SCIENTIFIC WRITING

Chang-hee Won, Ph.D.
Temple University
Czech Technical University Fulbright Scholar
Summer 2019

Courtesy of Nicole Kelley of MIT, Fall 2006
Lecture materials derived from *The Craft of Scientific Writing* © Michael Alley and "The Science of Scientific Writing" by Gopen and Swan
Courtesy of Nathan Sheffield, Duke University
Courtesy of Jeff Frolik, University of Massachusetts
Courtesy of Marc Tischler, University of Arizona

ANNOUNCEMENTS

- Explain archiving sites such as arXiv.org: Cornell University.
- Go over syllabus. Explain grading
- Last week presentation. Keep May 28th open
- Journal Citation Index: https://knihovna.cvut.cz/o-nas/archiv/blog/266-incites-a-jcr-meni-odkazy

What is the differences between scientific writing and other writing?

“The fundamental purpose of scientific discourse is not the mere presentation of information and thought but rather its actual communication. It does not matter how pleased an author might be to have converted all the right data into sentences and paragraphs; it matters only whether a large majority of the reading audience accurately perceives what the author had in mind.”

– George Gopen and Judith Swan
The Science of Scientific Writing

SCIENTIFIC WRITING

- Scientific writing is a subset of academic writing. Academic writing is formal, objective, and concise in comparison to other styles of writing.

- Scientific writing is an academic writing for people in sciences, mathematics, and engineering.

- The scientific paper is a written and published report describing original research results.

- A well-written scientific paper explains the scientist's motivation for doing an experiment, the experimental design and execution, and the meaning of the results. Scientific papers are written in a style that is exceedingly clear and concise.

	where it is used	language	slang	contraction	reference
General writing	letters, email, newspaper articles, diary, and journal	informal or semi-formal language	can use slang or jargon	can use contraction	does not usually use references or citations
Scientific writing	research projects, conferences, reports	use formal, objective, concise language	does not use slang nor jargon	does not use contraction	always use references and citations

SCIENTIFIC WRITING FLOW

- Planning
- Clarity
- Brevity
- Simplicity
- Word Choice
- Active Voice
- Committing to Writing as a Process

PLANNING: BEFORE YOU BEGIN

- Identify the audience
- Know your purpose and material
- Understand the writing task at hand
- Organize your thoughts and materials
- Budget adequate time to write, review, revise, and edit

CLARITY: AVOID JARGON

- Jargon: a vocabulary particular to a place of work (abbreviations, slang)
- Audience familiarity with the topic determines appropriate use of jargon
- Ex. 1: For the first year, the links with SDPC and the HAC were not connected, and all required OCS input data were artificially loaded. Thus CATCH22 and MERWIN were not available.
- **Ex. 2: Because some of the links in the computer system were not connected the first year, we could not run all the software codes.**

CLARITY: DEFINE THE UNFAMILIAR

- If you must abbreviate, define the term in its first occurrence, and put abbreviations in parentheses

- Ex: Edgartown Great Pond (EGP) is a vital body of water. Unfortunately, due to an unpredictable influx of saltwater, the delicate ecosystem is in danger of destabilizing.

- Italicize first occurrence of unfamiliar terms and define them right away

- Ex: *Retina* is a light-sensitive tissue, found at the back of the eye, that converts light impulses to nerve impulses.

CLARITY: PUT SUBJECT AND VERB CLOSE

- English readers expect doers to be near their actions.
- Complex subjects (subjects modified with essential clauses) can violate this expectation
- Ex. 0 The man who had followed us inside walked over to the telephone.
- Ex. 1. The assumption that all RNAs are polyadenylated is an oversimplification of the transcription process.
- Ex. 2. The model oversimplifies the transcription process because it assumes that all RNAs are polyadenylated.

CLARITY: PUT FAMILIAR IDEA FIRST

Old Information vs. New information

Farmers try to provide optimal growing conditions for crops by using soil additives to adjust soil pH. Garden lime, or agricultural limestone, is made from pulverized chalk, and can be used to raise the pH of the soil. Clay, which is a naturally acidic soil type, often requires addition of agricultural lime.

Farmers try to provide optimal growing conditions for crops by using soil additives to adjust soil pH. One way to raise the pH of the soil is an additive made from pulverized chalk called garden lime or agricultural limestone. Agricultural limestone is often added to naturally acidic soils, such as clay.

BREVITY: USE WORDS EFFICIENTLY

- Never use two words when one word will do.

- Ex. 1: The relationship between the nature of salt water to fresh water in the Edgartown Great Pond that fluctuates often is extremely important to everyone including scientists, residents, and environmentalists on Martha's Vineyard.

- **Ex. 2: The fluctuating salinity of EGP concerns many environmentalists, scientists, and residents.**

BREVITY: LESS IS MORE

- Trim your language down to the essential message you want to get across to your readers:
- Ex: Earthquakes can occur at predictable intervals along a given fault segment. Depending on the length and slip in each main shock, the exact interval can vary by a factor of two.
- Ex: The southern segment of the San Andreas fault has an interval of 145 years, plus or minus a few decades

BREVITY: MOST IMPORTANT FIRST

- Place key information in the main clause
- Ex. 1: Despite winning the game, the Patriots made several errors in the first half.
- Ex. 2: Despite making several errors in the first half, the Patriots won the game.
- **Ex. 3: The Patriots won the game, despite making several errors in the first half.**

BREVITY: REMOVE REDUNDANCY

- Combine overlapping sentences when possible
- Ex. 1: Water quality in Hawk River declined in March. This decline occurred because of the heavy rainfall that month. All the extra water overloaded Tomlin county's water treatment plant.
- **Ex. 2: Water quality in Hawk River declined in March because heavy rainfalls overloaded Tomlin County water treatment plant.**

SIMPLICITY: USE DETAILS WISELY

- Specific details are desirable, but be careful to balance detail with audience needs for clarity—significance is more important.
- Ex. 1: The average house in the area has a radon level of 0.4 picocuries per litre.
- **Ex. 2: The average house in the area has a radon level of 0.4 picocuries per litre, which is considered low by the EPA [Lafavore, 1987]. Levels between 20 and 200 picocuries per liter are considered high, and levels above 200 picocuries per liter are considered dangerous. For reference, the average radon level in outdoor air is about 0.2 pico curies per litre.**

SIMPLICITY: USE DETAILS WISELY

- Many engineers want to provide as much specific detail as possible, but this can come at the expense of readers understanding and their main point

- Ex. 1: The number of particular hydrocarbon combinations in our study is enormous. For example, the number of possible $C_{20}H_{42}$ is 366,319 and the number of $C_{40}H_{82}$ is 62,491,178,805,831.

