MSACL Early Career Network
26 February 2021

Scientific Writing

Alicia Williams, PhD
alicia.williams@rutgers.edu
Graduate Writing Program
Rutgers University

Agenda

What characterizes good scientific writing?

Revision workshop: sentences and paragraphs

Discussion: organizing manuscript sections

Q&A

How would you describe good scientific writing?

At the sentence level?

- Clear, exact
- Concise
- From Belcher 2019: "vigorous," "potent," "dynamic"

At the paragraph level?

- Cohesive (flows)
- Coherent (thematically unified)
- Exciting, but also easy to skim

At the section level?

- Title
- Abstract
- Intro
- Methods
- Results
- Discussion
- Conclusion

Each plays its role well.

At the manuscript level?

- Cohesive (flows)
- Coherent (thematically unified)
- Exciting, but also easy to skim
- Tells a compelling story

The story principle: “Science is a story. Tell it.”

At the sentence level?

- Tells a compelling story

At the paragraph level?

- Tells a compelling story

At the section level?

- Tells compelling story

At the manuscript level?

- Tells a compelling story

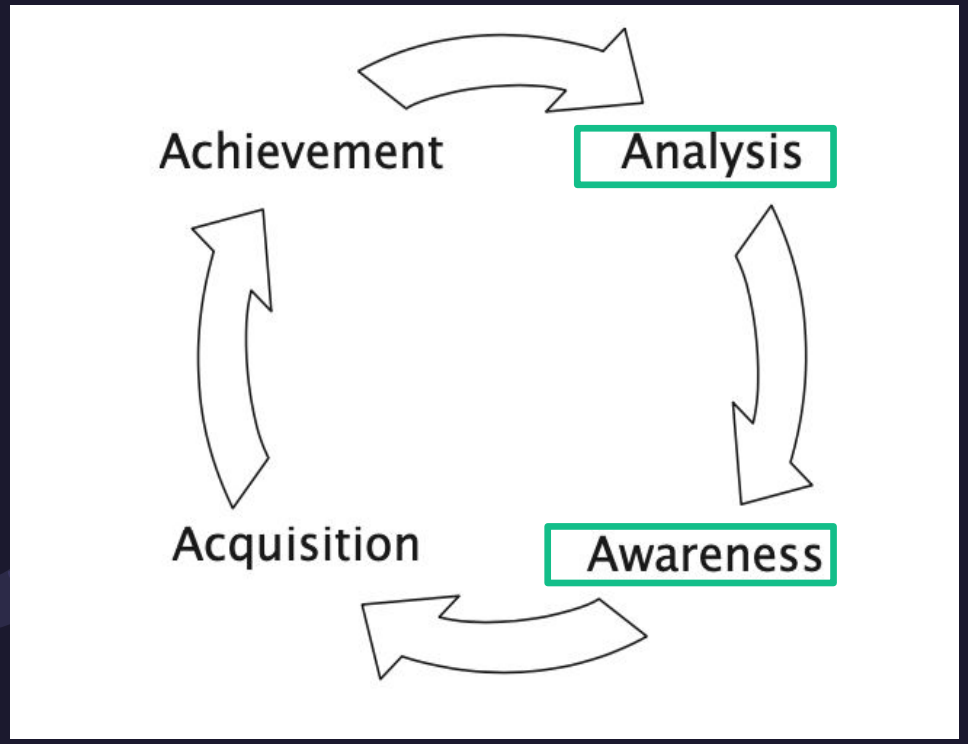
Stories allow you to see the forest through the trees.

"Remember [as you write] that everyone, even a scientist, thinks in narrative. Science is a story. Tell it" (Wells 2004 p. 757-58)

"[S]cience is not data. Data are the raw material of science. It is what you do with data that is science—the interpretation you make, the story you tell" (Zeiger 2000 p. 1)

Revision workshop method

- Examples come from *anonymized* recent publications, provided with titles for context; please do not recirculate the examples.
- Aim is to apply strategies learned via published work to works-in-progress.



"Rhetorical Consciousness Raising Cycle"(Swales and Feak 2009, p. ix)

Sentence-level storytelling

General principle: focus on subject-verb-completer relationships, which should mirror topic-to-stress "psychical geography" of sentence (Williams 2009).

In English, reading comprehension *depends on* knowing the relationship between the subject and predicate (Swan and Gopen 1990).

Subject of sentence (does or receives the action)

- Should be simple, short, prompt, and close to verb
- Should be a recognizable "character," where possible (Williams 2009)
- Avoid *nominalizations* where possible, especially in noun clusters
    - "Vigorous": "Don't use a noun when you could use a verb."

Verb ("predicate," does the action)

- "Potent": choose strong verbs over weak verbs
- "Dynamic": generally choose active voice over passive voice, except when actor is unknown, unimportant, or should be diminished

Completer (objects, direct and indirect)

- Introduce stress, complex vocabulary and points here.

