COMPANION TO THE CURRICULUM

Sources and notes

المصادر والملاحظات

Where the citations live.

We write our chapters, guides, sessions and workshops as plain prose, without citations in brackets, because they're meant for students who have never read a paper. The citations are here instead. Every claim in our material that can be traced to a source is listed against that source, programme by programme and section by section.

1	Research cohort البحث	141 claims
2	Skills cohort المهارات	172 claims
3	Principles of Research مبادئ البحث	140 claims
4	Workshops الورش	133 claims
5	Free guides الأدلة المجانية	15 claims

EDITION Third edition · October 2026
IN THIS EDITION 601 claims · 233 sources
CORRECTIONS contact@researchbasira.com

Gomosani, A., & Marghalani, H. (2026). *Basira Curriculum*. Zenodo. https://doi.org/10.5281/zenodo.21791751

كل ادعاء يرجع لمصدره.

Every claim goes back to its source.

Each source has its DOI where one exists, and we checked every DOI against the record it points to: authors, year, title, journal and pages. Sources without a DOI, such as standards, guidelines, laws and software, were checked against their official page. Claims that are there to illustrate a point (a worked example, made-up data, an analogy, a teaching device) aren't listed, because there's no source to give for them.

KEY IDEA

Read this alongside “Go deeper”

The “Go deeper” list at the end of a chapter or guide is a reading list, picked for what a student should read next. This document is the evidence behind the text, arranged by claim. Where one source supports several sections, it's listed once with all of them.

PROGRAMME 1

Research cohort البحث

The seven chapters (0 to 6), the session plans and teaching notes, and the six assignments. 141 claims, 88 sources.

0 Arriving الوصول

SECTION	CLAIM IN THE TEXT	SOURCE
0.2, 0.8	Zotero is one of the three Stage 0 accounts (reference manager).	Corporation for Digital Scholarship. Zotero. zotero.org
0.2, 0.8	OSF (Open Science Framework) is one of the three Stage 0 accounts.	Foster, E. D., & Deardorff, A. (2017). Open Science Framework (OSF). Journal of the Medical Library Association 105(2), 203–206. doi.org/10.5195/jmla.2017.88
0.2, 0.8	GitHub is one of the three Stage 0 accounts (version control and code sharing).	GitHub, Inc. About GitHub and Git. GitHub Docs. docs.github.com
0.5	Many tumours burn glucose inefficiently through glycolysis even when oxygen is available (the Warburg effect).	Warburg, O. (1956). On the Origin of Cancer Cells. Science 123(3191), 309–314. doi.org/10.1126/science.123.3191.309
0.5	A single metabolite made by a mutated enzyme (an oncometabolite) can help drive cancer.	Dang, L., et al. (2009). Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. Nature 462(7274), 739–744. doi.org/10.1038/nature08617 The founding example: mutant IDH1 produces 2-hydroxyglutarate. IDH1 is also the gene suggested for the 0.7 data tour.
0.5	Tumours compete with surrounding tissue and immune cells for nutrients.	Chang, C.-H., et al. (2015). Metabolic Competition in the Tumor Microenvironment Is a Driver of Cancer Progression. Cell 162(6), 1229–1241. doi.org/10.1016/j.cell.2015.08.016
0.5, 0.7, 0.10, 0.11	DepMap profiles hundreds of cancer cell lines with gene-dependency screens (genes a line cannot survive without) and molecular profiles.	Tsherniak, A., et al. (2017). Defining a Cancer Dependency Map. Cell 170(3), 564–576.e16. doi.org/10.1016/j.cell.2017.06.010 The molecular profiles come from the Cancer Cell Line Encyclopedia: Ghandi, M., et al. (2019), Nature 569(7757), 503–508, doi.org/10.1038/s41586-019-1186-3.
0.7	Gene expression is not metabolic activity: a highly expressed gene does not prove its pathway runs faster, because protein and enzyme activity are unmeasured steps.	Liu, Y., Beyer, A., & Aebersold, R. (2016). On the Dependency of Cellular Protein Levels on mRNA Abundance. Cell 165(3), 535–550. doi.org/10.1016/j.cell.2016.03.014 This covers the mRNA-to-protein step; enzyme activity and pathway flux are further steps the text also names.
0.7, 0.10, 0.11	cBioPortal is a point-and-click browser over many cancer genomics studies, including TCGA, needing no code.	Cerami, E., et al. (2012). The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. Cancer Discovery 2(5), 401–404. doi.org/10.1158/2159-8290.CD-12-0095 cBioPortal asks users to cite this paper together with Gao, J., et al. (2013), Science Signaling 6(269), pl1, doi.org/10.1126/scisignal.2004088.