- **Ex. 2: The number of hydrocarbon combinations in our study is enormous. For example, the number of possible $C_{40}H_{82}$ is over 60 trillion.**

LANGUAGE: NEEDLESS COMPLEXITY

Category	Example	Substitute
nouns	utilization functionality	**use** **feature**
verbs	facilitate finalize	**cause** **end**
adjectives	aforementioned individualized	**mentioned** **individual**
adverbs	firstly, secondly, heretofore	**first, second** **previous**

LANGUAGE: ABSTRACTION

- Avoid too many abstract nouns (idea, quality, or state)
- Ex. 1: The existing nature of Mount St. Helens' volcanic ash spewage was handled through the applied use of computer modeling capabilities.
- Ex. 2: With Cray computers, we modeled how much ash spewed from Mount St. Helens.

LANGUAGE: NEEDLESS WORDS

(already) existing	never (before)
at (the) present (time)	none (at all)
(basic) fundamentals	now (at this time)
(completely) eliminate	period (of time)
(continue to) remain	(private) industry
currently (being)	(separate) entities
(currently) underway	start (out)
(empty) space	write (out)
had done (previously)	(still) persists
introduced (a new)	
mix (together)	

LANGUAGE: AMBIGUITY

- Choose words whose meanings are clear
- Ex. 1: T cells, rather than B cells, appeared as the lymphocytes migrated to the thymus gland.
- Ex. 2: T cells, rather than B cells, appeared because the lymphocytes migrated to the thymus gland.

LANGUAGE: AMBIGUITY

- Order the words in your sentences carefully
- Ex. 1: In low water temperatures and high toxicity levels of oil, we tested how well the microorganisms survived.
- **Ex. 2: We tested how well the microorganisms survived in low water temperatures and high toxicity levels of oil.**

LANGUAGE: AMBIGUITY

- Do not overuse pronouns—particularly "it" and "this"—because it is often difficult to identify the antecedent
- Ex: Because the receiver presented the radiometer with a high-flux environment, it was mounted in a silver-plated stainless steel container.
- **Because the receiver presented the radiometer with a high-flux environment, the receiver was mounted in a silver-plated stainless steel container.**

LANGUAGE: WEAK VS. STRONG

- Avoid too many "to be" verbs
  "is" "was" "were" "has been" "have been"

- Avoid excess words, which slow comprehension of the main point

made arrangements for	arranged
made the decision	decided
made the measurement of	measured
performed the development of	developed
is working as expected	works as expected

ACTIVE VOICE: STRONG VERBS

- Technical writers want to communicate as efficiently as possible, and active voice is more straightforward and is stronger than passive voice

- Ex 1: The feedthrough was composed of a sapphire optical fiber, which was pressed against the pyrotechnic that was used to confine the charge.

- **Ex 2: The feedthrough contained a sapphire optical fiber, which pressed against the pyrotechnic that contained the charge.**

PASSIVE VOICE

- Passive voice has several effects:
- Reverses the order of the sentence (A-B vs. B-A)
  - I stole the money
  - The money was stolen by me
- It can eliminate the actor (causing ambiguity)
- It often increases length
- Most scientific journals encourage authors to use active voice for the sake of clarity, conciseness, and cohesion!

PASSIVE VOICE

- "But using active voice means using first person pronouns, and first person pronouns aren't allowed in my field!"
- But compare:
  - The substrate surface was mapped using an Atomic Force Microscope.
  - We mapped the substrate surface using an AFM.
  - The AFM mapped the substrate surface.

ACTIVE VOICE: NATURAL SOUND

- When in doubt, read passages out loud to determine the natural sound*
- Ex 1: A new process for eliminating nitrogen oxides from diesel exhaust engines is presented. Flow tube experiments to test this process are discussed. A chemical reaction scheme to account for this process is proposed.
- **Ex 2: We present a new process for eliminating nitrogen oxides from the exhaust of diesel engines. To test this process, we performed experiments in flow tubes. To explain this process, we developed a scheme of chemical reactions.**

*always defer to your professor, your journal, or your company style guide for use of "I" and "we" in technical papers

PRESENTING RESULTS

PRESENTING RESULTS

- YOUR WORK IS OF A QUANTITATIVE NATURE AND PROBABLY THE BEST WAY TO CONCISELY SUMMARIZE YOUR WORK WILL INVOLVE. Not Text.
  - Images
  - Drawings
  - Equations
  - Tables
  - Plots