Nominalizations, i.e., "Zombie Nouns"

(Sword 2012)

- Action becomes part of noun, rather than verb. It is "nominalized," becomes inert.
- Deadening effect exacerbated by tendency to take "uninspiring verb" (Sword), especially inactive to-be verbs: is, are, am, was, were, be, being, been.
- Agent of the activity—a would-be actor in the story—can become unclear. *No* agents = no intervention.
- MS has a wide variety of nominalizations! Take note, and <u>try to avoid in in clusters, and in new subject positions.</u>

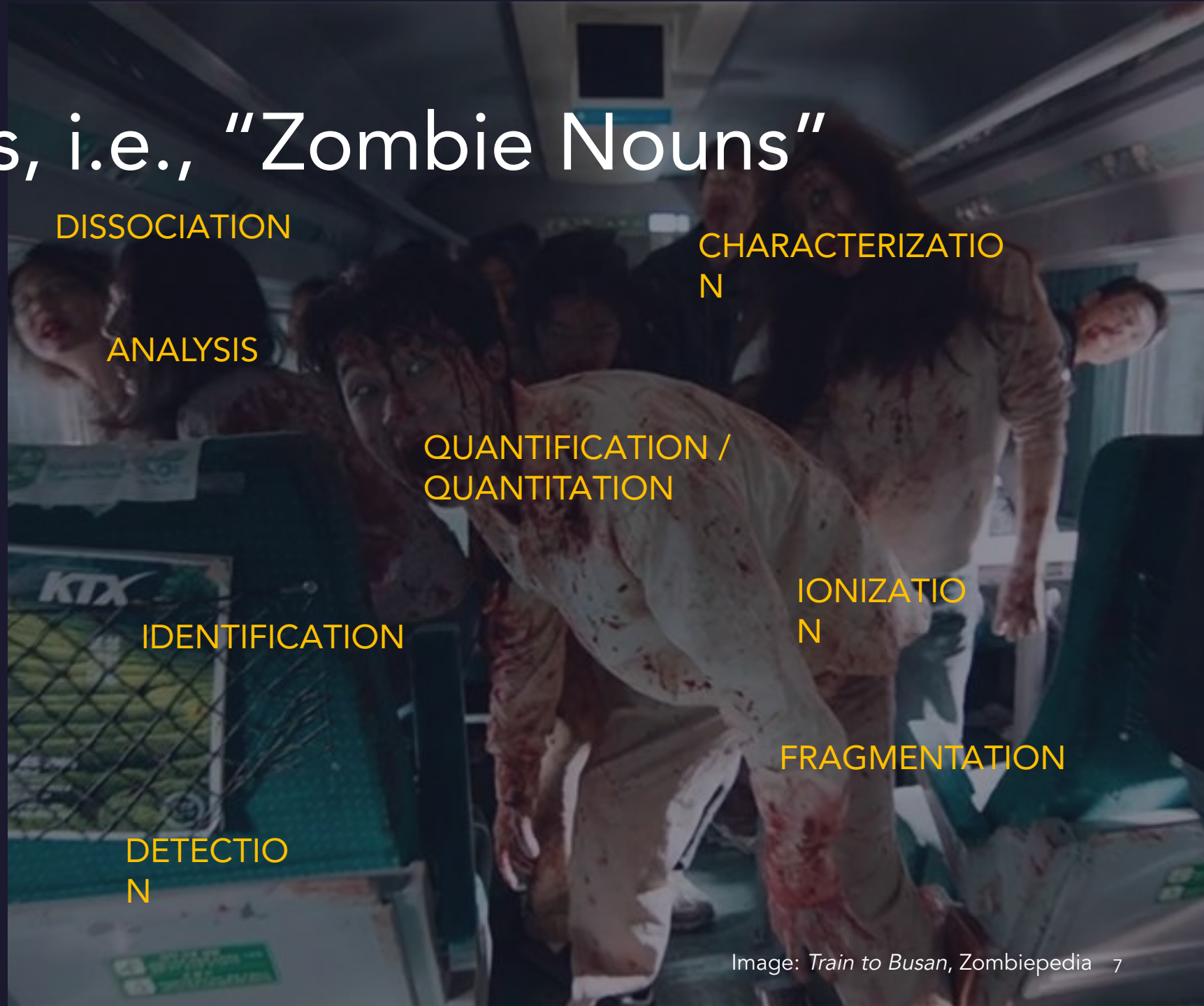


Image: *Train to Busan*, Zombiepedia

Revision examples

Select one (or two!) examples to revise; share to Zoom chat as we address that example.

Example 1

As proof-of-concept, the ability of the TDP-43 MRM LC–MS/MS assay to detect endogenous TDP-43 in human brain tissue (immunohistochemical-confirmed FTLD-TDP type A and an unaffected control) was assessed.

Title: Detection and characterization of TDP-43 in human cells and tissues by multiple reaction monitoring mass spectrometry

Example 2

When reference materials cannot be part of the biological sample, a separate QC sample, e.g. a spiked protein mixture of known composition, or a cell lysate or other mixture that mimics the biological material of interest, is interleaved between the biological samples.

Title: MSstatsQC: Longitudinal System Suitability Monitoring and Quality Control for Targeted Proteomic Experiments

Example 3

Application of this approach to clearly identify clinically-relevant variants among several species is presented for KPC-2, KPC-3, KPC-4 and KPC-5.

Title: Direct detection of intact *Klebsiella pneumoniae* carbapenemase variants from cell lysates: Identification, characterization and clinical implication

Revision Example 1

ORIGINAL

As proof-of-concept, the ability of the TDP-43 MRM LC–MS/MS assay to detect endogenous TDP-43 in human brain tissue (immunohistochemical-confirmed FTLD-TDP type A and an unaffected control) was assessed.

Words between main subject and verb: 20+

REVISION

For proof-of-concept, we assessed the ability of TDP-43 MRM LC–MS/MS assay to detect endogenous TDP-43 in human brain tissue (immunohistochemical-confirmed FTLD-TDP type A and an unaffected control).