SECTION	CLAIM IN THE TEXT	SOURCE
0.7, 0.11	TCGA holds mutations, expression, copy number and clinical data (stage, survival) on thousands of patient tumours across many cancer types.	The Cancer Genome Atlas Research Network, Weinstein, J. N., et al. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nature Genetics 45(10), 1113–1120. doi.org/10.1038/ng.2764
0.7, 0.10	TCGA data are served through the NCI Genomic Data Commons (GDC) portal.	Grossman, R. L., et al. (2016). Toward a Shared Vision for Cancer Genomic Data. New England Journal of Medicine 375(12), 1109–1112. doi.org/10.1056/NEJMp1607591
0.7, 0.10, 0.11	GEO is a large public archive of individual published expression studies.	Edgar, R., Domrachev, M., & Lash, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Research 30(1), 207–210. doi.org/10.1093/nar/30.1.207
0.10	Go deeper: World Clock Meeting Planner, to see a group's time-zone overlap.	timeanddate.com. World Clock Meeting Planner.

1 Thinking التفكير

SECTION	CLAIM IN THE TEXT	SOURCE
1.1, 1.10, 1.12	A hypothesis earns credibility by surviving serious attempts to refute it; one that cannot fail is not scientific (falsifiability).	Popper, K. (2002). The Logic of Scientific Discovery. Routledge (first English edition 1959). doi.org/10.4324/9780203994627 Popper is the origin of falsifiability as the mark of a scientific claim; the chapter does not name him.
1.3, 1.9, 1.13	Strong inference: list the rival explanations for a result, then design the analysis that distinguishes among them.	Platt, J. R. (1964). Strong Inference. Science 146(3642), 347–353. doi.org/10.1126/science.146.3642.347
1.4	Absence of evidence is not evidence of absence: a null result may mean no effect or a test unable to see one.	Altman, D. G., & Bland, J. M. (1995). Absence of evidence is not evidence of absence. BMJ 311(7003), 485. doi.org/10.1136/bmj.311.7003.485
1.4, 1.13	A false positive spreads: others build on it, so in a discovery science it does more damage than a false negative.	Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. PLoS Medicine 2(8), e124. doi.org/10.1371/journal.pmed.0020124
1.4, 1.11, 1.13	The first duty is not to fool yourself; report everything that might make your result invalid (leaning over backwards).	Feynman, R. P. (1974). Cargo Cult Science. Caltech commencement address, published in Engineering and Science 37(7). Not a DOI source; checked against the Caltech archive copy and the Engineering and Science 37(7), June 1974 issue record.
1.5.2, 1.12	Reproducibility (same data and code give the same numbers) is distinct from replication (a new study reaches the same conclusion).	National Academies of Sciences, Engineering, and Medicine (2019). Reproducibility and Replicability in Science. National Academies Press. doi.org/10.17226/25303 The report uses 'replicability'; usage of the two words differs across fields, which is why it fixes definitions.
1.7	Confirmation bias: unknowingly giving more weight to evidence for a favoured conclusion than against it.	Nickerson, R. S. (1998). Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. Review of General Psychology 2(2), 175–220. doi.org/10.1037/1089-2680.2.2.175

SECTION	CLAIM IN THE TEXT	SOURCE
1.7	HARKing (Hypothesising After the Results are Known) presents a story fitted to the data as a tested prediction, inflating apparent evidence.	Kerr, N. L. (1998). HARKing: Hypothesizing After the Results are Known. Personality and Social Psychology Review 2(3), 196–217. doi.org/10.1207/s15327957pspr0203_4
1.7, 1.13	Test enough things, with enough analytic freedom, and striking or significant patterns appear by chance for almost any hypothesis.	Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. Psychological Science 22(11), 1359–1366. doi.org/10.1177/0956797611417632
1.8	Label results exploratory or confirmatory; a pattern found after seeing the data is a hypothesis for the next study, not a finding.	Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. A. (2012). An Agenda for Purely Confirmatory Research. Perspectives on Psychological Science 7(6), 632–638. doi.org/10.1177/1745691612463078
1.9	Abduction, or inference to the best explanation: reasoning from an observation to the hypothesis that would best account for it.	Harman, G. H. (1965). The Inference to the Best Explanation. The Philosophical Review 74(1), 88–95. doi.org/10.2307/2183532 Harman coined 'inference to the best explanation'; the term 'abduction' is C. S. Peirce's.
1.11	Pre-specification: fix thresholds, comparisons and what counts as a hit before seeing results.	Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. PNAS 115(11), 2600–2606. doi.org/10.1073/pnas.1708274114
1.11	Provenance: every number in a final figure or table traces to the data and code that produced it; nothing typed by hand.	Sandve, G. K., Nekrutenko, A., Taylor, J., & Hovig, E. (2013). Ten Simple Rules for Reproducible Computational Research. PLoS Computational Biology 9(10), e1003285. doi.org/10.1371/journal.pcbi.1003285
1.13	Go deeper: Crash Course Scientific Thinking, a seven-episode series (episodes 2, 4, 5, 6, 7 named).	Crash Course (2026). Scientific Thinking [video series, 7 episodes]. YouTube playlist. Seven episodes confirmed on thecrashcourse.com; all five linked episode titles match.
1.13	Go deeper: Khan Academy Statistics & Probability.	Khan Academy. Statistics and Probability. khanacademy.org
1.13	Go deeper: Calling Bullshit, a free university course by Bergstrom & West, University of Washington.	Bergstrom, C. T., & West, J. D. Calling Bullshit: Data Reasoning in a Digital World. University of Washington course, callingbullshit.org The course site also lists the book of the same name by the same authors.