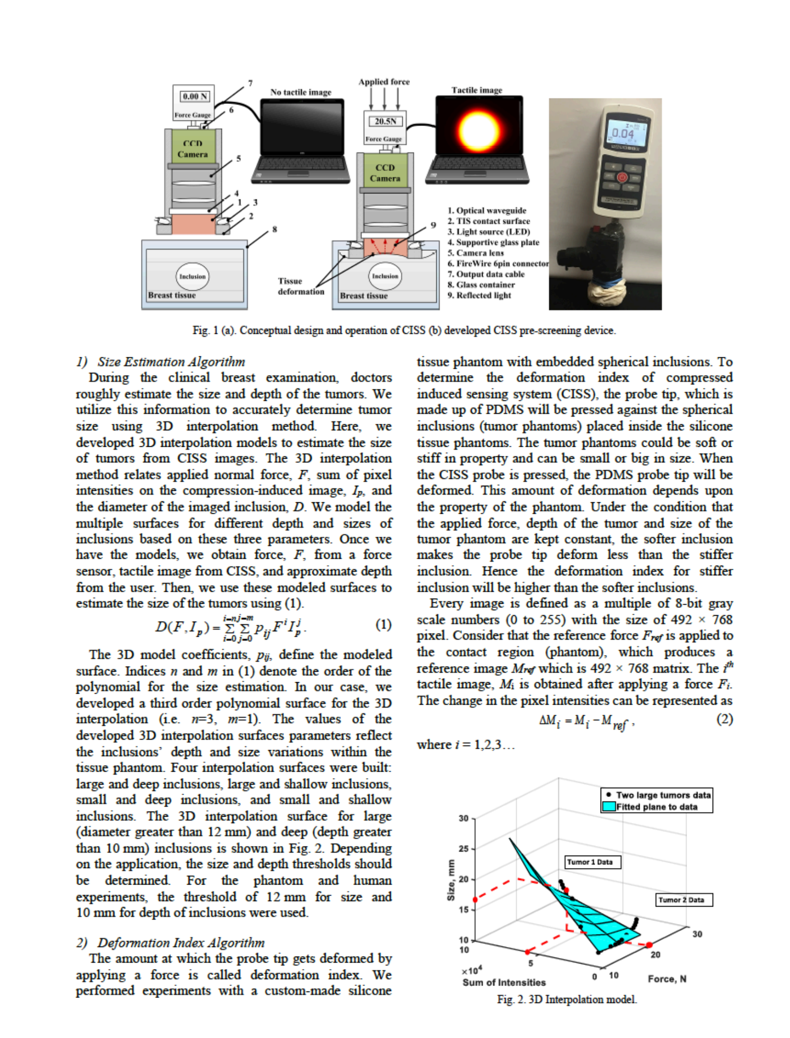


Fig. 1 (a). Conceptual design and operation of CISS (b) developed CISS pre-screening device.

1) Size Estimation Algorithm

During the clinical breast examination, doctors roughly estimate the size and depth of the tumors. We utilize this information to accurately determine tumor size using 3D interpolation method. Here, we developed 3D interpolation models to estimate the size of tumors from CISS images. The 3D interpolation method relates applied normal force, F, sum of pixel intensities on the compression-induced image, I_p, and the diameter of the imaged inclusion, D. We model the multiple surfaces for different depth and sizes of inclusions based on these three parameters. Once we have the models, we obtain force, F, from a force sensor, tactile image from CISS, and approximate depth from the user. Then, we use these modeled surfaces to estimate the size of the tumors using (1).

$$D(F, I_p) = \sum_{i=0}^{i=n} \sum_{j=0}^{j=m} p_{ij} F^i I_p^j . \quad (1)$$

The 3D model coefficients, p_{ij}, define the modeled surface. Indices n and m in (1) denote the order of the polynomial for the size estimation. In our case, we developed a third order polynomial surface for the 3D interpolation (i.e. n=3, m=1). The values of the developed 3D interpolation surfaces parameters reflect the inclusions' depth and size variations within the tissue phantom. Four interpolation surfaces were built: large and deep inclusions, large and shallow inclusions, small and deep inclusions, and small and shallow inclusions. The 3D interpolation surface for large (diameter greater than 12 mm) and deep (depth greater than 10 mm) inclusions is shown in Fig. 2. Depending on the application, the size and depth thresholds should be determined. For the phantom and human experiments, the threshold of 12 mm for size and 10 mm for depth of inclusions were used.

2) Deformation Index Algorithm

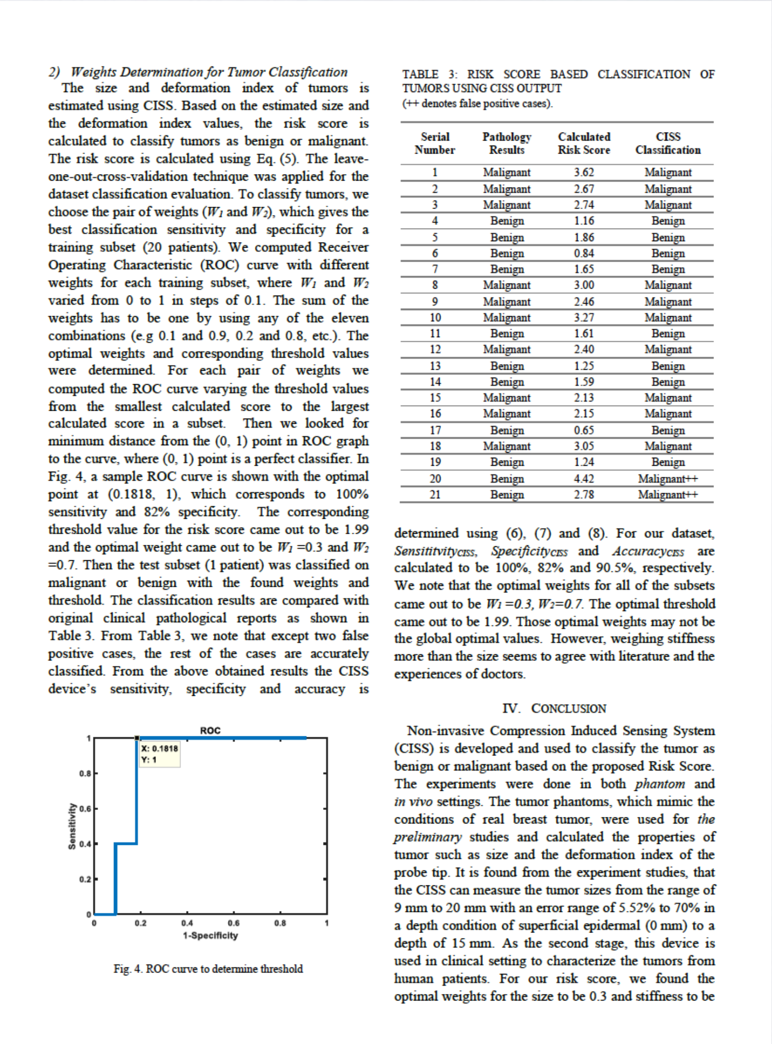
The amount at which the probe tip gets deformed by applying a force is called deformation index. We performed experiments with a custom-made silicone tissue phantom with embedded spherical inclusions. To determine the deformation index of compressed induced sensing system (CISS), the probe tip, which is made up of PDMS will be pressed against the spherical inclusions (tumor phantoms) placed inside the silicone tissue phantoms. The tumor phantoms could be soft or stiff in property and can be small or big in size. When the CISS probe is pressed, the PDMS probe tip will be deformed. This amount of deformation depends upon the property of the phantom. Under the condition that the applied force, depth of the tumor and size of the tumor phantom are kept constant, the softer inclusion makes the probe tip deform less than the stiffer inclusion. Hence the deformation index for stiffer inclusion will be higher than the softer inclusions.