Words between main subject and verb: 0

Revision Example 2

ORIGINAL

When reference materials cannot be part of the biological sample, a separate QC sample, e.g. a spiked protein mixture of known composition, or a cell lysate or other mixture that mimics the biological material of interest, is interleaved between the biological samples.

REVISION

When reference materials cannot be part of the biological sample, a separate QC sample is interleaved between them: e.g., a spiked protein mixture of known composition, or a cell lysate or other mixture that mimics the biological material of interest.

Words between main subject and verb: 20+

Words between main subject and verb: 0

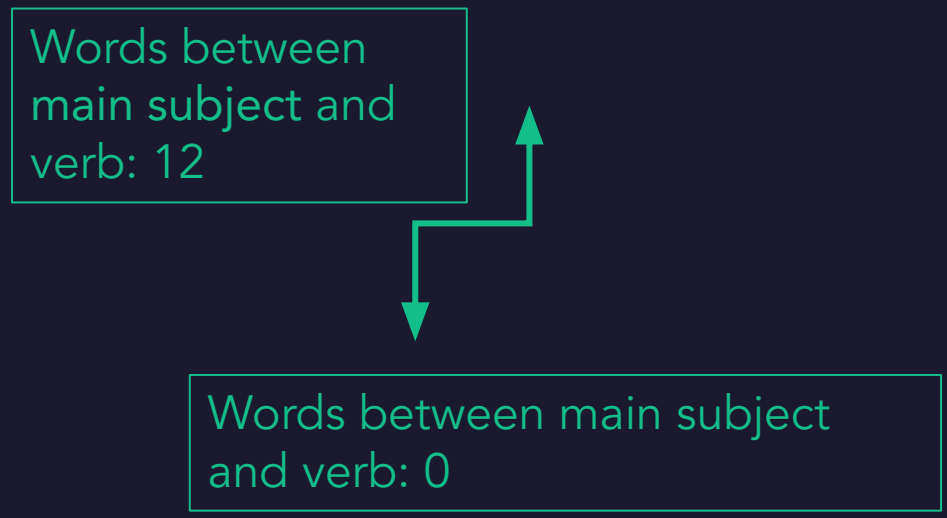
Revision Example 3

ORIGINAL

Characterization of the mature KPC protein revealed an unexpected signal peptide cleavage site preceding an AXA signal peptide motif, modifying the molecular weight (MW) of the mature protein. Taking the additional AXA residues into account allowed for direct detection of the intact protein using top-down proteomic methods. Further validation was performed by transforming a KPC-harboring plasmid into a negative control strain, followed by MS detection of the KPC variant from the transformed cell line. Application of this approach to clearly identify clinically-relevant variants among several species is presented for KPC-2, KPC-3, KPC-4 and KPC-5.

REVISION

We applied this approach to clearly identify clinically relevant variants among several species for KPC-2, KPC-3, KPC-4, and KPC-5.



More extreme cases: no main subject

ORIGINAL

Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometry, native top-down MS with various fragmentation methods, including electron capture dissociation (ECD), collisional activated dissociation (CAD), and multistage tandem MS (MS^3), deduced the binding sites of cobalt and manganese to the C-terminal region of the protein.

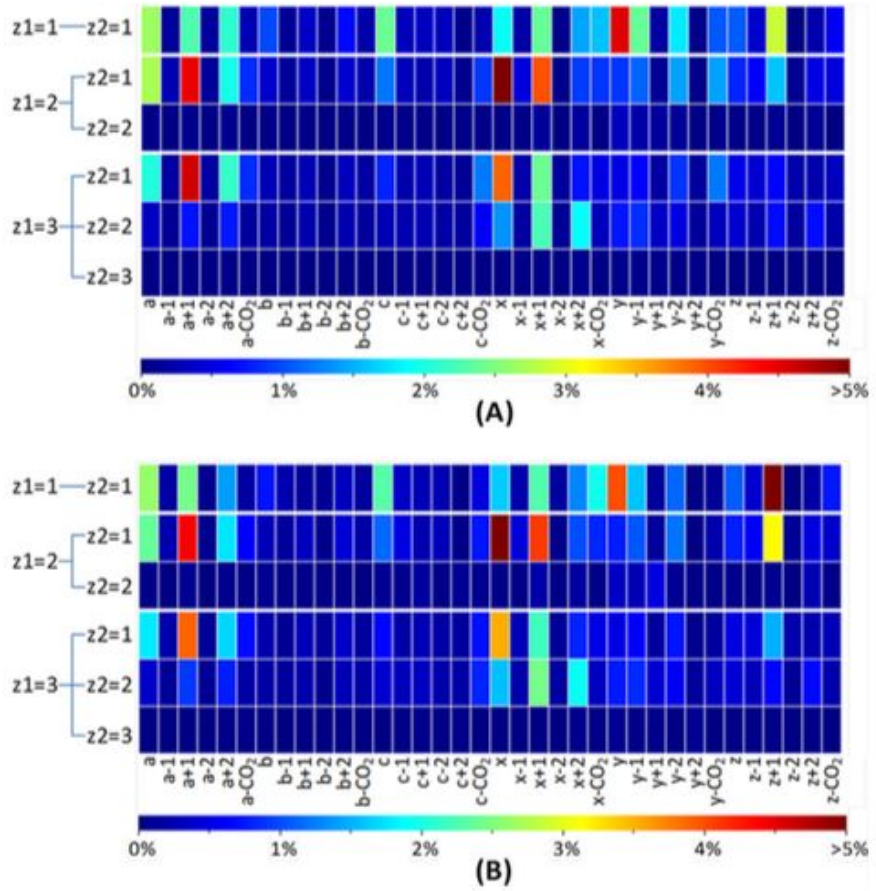
REVISION

To deduce the binding sites of cobalt and manganese to the C-terminal region of the protein, we performed top-down native mass spectrometry with various fragmentation methods on a Fourier transform ion cyclotron resonance (FT-ICR) mass analyzer.