2 Practice الممارسة

SECTION	CLAIM IN THE TEXT	SOURCE
2.0	HARKing named among the practices students should recognise in their own work.	Kerr, N. L. (1998). HARKing: Hypothesizing After the Results are Known. Personality and Social Psychology Review 2(3), 196–217. doi.org/10.1207/s15327957pspr0203_4

SECTION	CLAIM IN THE TEXT	SOURCE
2.0, 2.3, 2.13, 2.14	Questionable research practices are common and admitted often, doing more damage to the literature than outright fraud.	John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling. Psychological Science 23(5), 524–532. doi.org/10.1177/0956797611430953 Anonymous survey of over 2,000 psychologists, with incentives for honest answers.
2.0, 2.3	Fabrication is rare, but its quieter relatives are not.	Fanelli, D. (2009). How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data. PLoS ONE 4(5), e5738. doi.org/10.1371/journal.pone.0005738 Pooled 1.97% admitted fabricating or falsifying at least once; up to 33.7% admitted other questionable practices.
2.0, 2.3	p-hacking named as a practice that manufactures findings from noise.	Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file-drawer. Journal of Experimental Psychology: General 143(2), 534–547. doi.org/10.1037/a0033242 Paper that introduced and defined the term p-hacking.
2.0, 2.14	The person you are most likely to deceive is yourself.	Feynman, R. P. (1974). Cargo Cult Science. Caltech commencement address, published in Engineering and Science 37(7).
2.2, 2.6	Fabrication, falsification and plagiarism are the three categorical violations; honest error is not misconduct.	US Department of Health and Human Services. 42 CFR Part 93: Public Health Service Policies on Research Misconduct (§93.234 Research misconduct; §§93.211, 93.212, 93.227; §93.103). ecfr.gov Revised 2024; the definition is now §93.234. A finding requires conduct 'intentionally, knowingly, or recklessly', so recklessness counts as well as deliberate intent.
2.3.1	Selecting candidates by a criterion and then validating them on the same criterion guarantees the result (circular analysis).	Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F., & Baker, C. I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. Nature Neuroscience 12(5), 535–540. doi.org/10.1038/nn.2303
2.3, 2.14	Researcher degrees of freedom: enough flexibility lets a diligent scientist find significance for almost any hypothesis, even an impossible one.	Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. Psychological Science 22(11), 1359–1366. doi.org/10.1177/0956797611417632 The term is coined in this paper; the 'impossible' result is that a song made listeners younger.
2.4	Motivated reasoning works below awareness, so willpower is an unreliable defence when you want a result.	Kunda, Z. (1990). The case for motivated reasoning. Psychological Bulletin 108(3), 480–498. doi.org/10.1037/0033-2909.108.3.480
2.4, 2.9	Version control, code that produces every number from source, and unit tests make shared analyses trustworthy and safe to change.	Wilson, G., et al. (2014). Best Practices for Scientific Computing. PLoS Biology 12(1), e1001745. doi.org/10.1371/journal.pbio.1001745
2.5	Gift (honorary) and ghost authorship are common enough that students should expect to meet them.	Wislar, J. S., Flanagin, A., Fontanarosa, P. B., & DeAngelis, C. D. (2011). Honorary and ghost authorship in high impact biomedical journals: a cross sectional survey. BMJ 343, d6128. doi.org/10.1136/bmj.d6128 21% of 2008 articles in six leading journals had honorary or ghost authors, or both.