Every image is defined as a multiple of 8-bit gray scale numbers (0 to 255) with the size of 492 × 768 pixel. Consider that the reference force F_{ref} is applied to the contact region (phantom), which produces a reference image M_{ref} which is 492 × 768 matrix. The i^{th} tactile image, M_i is obtained after applying a force F_i. The change in the pixel intensities can be represented as

$$\Delta M_i = M_i - M_{ref} , \quad (2)$$

where $i = 1,2,3\ldots$

Fig. 2. 3D Interpolation model.



2) Weights Determination for Tumor Classification

The size and deformation index of tumors is estimated using CISS. Based on the estimated size and the deformation index values, the risk score is calculated to classify tumors as benign or malignant. The risk score is calculated using Eq. (5). The leave-one-out-cross-validation technique was applied for the dataset classification evaluation. To classify tumors, we choose the pair of weights (W_1 and W_2), which gives the best classification sensitivity and specificity for a training subset (20 patients). We computed Receiver Operating Characteristic (ROC) curve with different weights for each training subset, where W_1 and W_2 varied from 0 to 1 in steps of 0.1. The sum of the weights has to be one by using any of the eleven combinations (e.g 0.1 and 0.9, 0.2 and 0.8, etc.). The optimal weights and corresponding threshold values were determined. For each pair of weights we computed the ROC curve varying the threshold values from the smallest calculated score to the largest calculated score in a subset. Then we looked for minimum distance from the (0, 1) point in ROC graph to the curve, where (0, 1) point is a perfect classifier. In Fig. 4, a sample ROC curve is shown with the optimal point at (0.1818, 1), which corresponds to 100% sensitivity and 82% specificity. The corresponding threshold value for the risk score came out to be 1.99 and the optimal weight came out to be W_1 =0.3 and W_2 =0.7. Then the test subset (1 patient) was classified on malignant or benign with the found weights and threshold. The classification results are compared with original clinical pathological reports as shown in Table 3. From Table 3, we note that except two false positive cases, the rest of the cases are accurately classified. From the above obtained results the CISS device's sensitivity, specificity and accuracy is

Fig. 4. ROC curve to determine threshold

TABLE 3: RISK SCORE BASED CLASSIFICATION OF TUMORS USING CISS OUTPUT
(++ denotes false positive cases).

Serial Number	Pathology Results	Calculated Risk Score	CISS Classification
1	Malignant	3.62	Malignant
2	Malignant	2.67	Malignant
3	Malignant	2.74	Malignant
4	Benign	1.16	Benign
5	Benign	1.86	Benign
6	Benign	0.84	Benign
7	Benign	1.65	Benign
8	Malignant	3.00	Malignant
9	Malignant	2.46	Malignant
10	Malignant	3.27	Malignant
11	Benign	1.61	Benign
12	Malignant	2.40	Malignant
13	Benign	1.25	Benign
14	Benign	1.59	Benign
15	Malignant	2.13	Malignant
16	Malignant	2.15	Malignant
17	Benign	0.65	Benign
18	Malignant	3.05	Malignant
19	Benign	1.24	Benign
20	Benign	4.42	Malignant++
21	Benign	2.78	Malignant++

determined using (6), (7) and (8). For our dataset, *Sensititvity$_{CISS}$*, *Specificity$_{CISS}$* and *Accuracy$_{CISS}$* are calculated to be 100%, 82% and 90.5%, respectively. We note that the optimal weights for all of the subsets came out to be $W_1 =0.3, W_2=0.7$. The optimal threshold came out to be 1.99. Those optimal weights may not be the global optimal values. However, weighing stiffness more than the size seems to agree with literature and the experiences of doctors.

IV. CONCLUSION

Non-invasive Compression Induced Sensing System (CISS) is developed and used to classify the tumor as benign or malignant based on the proposed Risk Score. The experiments were done in both *phantom* and *in vivo* settings. The tumor phantoms, which mimic the conditions of real breast tumor, were used for *the preliminary* studies and calculated the properties of tumor such as size and the deformation index of the probe tip. It is found from the experiment studies, that the CISS can measure the tumor sizes from the range of 9 mm to 20 mm with an error range of 5.52% to 70% in a depth condition of superficial epidermal (0 mm) to a depth of 15 mm. As the second stage, this device is used in clinical setting to characterize the tumors from human patients. For our risk score, we found the optimal weights for the size to be 0.3 and stiffness to be

PRESENTING RESULTS

- However, each of the above needs text to describe and/or provide context.
- If you don’t reference it, don’t include it.