Tip: don't let lists occupy subject position

More extreme cases: no verb

FIG. 1. **The fragmentation behavior of tryptic HeLa and Halo peptide anions as determined via UVPD from the optimized search algorithm in MassMatrix: ion intensities of various product ions normalized by total ion intensities of the spectrum (after precursor and intact charge-reduced peak filtering) for tryptic peptide matches with FDR < 1% from (*A*) Halo and (*B*) HeLa proteome samples.** For the Halo proteome sample, 14,802 peptide spectral matches (PSMs) from 3663 unique peptides were identified. Among them, 726 PSMs from 328 unique peptides were singly charged, 12,112 PSMs from 3121 unique peptides were doubly charged, and 1964 PSMs from 617 unique peptides were triply charged. For the HeLa proteome sample, there were 7886 PSMs from 2349 unique peptides identified. Among them, 232 PSMs from 121 unique peptides were singly charged, 5972 PSMs from 1991 unique peptides were doubly charged, and 1682 PSMs from 568 unique peptides were triply charged. Only one representative PSM for each unique peptide with a given charge state was used in plotting the heat maps.



Original

The fragmentation behavior of tryptic HeLa and Halo peptide anions as determined via UVPD from the optimized search algorithm in MassMatrixned: ion intensities of various product ions normalized by total ion intensities of the spectrum (after precursor and intact charge-reduced peak filtering) for tryptic peptide matches with FDR <1% for *(A)* Halo and *(B)* HeLa proteome samples.

Revision

For *(A)* Halo and *(B)* HeLa proteome samples, UVPD from the optimized search algorithm in MassMatrix determined the fragmentation behavior of Halo and HeLa peptide anions, respectively. Ion intensities of various product ions were normalized by total ion intensities of the spectrum (after precursor and intact charge-reduced peak filtering) for tryptic peptide matches with FDR <1%

Image: MS C 624, box 2, Bernadine Healy Papers, NIH

Sentence-level storytelling takeaway: “strong verbs, short sentences.”

—Bernadine Healy, first female president of NIH

Paragraph-level storytelling

Three main principles: cohesion, coherence, structure

Cohesion: flow

- Transitions: interpretive transitions are important in Introduction and Discussion
- Use “bidirectional continuity” (Zeiger 2000): end of sentence *A* re-appears in short form at beginning of sentence *B*
- Move from old to new information

Coherence: thematic unity

- Repeat *exact* key terms, ideally in “topic” positions (early in sentences)
- Decide on *one* key message per paragraph
    - The message should represent both a topic (of the research) and a purpose (have a distinct, justified function your paper)

Structure (for ease of skimming, spotting significance vs. detail)

- Frame details that develop message with topic (opening), stress (ending)

Paragraph-level storytelling

Cohesion: flow

- Use “bidirectional continuity” (Zeiger 2000): end of sentence *A* re-appears in short form at beginning of sentence *B*

Coherence: thematic unity

- Repeat *exact* key terms, ideally in “topic” positions (early in sentences)

Example from Zeiger 2000, p. 61

Digitalis increases the contractility of the mammalian heart. Changes in the calcium flux through the muscle cell membrane cause this increased contractility.

VS.

Digitalis increases the contractility of the mammalian heart. This change in inotropic state is a result of changes in calcium flux through the muscle cell membrane.

VS.

Digitalis increases the contractility of the mammalian heart. This increased contractility is a result of changes in calcium flux through the muscle cell membrane.

Revision example

A) As proof-of-concept, the ability of the TDP-43 MRM LC–MS/MS assay to detect endogenous TDP-43 in human brain tissue (immunohistochemical-confirmed FTLD-TDP type A and an unaffected control) was assessed.

B) While this methodology has been valuable in the characterization of TDP-43 in disease, it has limited multiplexing capabilities, relies on indirect detection (resulting in a lack of specificity) and provides low-resolution structural information.

C) Common contemporary approaches for characterizing TDP-43 structure in brain tissue include immunohistochemical staining and western blot analyses, both of which are ligand binding methods dependent on antibody-antigen interactions.

D) The availability of a higher resolution method that directly detects the measurand of interest, e.g., MRM LC–MS/MS, would be helpful in routine characterization of tissues, and complement information obtained from immunometric approaches.

Q: In what sequence would these sentences occur in publication?

Answer: C – B - D - A

C) Common contemporary approaches for characterizing TDP-43 structure in brain tissue include immunohistochemical staining and western blot analyses, both of which are ligand binding methods dependent on antibody-antigen interactions. B) While this methodology has been valuable in the characterization of TDP-43 in disease, it has limited multiplexing capabilities, relies on indirect detection (resulting in a lack of specificity) and provides low-resolution structural information. D) The availability of a higher resolution method that directly detects the measurand of interest, e.g., MRM LC–MS/MS, would be helpful in routine characterization of tissues, and complement information obtained from immunometric approaches. A) As proof-of-concept, the ability of the TDP-43 MRM LC–MS/MS assay to detect endogenous TDP-43 in human brain tissue (immunohistochemical-confirmed FTLD-TDP type A and an unaffected control) was assessed.

