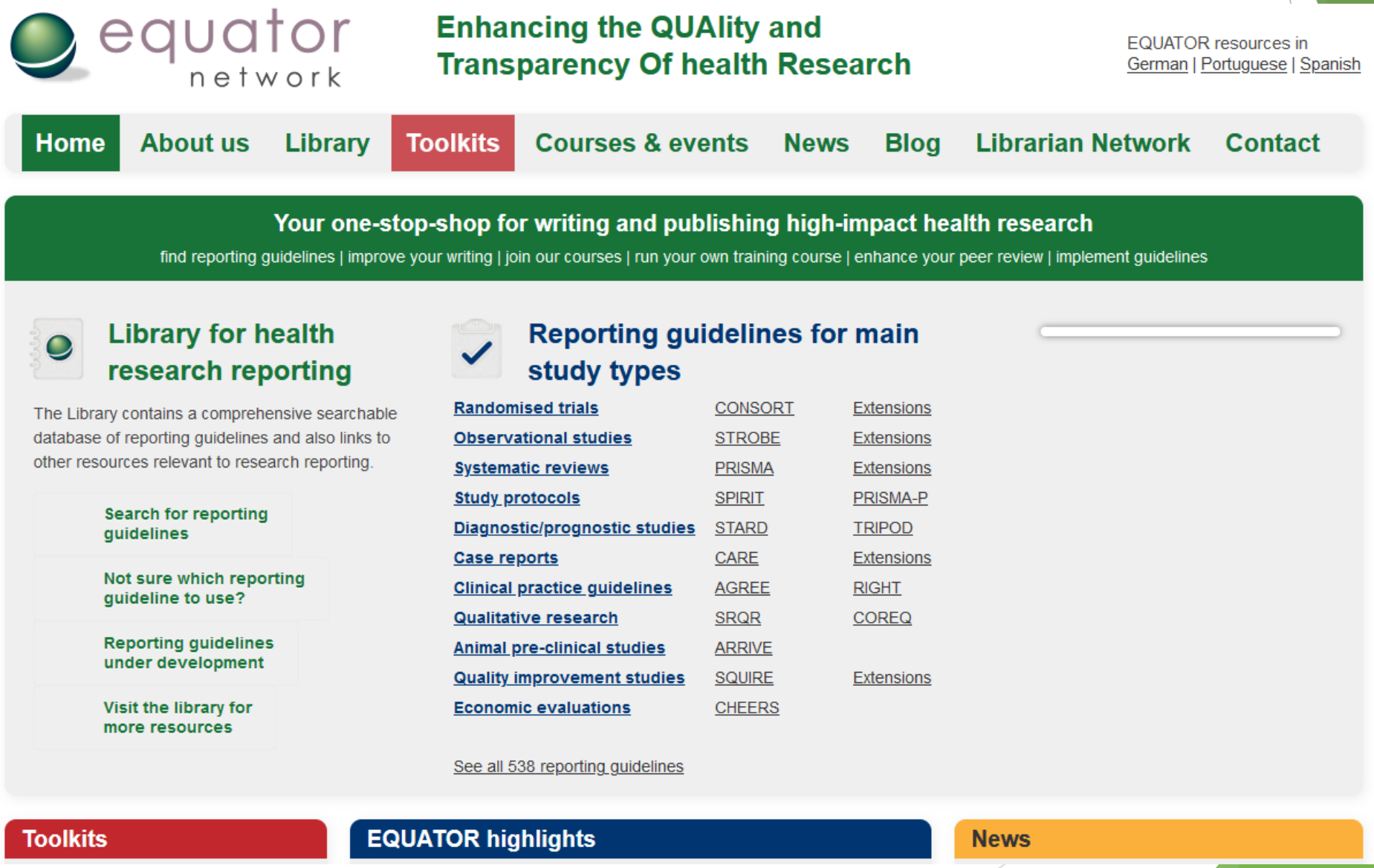
Figure 2. Adjusted hazard ratios for the primary outcome (cardiovascular mortality, nonfatal myocardial infarction, or nonfatal stroke) and secondary outcomes for regular nonsteroidal anti-inflammatory drug (NSAID) use versus no NSAID use. Group 1=cox-2 selective agents, group 2=nonselective agents with cox-2>cox-1 inhibition, and group 3=nonselective agents with cox-1>cox-2 inhibition.

Adjustments need to be made for all major known risk factors and confounders.

Remarks:

- Confounding by indication
- Immortal time bias
- Time-varying exposure

Manuscripts
https://www.equator-network.org/



Manuscripts
https://www.equator-network.org/

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported
Objectives	3	State specific objectives, including any prespecified hypotheses
Methods		
Study design	4	Present key elements of study design early in the paper
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group
Bias	9	Describe any efforts to address potential sources of bias
Study size	10	Explain how the study size was arrived at
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding
		(*b*) Describe any methods used to examine subgroups and interactions
		(*c*) Explain how missing data were addressed
		(*d*) If applicable, explain how loss to follow-up was addressed
		(*e*) Describe any sensitivity analyses

https://www.equator-network.org/

Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed
		(b) Give reasons for non-participation at each stage
		(c) Consider use of a flow diagram
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders
		(b) Indicate number of participants with missing data for each variable of interest
		(c) Summarise follow-up time (eg, average and total amount)
Outcome data	15*	Report numbers of outcome events or summary measures over time
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included
		(*b*) Report category boundaries when continuous variables were categorized
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses
Discussion		

https://www.equator-network.org/

Many study designs and specialties available

Collaboration vs. consultation

What is the intellectual contribution?

Who makes decisions?
Who influences the details of the work?

International Committee of Medical Journal Editors (ICMJE)

https://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html; accessed 10/26/2022

International Committee of Medical Journal Editors (ICMJE)

2. Who Is an Author?

The ICMJE recommends that authorship be based on the following 4 criteria:

- Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; AND
- Drafting the work or revising it critically for important intellectual content; AND
- Final approval of the version to be published; AND
- Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Non-author contributions (by themselves): technical editing, language editing, proofreading, measurement collection (unless specialized)

https://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html; accessed 10/26/2022

How many co-authors can you have?

Current “record” seems to be a physics paper with 5,154 authors

Combined Measurement of the Higgs Boson Mass in pp Collisions at √s=7 and 8 TeV with the ATLAS and CMS Experiments
G. Aad et al. (ATLAS Collaboration, CMS Collaboration)
Physical Review Letters 114, 191803 – Published 14 May 2015

33 pages
– 9 for research findings (incl. references)
- 24 listing authors and their institutions

Two research groups at the European Organization for Nuclear Research (CERN) pooled their data from the Large Hadron Collider (LHC) to publish the so far most precise estimate of the mass of the Higgs boson (theorized in 1964 and “discovered” at LHC 2011-2013).