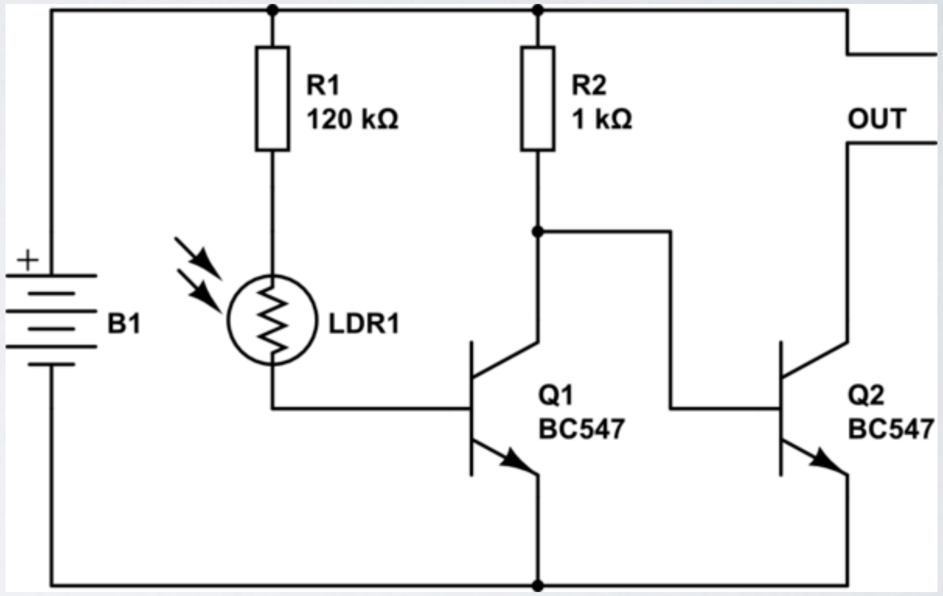
IMAGES

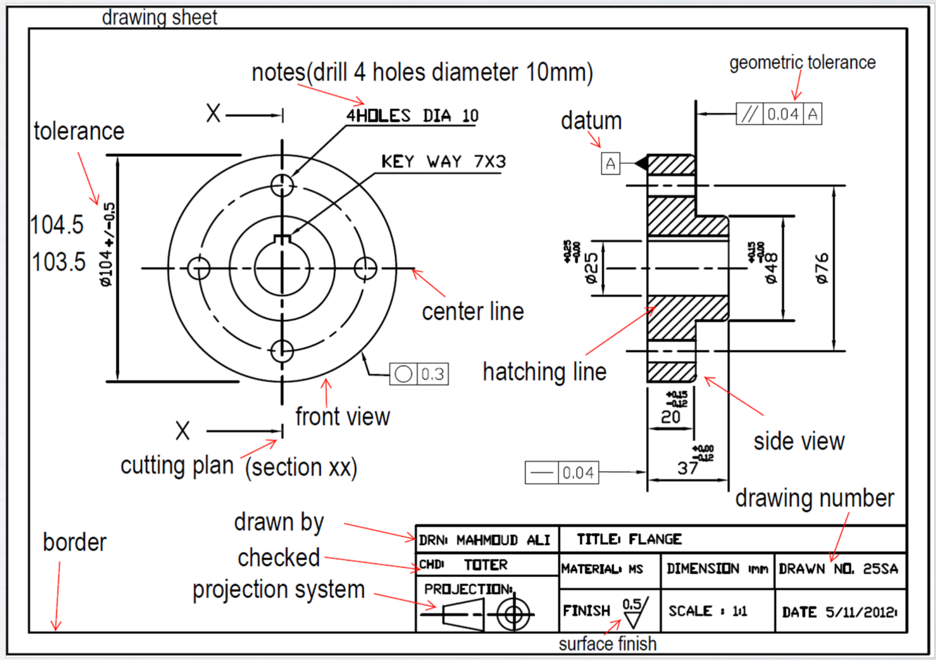
- Make Sure Image Is Clear And Will Show Well In Black & White
- Assume The Reader Knows Nothing About What You Are Showing. Use Labels/Arrows To Point At Key Components, Features, Etc.
- Make Sure Image Fonts Are Of Legible Size.
- Provide A Descriptive Caption Referencing Any Thing That You Have Highlighted
- Try To Limit The Size In Manuscript To Something Reasonable.
- Recognize These Probably Have The Least Value In Terms Of Understanding
- What You Did. Really Just Shows That You Actually Built Something.
- Probably Does Not Need A Lot Of Descriptive Text But Should Be Reference In The Text.
- Provide A Figure Number And Reference By "Fig. X".

Samsung S5K3P3SX
16 MP ISOCELL CIS

DRAWINGS

- Very Useful In Showingwhat You Did
- Make Sure All Aspects Are Well-Labeled.
- Make Sure Drawing Fonts Are Of Legible Size.
- Informed Reader Should Be Able To 'Build' What You Are Presenting.
- If Schematic, Provide Details On Components.
- Provide A Descriptive Caption
- Probably Can Give More Room In Your Manuscript To Drawing Vs. Image.
- Discuss In Detail In Text (One Paragraph).
- Provide A Figure Number And Reference By "Fig. X".