SECTION	CLAIM IN THE TEXT	SOURCE
2.5, 2.14	Authorship requires substantial contribution together with accountability; adding non-contributors or omitting contributors misrepresents credit.	International Committee of Medical Journal Editors (ICMJE). Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals: Defining the Role of Authors and Contributors. icmje.org
2.6, 2.14	Correct the record promptly after an honest error; correction is distinct from misconduct.	Barbour, V., Kleinert, S., Wager, E., & Yentis, S. (2009). Guidelines for retracting articles. Committee on Publication Ethics. doi.org/10.24318/cope.2019.1.4 COPE's retraction guidelines; COPE's flowcharts and core practices are at publicationethics.org.
2.8	Disagree and commit: argue your case, then fully support the decision once it is made.	Amazon. Leadership Principles: Have Backbone; Disagree and Commit. amazon.jobs Popularised by Amazon; see also J. Bezos, 2016 Letter to Shareholders.
2.9	Provenance from source to figure lets anyone check a claim without asking the author.	Sandve, G. K., Nekrutenko, A., Taylor, J., & Hovig, E. (2013). Ten Simple Rules for Reproducible Computational Research. PLoS Computational Biology 9(10), e1003285. doi.org/10.1371/journal.pcbi.1003285
2.11	Show the data, not only a summary that hides spread; bar charts of means can conceal very different distributions.	Weissgerber, T. L., Milic, N. M., Winham, S. J., & Garovic, V. D. (2015). Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm. PLOS Biology 13(4), e1002128. doi.org/10.1371/journal.pbio.1002128
2.11	Truncated axes make a modest difference look dramatic and mislead readers.	Pandey, A. V., Rall, K., Satterthwaite, M. L., Nov, O., & Bertini, E. (2015). How Deceptive are Deceptive Visualizations? An Empirical Analysis of Common Distortion Techniques. Proceedings of CHI 2015, 1469–1478. doi.org/10.1145/2702123.2702608
2.12	Scientific writing follows a standard logic: question and why, how it was answered, what was found, what it means (IMRaD).	Sollaci, L. B., & Pereira, M. G. (2004). The introduction, methods, results, and discussion (IMRAD) structure: a fifty-year survey. Journal of the Medical Library Association 92(3), 364–367. No DOI; verified on PubMed (PMID 15243643).
2.14	Go deeper: Crash Course Scientific Thinking #4, Peer Review and the Quest for Truth, and the full series.	Crash Course (2026). Peer Review and the Quest for Truth: Crash Course Scientific Thinking #4. YouTube.
2.14	Go deeper: Retraction Watch, a running record of withdrawn papers and why.	Retraction Watch. Tracking retractions as a window into the scientific process. retractionwatch.com

3 The Science العلم

SECTION	CLAIM IN THE TEXT	SOURCE
3.4	Both copies of a tumour suppressor usually have to be disabled before the brake fails: the two-hit requirement.	Knudson, A. G. (1971). Mutation and Cancer: Statistical Study of Retinoblastoma. PNAS 68(4), 820–823. doi.org/10.1073/pnas.68.4.820

SECTION	CLAIM IN THE TEXT	SOURCE
3.4, 3.5	Drivers confer a growth advantage; passengers are carried along. The two classes are altered by different kinds of mutation.	Vogelstein, B., et al. (2013). Cancer Genome Landscapes. Science 339(6127), 1546–1558. doi.org/10.1126/science.1235122
3.5	The first principle is that you must not fool yourself.	Feynman, R. P. (1974). Cargo Cult Science. Caltech commencement address, published in Engineering and Science 37(7), 10–13. calteches.library.caltech.edu/51/2/CargoCult.htm
3.5	Cancer as evolution by natural selection inside a tissue, ending in a mixed rather than uniform population.	Nowell, P. C. (1976). The Clonal Evolution of Tumor Cell Populations. Science 194(4260), 23–28. doi.org/10.1126/science.959840
3.7	The eight capabilities a cell must acquire, and the two enabling characteristics.	Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of Cancer: The Next Generation. Cell 144(5), 646–674. doi.org/10.1016/j.cell.2011.02.013 · Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. Cancer Discovery 12(1), 31–46. doi.org/10.1158/2159-8290.CD-21-1059
3.7, 3.17	The hallmarks of cancer.	Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. Cancer Discovery 12(1), 31-46. doi.org/10.1158/2159-8290.cd-21-1059
3.12	Thirteen mitochondrial proteins are encoded in mitochondrial DNA and roughly 1,100 in the nucleus.	Rath, S., et al. (2021). MitoCarta3.0. Nucleic Acids Research 49(D1), D1541–D1547. doi.org/10.1093/nar/gkaa1011. MitoCarta3.0 lists 1,136 human genes, which is the source of the figure quoted.
3.13	Metabolism for building rather than only for energy; the metabolic hallmarks.	Pavlova, N. N., & Thompson, C. B. (2016). The Emerging Hallmarks of Cancer Metabolism. Cell Metabolism 23(1), 27–47. doi.org/10.1016/j.cmet.2015.12.006
3.14	Warburg's original observation in the 1920s: aerobic glycolysis with lactate secretion despite available oxygen.	Warburg, O. (1956). On the Origin of Cancer Cells. Science 123(3191), 309–314. doi.org/10.1126/science.123.3191.309
3.14	The correction: the Warburg effect does not mean mitochondria are switched off, and mitochondrial metabolism remains essential in most tumours.	Vander Heiden, M. G., Cantley, L. C., & Thompson, C. B. (2009). Understanding the Warburg Effect. Science 324(5930), 1029–1033. doi.org/10.1126/science.1160809
3.15	Strong inference: list the rival explanations and design the test that separates them.	Platt, J. R. (1964). Strong Inference. Science 146(3642), 347-353. doi.org/10.1126/science.146.3642.347