Storytelling in IMRAD-style Manuscript Sections

ABSTRACT

INTRO

METHODS

RESULTS

DISCUSSION

Abstract


ABSTRACT
INTRO
METHODS
RESULTS
DISCUSSION
Context
Purpose
Common obstacle to story: relevance of purpose, major takeaways left implicit
Impact for field

Abstract examples

First sentence

The purpose of this study was to evaluate whether the risk of *Streptococcus uberis* clinical mastitis at cow level could be predicted from the historical presence of specific strains of *S. uberis* on dairy farms. […]

Title: Prediction of *Streptococcus uberis* clinical mastitis risk using Matrix-assisted laser desorption ionization time of flight mass spectrometry (MALDI-TOF MS) in dairy herds

Last sentence

[…] With this FAIMS setup, 60% of the proteins identified are classified as very low abundant proteins.

FAIMS-MS might contribute to phosphopeptides identification in plasma. Journal of Proteomics

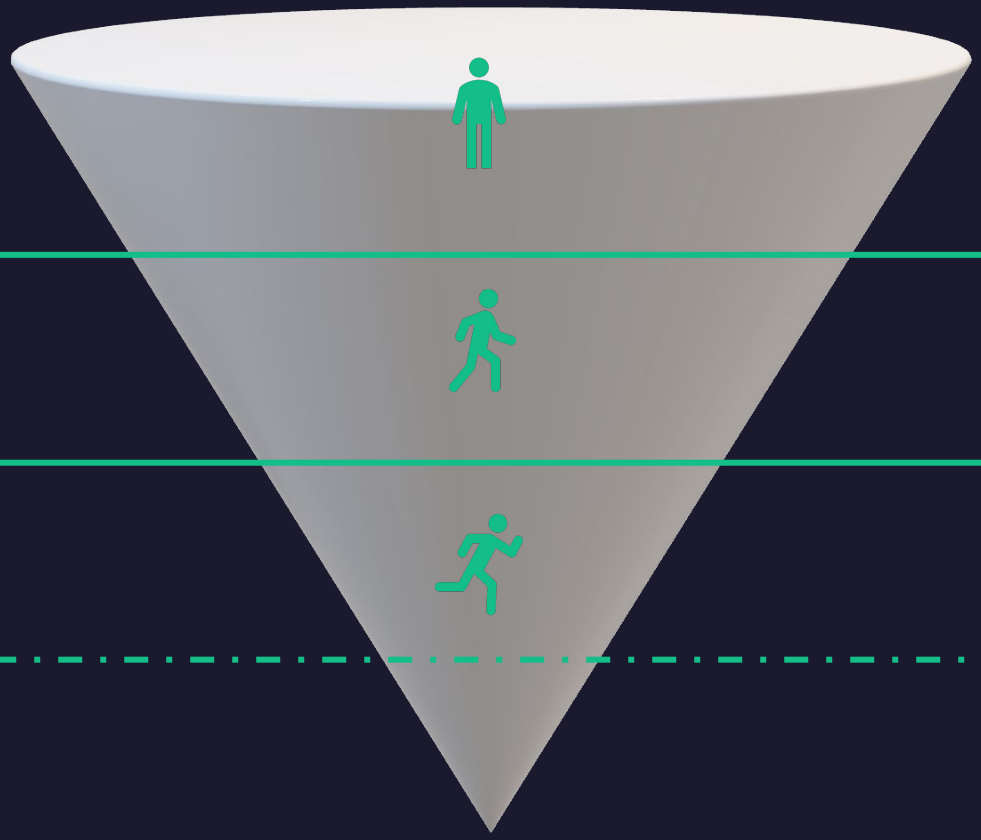
Introduction (4)

Background: known information, via review of lit.

Problem: unknown information, i.e., knowledge gap

Purpose statement, **question** to be answered

Briefly: "plan of attack" to carry out purpose



Introduction summary

Obstacles to story

Murky motivation or direction towards purpose of study

- Lack of cohesion
  - Missing transitions
- Lack of coherence
- Length
  - Overly extensive description and/or literature review

Solutions to obstacles

To shorten (Annesley 2010):

- Consider target readership of journal's background knowledge
- Cite only most relevant, recent literature