EQUATIONS

- Use Simplest Nomenclature Possible
- Use Standard Variables (E.G., ’T’ For Time; ’X,Y,Z’ For 3D Space)
- Make Sure All Variables Are Defined At First Appearance.
- Refresh A Readers Memory On What A Particular Term Means.
- Break Up The Presentation Of Equations With Text Discussing: What? Why? How?
- Avoid Combining Superscripts And Subscripts.
- Make Sure Equations You Reference Latter (I.E., The Important Ones)
- Have Equations Numbers. Reference As “Eq. (X)”.

$$
\begin{aligned}
\sigma^2 &= \frac{\sum(X-\mu)^2}{N} && (1)\\
&= \frac{\sum(X^2 - 2\mu X + \mu^2)}{N} && (2)\\
&= \frac{\sum X^2}{N} - \frac{2\mu \sum X}{N} + \frac{N\mu^2}{N} && (3)\\
&= \frac{\sum X^2}{N} - 2\mu^2 + \mu^2 && (4)\\
&= \frac{\sum X^2}{N} - \mu^2 && (5)
\end{aligned}
$$

TABLES

- Great Way To Summarize Cases And Results
- Be Sure You Have Clear Labels For Rows And Columns
- Limit The Size. If You Have So Much Data, Then A Plot Is More Appropriate
- Use Divider Lines (And Double Lines) As Needed
- Highlight In Italics/Bold Any Particular Result That You Will Discuss In Detail.
- Title Should Be Short (One Line) But Descriptive
- Discuss With A Dedicated Paragraph.
- Provide A Table Number And Reference By "Table X".

Clinical Values	**Intervention (n=200)**	**Control (n=201)**
No. (%) with data	175 (88)	180 (90)
Weight, mean (SD), kg	70 (12)	68 (12)
Cholesterol, mean (SD), mg/dL	212 (10)	214 (13)
Blood pressure, mean (SD), mm Hg		
Systolic	118 (20)	117 (19)
Diastolic	70 (13)	69 (20)

PLOTS

- Arguably The Most Useful Tool You Have To Show Your Results
- Allows You To Compare Performance (Y-Axis) Over Multiple Cases (X-3Axis) And Methods (Different Lines)
- Unless You Have A Good Reason, Present Data Consistent To Prior Work (Standards). E.G., Voltage Vs. Time And Not Time Vs. Voltage For Oscilloscope Results.
- Think About Whether Scatter Plots, Fitted Curves, Bar Charts Are The Best Way To Represent Your Results. What Have Other People Done?
- Individual Data Points Should Be Shown As Long As The Number Is Small Enough (< 20 Points ???)

PLOTS

- Make Sure Axis Numbers, Fonts Are Of Legible Size.
- Try To Limit Plots To 3 Or Less Curves.
- 3D Plots Are Hard To Interpret. Try To Represent In 2D.
- Avoid Double-Y Plots If Possible.
- Use Different Line Styles As Appropriate. Solid For Known Or Theory. Dashed For Data. Dots For Thresholds.
- Make Sure Plot Looks Good In Black & White
- Legends Are Ok, But Labeled Arrows On Curves Are Better
- If There's Some Particular Point Of Interest On The Curve, Point It Out.

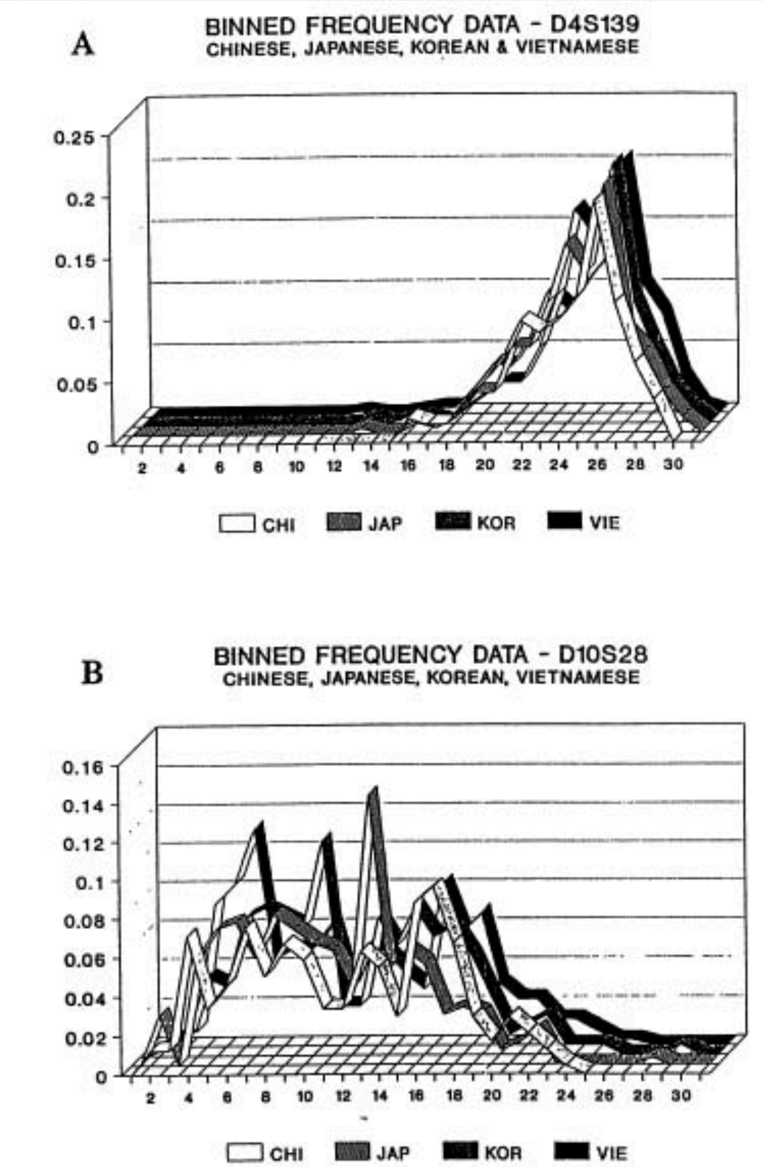
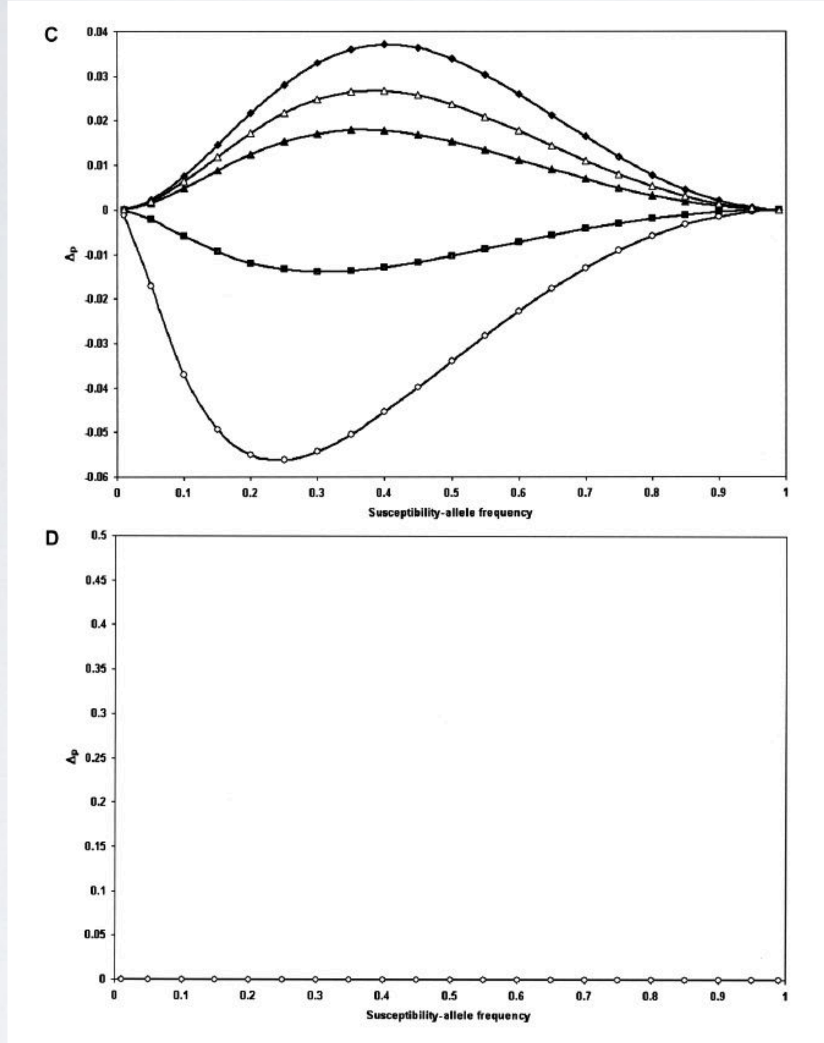