4 Data & Tools البيانات والأدوات

SECTION	CLAIM IN THE TEXT	SOURCE
4.2	The Cancer Genome Atlas as a source of matched tumour, expression and clinical data.	The Cancer Genome Atlas Research Network, Weinstein, J. N., et al. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nature Genetics 45(10), 1113–1120. doi.org/10.1038/ng.2764
4.2	Gene Expression Omnibus as the repository behind most reusable expression datasets.	Edgar, R., Domrachev, M., & Lash, A. E. (2002). Gene Expression Omnibus. Nucleic Acids Research 30(1), 207–210. doi.org/10.1093/nar/30.1.207

SECTION	CLAIM IN THE TEXT	SOURCE
4.2	cBioPortal as the browsing layer over these cohorts.	Cerami, E., et al. (2012). The cBio Cancer Genomics Portal. Cancer Discovery 2(5), 401–404. doi.org/10.1158/2159-8290.CD-12-0095
4.2	DepMap and the dependency-screen data behind it.	Tsherniak, A., et al. (2017). Defining a Cancer Dependency Map. Cell 170(3), 564–576. doi.org/10.1016/j.cell.2017.06.010
4.5.1	Count models require raw counts; a log-transformed matrix run through one produces nonsense without an error.	Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15(12), 550. doi.org/10.1186/s13059-014-0550-8
4.6	Posit Cloud, where we run R in the browser.	Posit Software. Posit Cloud. posit.cloud
4.6, 4.12	The tidyverse packages (dplyr and friends) used for data handling in R.	Wickham, H., et al. (2019). Welcome to the Tidyverse. Journal of Open Source Software 4(43), 1686. doi.org/10.21105/joss.01686
4.7	Reproducible practice: version control, environment capture, and producing every number from source.	The Turing Way Community. The Turing Way: A Handbook for Reproducible, Ethical and Collaborative Research. book.the-turing-way.org
4.9	Confounding: a third factor that can explain an apparent link.	Greenland, S., Pearl, J., & Robins, J. M. (1999). Causal Diagrams for Epidemiologic Research. Epidemiology 10(1), 37-48. doi.org/10.1097/00001648-199901000-00008
4.9, 4.10, 4.11	Batch effects can masquerade as biology.	Leek, J. T., et al. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. Nature Reviews Genetics 11(10), 733-739. doi.org/10.1038/nrg2825
4.12	MitoCarta3.0, the inventory of mitochondrial genes.	Rath, S., et al. (2021). MitoCarta3.0: an updated mitochondrial proteome now with sub-organelle localization and pathway annotations. Nucleic Acids Research 49(D1), D1541-D1547. doi.org/10.1093/nar/gkaa1011

5 The Toolkit عدّة العمل

SECTION	CLAIM IN THE TEXT	SOURCE
5.1, 5.10	The t-test for comparing the means of two groups.	Student (1908). The Probable Error of a Mean. Biometrika 6(1), 1. doi.org/10.2307/2331554
5.1, 5.10	The Mann–Whitney U test for comparing two independent groups by rank.	Mann, H. B., & Whitney, D. R. (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. The Annals of Mathematical Statistics 18(1), 50-60. doi.org/10.1214/aoms/1177730491
5.1, 5.14	The Kaplan–Meier estimate of a survival curve, which handles censored follow-up.	Kaplan, E. L., & Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. Journal of the American Statistical Association 53(282), 457-481. doi.org/10.1080/01621459.1958.10501452
5.1, 5.14	The log-rank test for comparing survival curves.	Peto, R., & Peto, J. (1972). Asymptotically Efficient Rank Invariant Test Procedures. Journal of the Royal Statistical Society. Series A (General) 135(2), 185. doi.org/10.2307/2344317