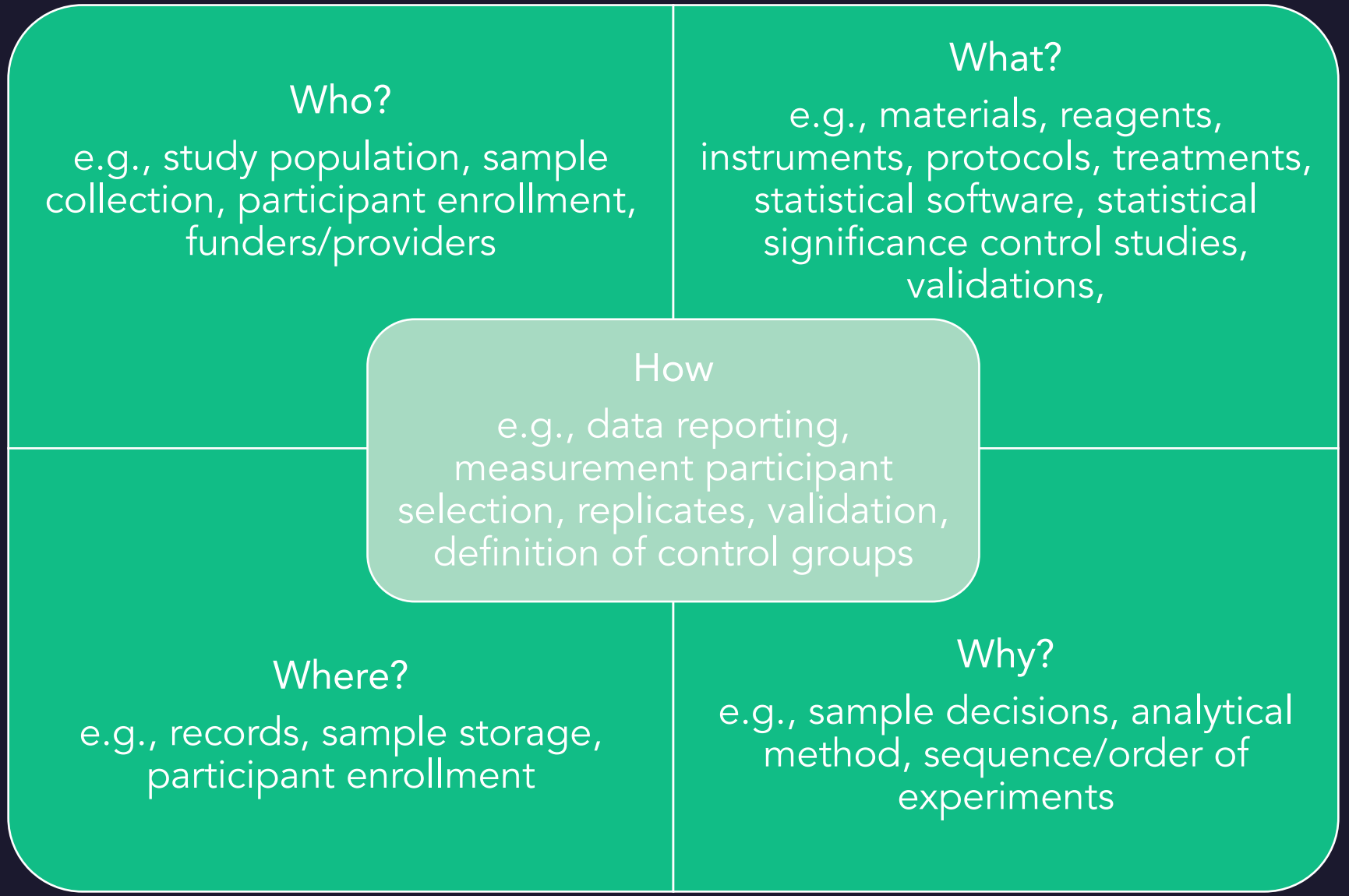
To add cohesion and coherence:

- Add transitions
- Follow logical sentence patterns
- Repeat key terms from known/unknown in the purpose statement

(2-3) Methods

Experimental procedure, materials & methods, protocol

The common "recipe" metaphor points to importance of "how," but other questions matter too for clarity, reproducibility (Annesley 2010).



Results & Discussion

RESULTS

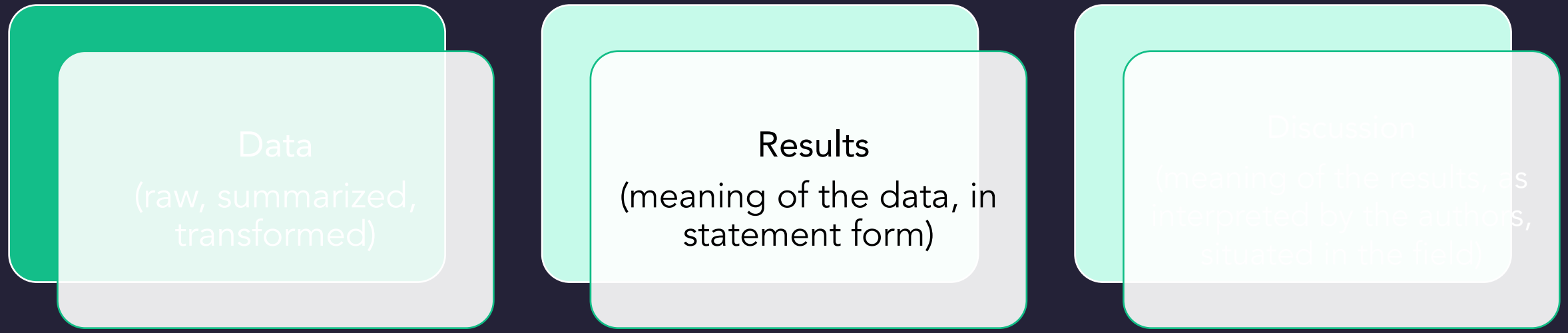
"a clear presentation of the results means that readers should readily be able to see their connection to both the just-described methods and the study's purpose." (Goldbort 2006, p. 231)

DISCUSSION

"The discussion section as a whole essentially is the place for the writer to demonstrate not only critical competence in interpreting and assessing the findings, but also the professional authority necessary to argue for their significance convincingly." (p. 238)

(Data &) Results & Discussion

Zeiger 2000



1-2 2-3 4-5

Discussion (4-5)