FIG. 4. Fixed bin distribution (histogram) for two loci and four Asian subpopulations (used with permission from John Hartmann): the boundaries of the 30 bins (vertical axis) are determined by the FBI; these bins are not of equal length. Sample sizes (numbers of individuals) for Chinese, Japanese, Korean and Vietnamese are 103, 125, 93 and 215 for D4S139 and 120, 137, 100 and 193 for D10S28. The horizontal axis is the bin number; bins are not of equal length.



Don't believe me?
Here's the graph from a paper on the subject.

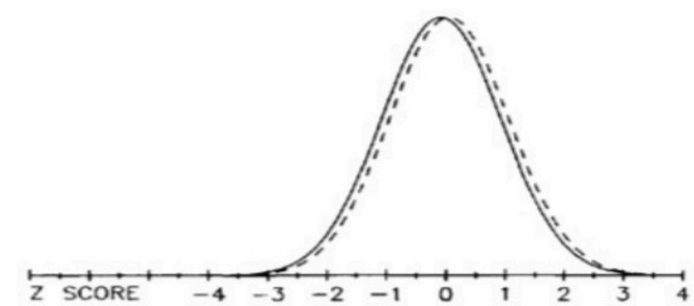
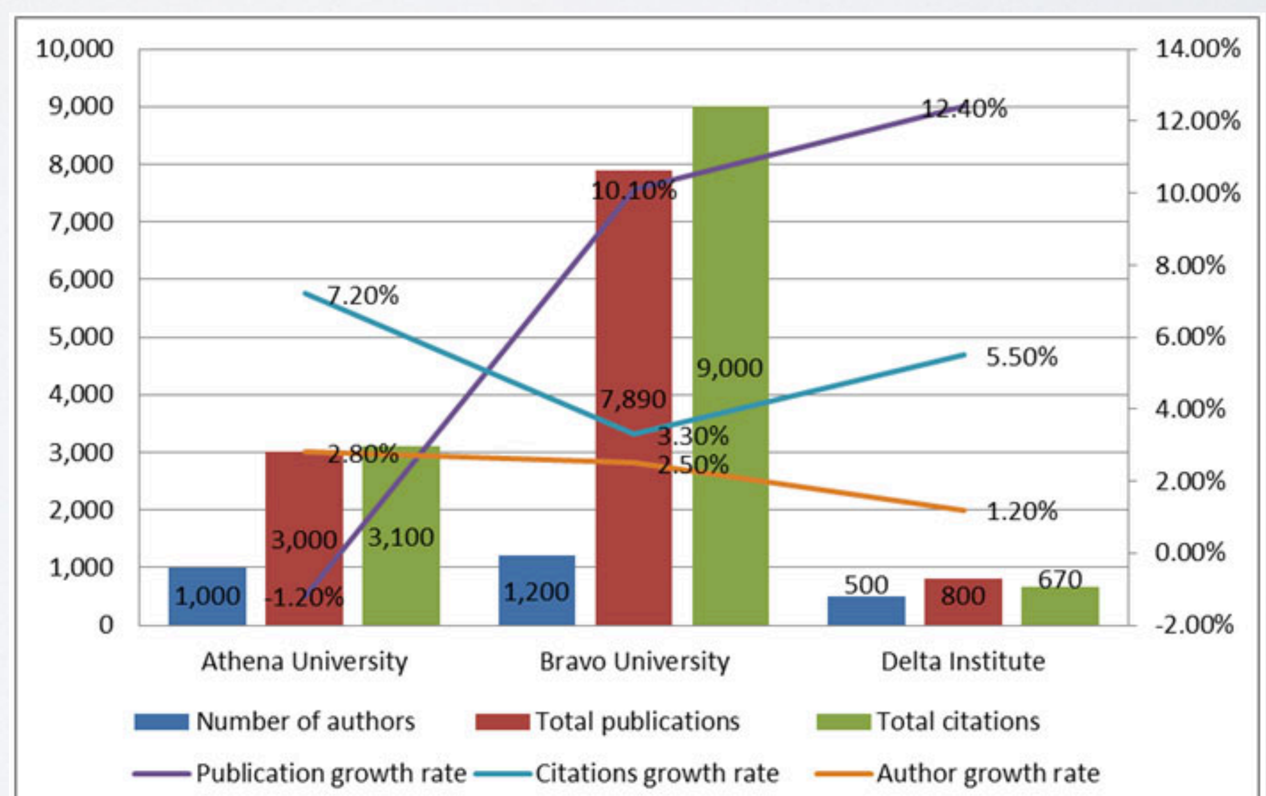