SECTION	CLAIM IN THE TEXT	SOURCE
5.1, 5.6	Effect size (Cohen's d) as a measure of how big a difference is.	Cohen, J. (1992). A power primer. Psychological Bulletin 112(1), 155-159. doi.org/10.1037/0033-2909.112.1.155
5.1, 5.11	DESeq2 for testing differences in gene expression from raw counts.	Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15(12), 550. doi.org/10.1186/s13059-014-0550-8
5.4, 5.11	The negative binomial as the count distribution for RNA-seq, allowing more variability than Poisson.	Love, Huber & Anders (2014), as above.
5.5	The three things a p-value is not.	Wasserstein, R. L., & Lazar, N. A. (2016). The ASA Statement on p-Values. The American Statistician 70(2), 129–133. doi.org/10.1080/00031305.2016.1154108
5.7	False discovery rate control, and why an uncorrected threshold is meaningless at genome scale.	Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate. Journal of the Royal Statistical Society: Series B 57(1), 289–300. doi.org/10.1111/j.2517-6161.1995.tb02031.x
5.8, 5.9, 5.12	Confounding: a third factor that can explain an apparent link.	Greenland, S., Pearl, J., & Robins, J. M. (1999). Causal Diagrams for Epidemiologic Research. Epidemiology 10(1), 37-48. doi.org/10.1097/00001648-199901000-00008
5.10	Analytical flexibility (p-hacking) can make almost any result look significant.	Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology. Psychological Science 22(11), 1359-1366. doi.org/10.1177/0956797611417632
5.12	Spearman's rank correlation.	Spearman, C. (1904). The Proof and Measurement of Association between Two Things. The American Journal of Psychology 15(1), 72. doi.org/10.2307/1412159
5.14	Cox proportional hazards regression and the hazard ratio.	Cox, D. R. (1972). Regression Models and Life-Tables. Journal of the Royal Statistical Society Series B: Statistical Methodology 34(2), 187-202. doi.org/10.1111/j.2517-6161.1972.tb00899.x
5.15	The tidyverse packages (dplyr and friends) used for data handling in R.	Wickham, H., et al. (2019). Welcome to the Tidyverse. Journal of Open Source Software 4(43), 1686. doi.org/10.21105/joss.01686
5.15	ggplot2, the plotting package used in the course.	Wickham, H. (2016). ggplot2. Springer International Publishing. doi.org/10.1007/978-3-319-24277-4

6 Build Your Own بناء سؤالك

SECTION	CLAIM IN THE TEXT	SOURCE
6.4	Controlling the false discovery rate when many tests are run.	Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B: Statistical Methodology 57(1), 289-300. doi.org/10.1111/j.2517-6161.1995.tb02031.x
6.4	Effect size (Cohen's d) as a measure of how big a difference is.	Cohen, J. (1992). A power primer. Psychological Bulletin 112(1), 155-159. doi.org/10.1037/0033-2909.112.1.155

SECTION	CLAIM IN THE TEXT	SOURCE
6.4	Pre-registration as a defence against analytical flexibility.	Center for Open Science, Preregistration. cos.io/initiatives/prereg · The underlying argument is Simmons, Nelson & Simonsohn (2011), as above.
6.4, 6.9	Pre-registration as a defence against analytical flexibility.	Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. Proceedings of the National Academy of Sciences 115(11), 2600-2606. doi.org/10.1073/pnas.1708274114
6.7	The garden of forking paths: many possible analyses inflate false positives.	Gelman, A., & Loken, E. (2014). The Statistical Crisis in Science. American Scientist 102(6), 460. doi.org/10.1511/2014.111.460
6.9	Strong inference: list the rival explanations and design the test that separates them.	Platt, J. R. (1964). Strong Inference. Science 146(3642), 347-353. doi.org/10.1126/science.146.3642.347
6.9	PubMed, the free search engine for the biomedical literature.	National Library of Medicine. PubMed. pubmed.ncbi.nlm.nih.gov