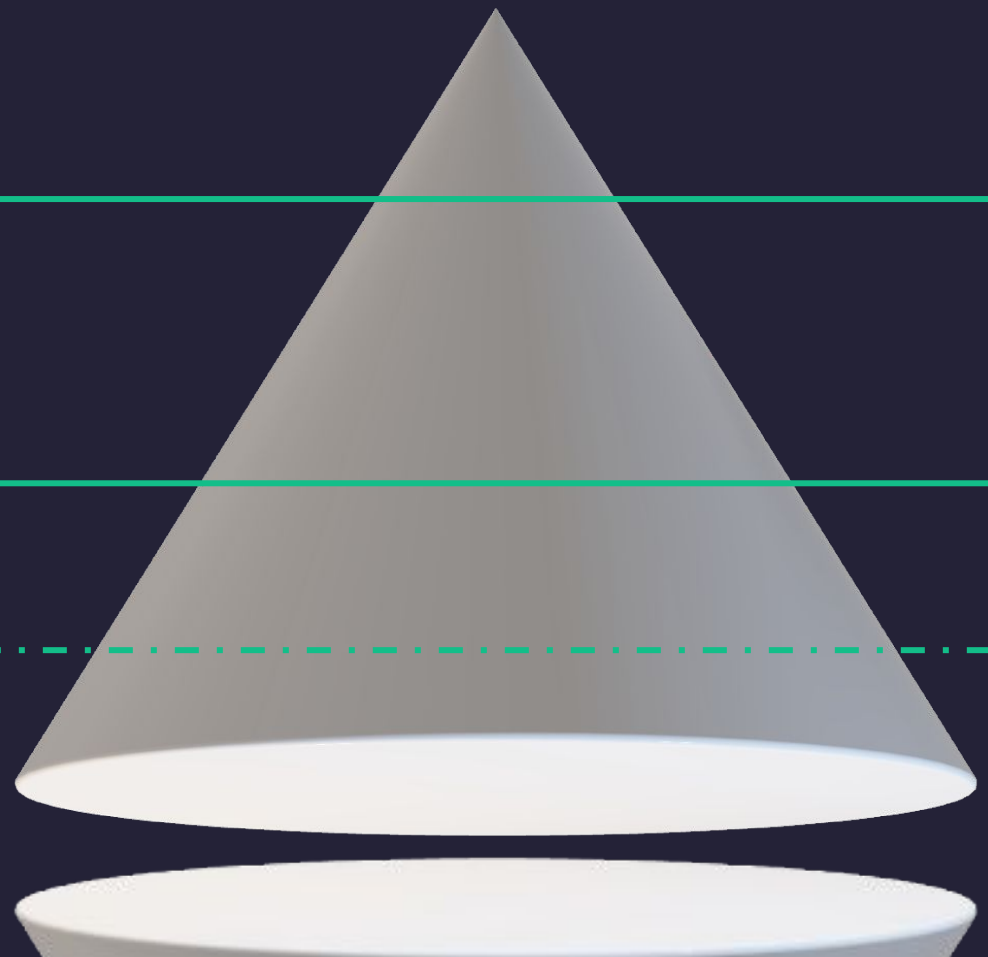
Repeat and answer main question from end of introduction

Interpret meaning of results

- Typically mirror the sequence of their presentation
- Cite other research as needed to illuminate meaning for field
- Acknowledge limitations

Major statement about value added to field

Conclusion (if standard for journal): summary statement, broader implications, future directions



Methods, results, discussion summary

Obstacles to story

Illogical sequence of subsections

The goldilocks problem: too little or too much detail (methods) and repetition (results and discussion)

Too little: "steps" are jumped, creating confusion.

Too much: contribution becomes muddled

Solutions to obstacles

Review examples in target journal.

Have colleagues review this section to identify any jumps or unnecessary detail – ideally from an adjacent field.

Paragraph-level editing.