Figure 1. Two normal distributions that are 0.15 standard deviations apart (i.e., $d = 0.15$. This is the approximate magnitude of the gender difference in mathematics performance, averaging over all samples.)

Hyde, J; Fennema, E; Lamon, S. “Gender differences in mathematics performance: A meta-analysis.” Psychological Bulletin. Vol 107(2), Mar 1990, 139-155.



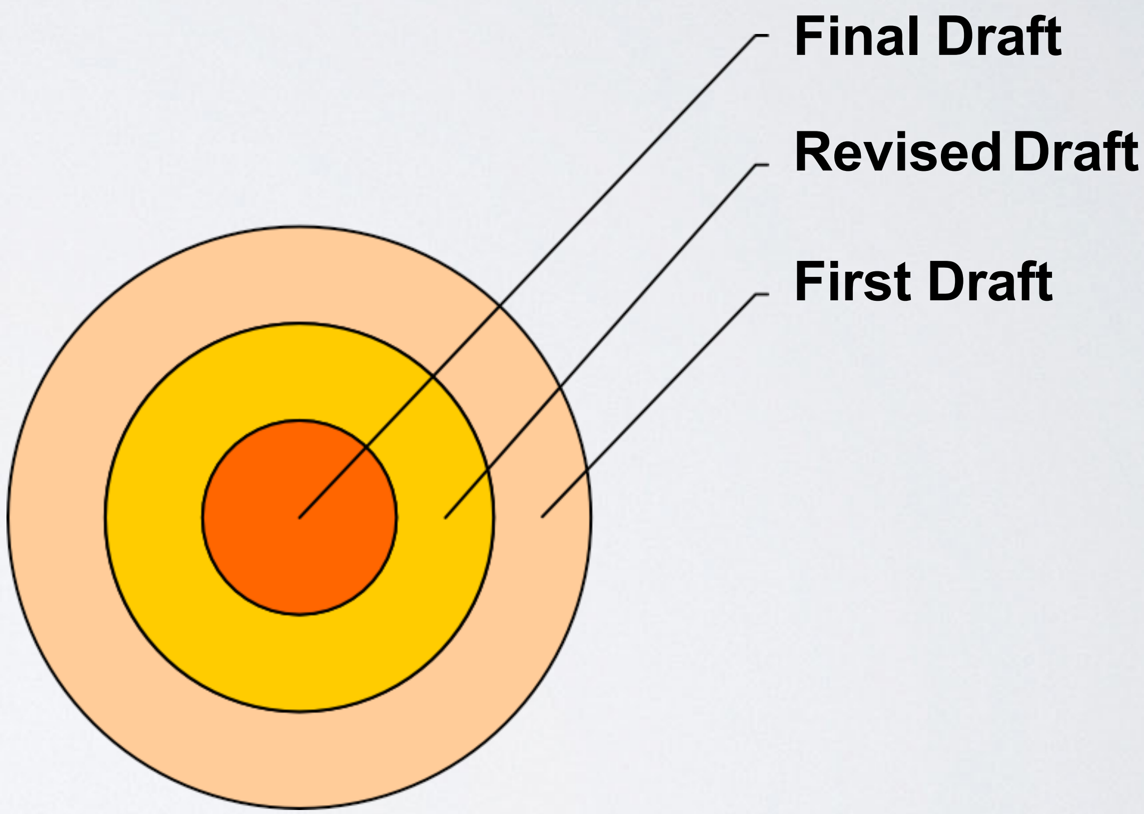
PLOTS

- Use Vertical Dotted Lines To Break Plots Into Zones If Appropriate
- Use Horizontal Dotted Lines To Show Thresholds
- Provide A Descriptive Caption That Distinguishes This Plot From The Rest.
- Discuss Result In Great Detail. You Are Using The Plot To Summarize A Lot Of Data. Each Plot Should Be Thought Of As A Key Result From Your Paper. Detail In Text (At Least One Paragraph).
- Present Only The Data You Need To Make Your Key Points. If You Don't Have A Good Discussion For A Particular Curve, Then It May Only Serve As A Distraction.
- Provide A Figure Number And Reference By "Fig. X".

SUMMARY

WRITING IS A PROCESS

- Good writing doesn't happen overnight; it requires planning, drafting, rereading, revising, and editing.
- Learning and improvement requires self-review, peer-review, subject-matter expert feedback, and practice.
- There are no shortcuts; practice makes perfect!



TO SUMMARIZE

- Plan your project before you begin drafting.
- Understand basic qualities of good technical writing; use the examples presented to guide you in your writing and revising process.
- Good writing is a habit that takes time to develop; practice makes perfect.
- Present your results effectively.

?