Research sessions and assignments

23 CLAIMS

SECTION	CLAIM IN THE TEXT	SOURCE
Session 2	cBioPortal: open it and have a look round.	Cerami, E., et al. (2012). The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. Cancer Discovery 2(5), 401–404. doi.org/10.1158/2159-8290.CD-12-0095
Assignment 3	Look for claims everybody cites and nobody has tested; repeated citation can turn an untested claim into accepted fact.	Greenberg, S. A. (2009). How citation distortions create unfounded authority: analysis of a citation network. BMJ 339, b2680. doi.org/10.1136/bmj.b2680
Assignment 3	Followed back, a cited source frequently does not show what the citing paper says it shows.	Jergas, H., & Baethge, C. (2015). Quotation accuracy in medical journal articles: a systematic review and meta-analysis. PeerJ 3, e1364. doi.org/10.7717/peerj.1364 Total quotation error rate 25.4%, major errors 11.9%.
Session 3 (Cohort 2 session 1); Session 18	Crash Course Scientific Thinking, full series.	Crash Course (2026). Scientific Thinking [video series, 7 episodes]. YouTube playlist.
Session 4 (Cohort 2 session 2)	Feynman, Cargo Cult Science.	Feynman, R. P. (1974). Cargo Cult Science. Caltech commencement address, published in Engineering and Science 37(7).
Assignment 5	PECO worksheet: Population, Exposure, Comparator, Outcome.	Morgan, R. L., Whaley, P., Thayer, K. A., & Schünemann, H. J. (2018). Identifying the PECO: A framework for formulating good questions to explore the association of environmental and other exposures with health outcomes. Environment International 121, 1027–1031. doi.org/10.1016/j.envint.2018.07.015

SECTION	CLAIM IN THE TEXT	SOURCE
Assignment 5	FINER: is the question Feasible, Interesting, Novel, Ethical and Relevant?	Hulley, S. B., Cummings, S. R., Browner, W. S., Grady, D. G., & Newman, T. B. (2013). Designing Clinical Research (4th ed.). Wolters Kluwer / Lippincott Williams & Wilkins. ISBN 9781608318049 FINER comes from this textbook (first edition 1988); edition details checked against a library catalogue record.
Assignment 5	Fix the FDR threshold in advance.	Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B 57(1), 289–300. doi.org/10.1111/j.2517-6161.1995.tb02031.x
Session 5 (Cohort 2 session 3)	Retraction Watch: read a week of it.	Retraction Watch. Tracking retractions as a window into the scientific process. retractionwatch.com
Session 6 (Cohort 2 session 4)	Crash Course Scientific Thinking #4: Peer Review.	Crash Course (2026). Peer Review and the Quest for Truth: Crash Course Scientific Thinking #4. YouTube.
Session 7 (Cohort 2 session 5)	Crash Course Biology: DNA, transcription, translation.	Crash Course (2012). Biology [video series]. YouTube playlist.
Session 8 (Cohort 2 session 6)	NCI: What Is Cancer?	National Cancer Institute. What Is Cancer? cancer.gov
Session 9 (Cohort 2 session 7)	Human Protein Atlas: browse the Dictionary (histology and pathology).	Human Protein Atlas. Dictionary. proteinatlas.org
Session 10 (Cohort 2 session 8)	Vander Heiden et al.: Understanding the Warburg Effect.	Vander Heiden, M. G., Cantley, L. C., & Thompson, C. B. (2009). Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation. Science 324(5930), 1029–1033. doi.org/10.1126/science.1160809
Session 11 (Cohort 2 session 9)	Hanahan & Weinberg: Hallmarks of Cancer.	Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of Cancer: The Next Generation. Cell 144(5), 646–674. doi.org/10.1016/j.cell.2011.02.013
Session 12 (Cohort 2 session 10)	DepMap portal: search a gene you know.	Tsherniak, A., et al. (2017). Defining a Cancer Dependency Map. Cell 170(3), 564–576.e16. doi.org/10.1016/j.cell.2017.06.010
Session 13 (Cohort 2 session 11)	The Turing Way: reproducible research.	The Turing Way Community. The Turing Way: A handbook for reproducible, ethical and collaborative research. Zenodo. doi.org/10.5281/zenodo.3233853 Concept DOI (all versions), verified on DataCite; read online at book.the-turing-way.org.
Session 14 (Cohort 2 session 12)	Seeing Theory: interactive probability.	Kunin, D. Seeing Theory: A visual introduction to probability and statistics. Brown University.
Session 15 (Cohort 2 session 13)	StatQuest: p-values and FDR.	Starmer, J. StatQuest with Josh Starmer [YouTube channel]. The link goes to the channel, not to specific p-value or FDR videos.