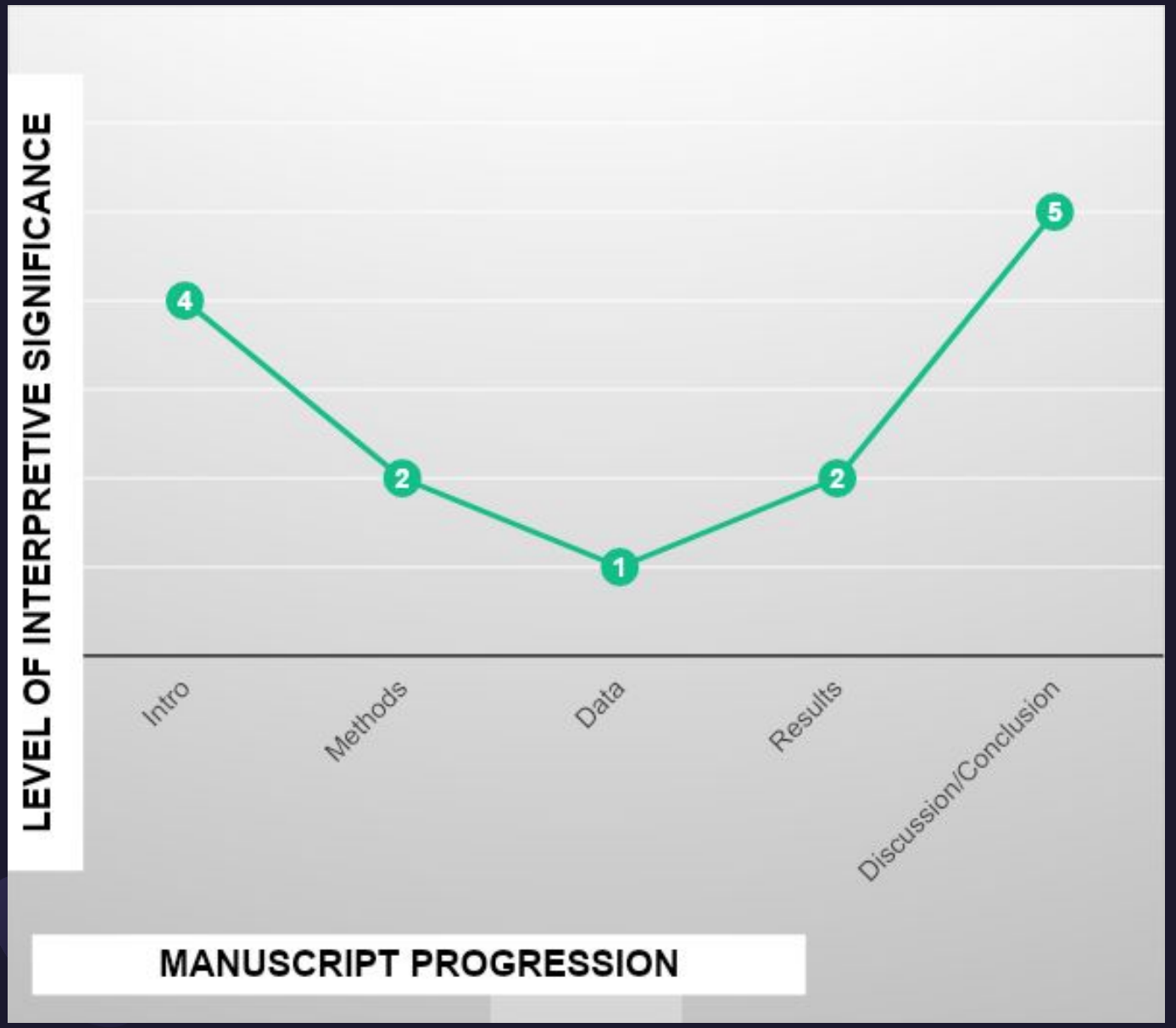
Storytelling at the Manuscript Level

The "UNEVEN U"

(adapted from Hayot 2015)

Unevenness: addressing a major problem (4) by way of a research methodology (2) that produces data (1) framed as results (2) and analyzed in discussion (3-4) to produce new knowledge (5)

Cohesiveness: ties each element of story into an overarching narrative



Q & A

References & suggested resources

- *The ACS Guide to Scholarly Communication. American Chemical Society*, 2020.
- *The CSE Manual for Authors, Editors, and Publishers, 8th edition*. University of Chicago Press, 2014.
- Annesley, T. (2010-2011). Clinical Chemistry Guide to Scientific Writing. *Clinical Chemistry.* (several issues). https://academic.oup.com/clinchem/pages/guide-to-scientific-writing

  ^ Translations in Chinese, Japanese, Spanish, Portuguese, Turkish, and Russian are available.
- Belcher, W.L. (2019). *Writing Your Journal Article in 12 Weeks: A Guide to Academic Publishing Success: second edition.* University of Chicago Press.
- Cameron, C., Lee, H. Y., Anderson, C., Byars-Winston, A., Baldwin, C. D., & Chang, S. (2015). The Role of Scientific Communication Skills in Trainees' Intention to Pursue Biomedical Research Careers: A Social Cognitive Analysis. *CBE life sciences education*, 14(4), ar46. https://doi.org/10.1187/cbe.14-09-0152
- Cameron, C., Collie, C. L., Baldwin, C. D., Bartholomew, L. K., Palmer, J. L., Greer, M., & Chang, S. (2013). The development of scientific communication skills: a qualitative study of the perceptions of trainees and their mentors. Academic medicine : journal of the Association of American Medical Colleges, 88(10), 1499–1506. https://doi.org/10.1097/ACM.0b013e3182a34f36
- Goldbort, R. (2006). *Writing for Science*. New Haven, Yale University Press.

References, suggested resources, & gratitude

- Hayot, Eric. *The Elements of Academic Style: Writing for the Humanities.* New York, Columbia University Press, 2014.
- Knight J. (2003). Scientific literacy: Clear as mud. Nature, 423(6938), 376–378. https://doi.org/10.1038/423376a
- Silyn-Roberts, Heather. *Writing for Science and Engineering.* Amsterdam, Elsevier, 2013.
- Swales, J. and Feak, C. (2012). *Academic Writing for Graduate Students.* Ann Arbor, University of Michigan Press.
- Swan, J., and Gopen, G. (1990). The Science of Scientific Writing. *American Scientist* 78(6), 550–558.
- Sword, Helen. "Zombie Nouns." *New York Times,* 23 July 2012.
- Wells WA. Me write pretty one day: how to write a good scientific paper. J Cell Biol. 2004 Jun 21;165(6):757-8. doi: 10.1083/jcb.200403137
- Williams, J. (2009). *Style: Lessons Toward Clarity and Grace, 11th edition.* Pearson.
- Zeiger, M. (2000). *The Essentials of Writing Biomedical Research Papers.* McGraw Hill.

Many thanks to Hoda Safari Yazd, MSACL Early Career Network Lead, America Region, and all organizers at MSACL Early Careers Network.