SECTION	CLAIM IN THE TEXT	SOURCE
Session 16 (Cohort 2 session 14)	Calling Bullshit: the causality lectures.	Bergstrom, C. T., & West, J. D. Calling Bullshit: Data Reasoning in a Digital World. University of Washington course, callingbullshit.org
Session 17 (Cohort 2 session 15)	Nature Methods: Points of Significance.	Nature. Statistics for Biologists [collection, including the Nature Methods 'Points of Significance' column]. nature.com The linked page is titled 'Statistics for Biologists'; it collects the Points of Significance columns.
Session 18 (Cohort 2 session 16); Assignment 5	Center for Open Science: preregistration; a dated plan, fixed before any result, that cannot be revised.	Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. PNAS 115(11), 2600–2606. doi.org/10.1073/pnas.1708274114 Session 18 links COS's page, cos.io/initiatives/prereg (checked).
Session 19 (Cohort 2 session 17)	Platt: Strong Inference (read it again).	Platt, J. R. (1964). Strong Inference. Science 146(3642), 347–353. doi.org/10.1126/science.146.3642.347

Session teaching notes

25 CLAIMS

SECTION	CLAIM IN THE TEXT	SOURCE
Session 2	Zotero, the free reference manager.	Corporation for Digital Scholarship. Zotero. zotero.org
Session 2, Session 3	cBioPortal as a way to browse and download cancer genomics data.	Cerami, E., et al. (2012). The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. Cancer Discovery 2(5), 401-404. doi.org/10.1158/2159-8290.cd-12-0095
Session 2, Session 12	The Cancer Genome Atlas: matched tumour, molecular and clinical data.	The Cancer Genome Atlas Research Network, et al. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nature Genetics 45(10), 1113-1120. doi.org/10.1038/ng.2764
Session 3, Session 4, Session 13, Session 18	Confounding: a third factor that can explain an apparent link.	Greenland, S., Pearl, J., & Robins, J. M. (1999). Causal Diagrams for Epidemiologic Research. Epidemiology 10(1), 37-48. doi.org/10.1097/00001648-199901000-00008
Session 3, Session 11	Strong inference: list the rival explanations and design the test that separates them.	Platt, J. R. (1964). Strong Inference. Science 146(3642), 347-353. doi.org/10.1126/science.146.3642.347
Session 3, Session 13	Batch effects can masquerade as biology.	Leek, J. T., et al. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. Nature Reviews Genetics 11(10), 733-739. doi.org/10.1038/nrg2825

SECTION	CLAIM IN THE TEXT	SOURCE
Session 4	HARKing: presenting a hypothesis made after seeing the results as if it came first.	Kerr, N. L. (1998). HARKing: Hypothesizing After the Results are Known. Personality and Social Psychology Review 2(3), 196-217. doi.org/10.1207/s15327957pspr0203_4
Session 4	Confirmation bias: noticing what fits what we already believe.	Nickerson, R. S. (1998). Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. Review of General Psychology 2(2), 175-220. doi.org/10.1037/1089-2680.2.2.175
Session 5	Fabrication, falsification and plagiarism as research misconduct.	US Office of Research Integrity. Definition of Research Misconduct, 42 CFR Part 93. ori.hhs.gov/definition-research-misconduct
Session 5	A hypothesis earns credibility by surviving serious attempts to refute it.	Popper, K. (2002). The Logic of Scientific Discovery. Routledge (first English edition 1959). ISBN 9780415278447
Session 5, Session 14, Session 15, Session 16, Session 17	What a p-value is and is not.	Wasserstein, R. L., & Lazar, N. A. (2016). The ASA Statement on p - Values: Context, Process, and Purpose. The American Statistician 70(2), 129-133. doi.org/10.1080/00031305.2016.1154108
Session 6	Spin: reporting that makes results look better than they are.	Boutron, I. (2010). Reporting and Interpretation of Randomized Controlled Trials With Statistically Nonsignificant Results for Primary Outcomes. JAMA 303(20), 2058. doi.org/10.1001/jama.2010.651
Session 8	The hallmarks of cancer.	Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. Cancer Discovery 12(1), 31-46. doi.org/10.1158/2159-8290.cd-21-1059
Session 10	The Warburg effect: tumours take up glucose and make lactate even with oxygen present.	Vander Heiden, M. G., Cantley, L. C., & Thompson, C. B. (2009). Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation. Science 324(5930), 1029-1033. doi.org/10.1126/science.1160809
Session 12	PubMed, the free search engine for the biomedical literature.	National Library of Medicine. PubMed. pubmed.ncbi.nlm.nih.gov
Session 12, Session 17	DESeq2 for testing differences in gene expression from raw counts.	Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15(12), 550. doi.org/10.1186/s13059-014-0550-8
Session 13	The tidyverse packages (dplyr and friends) used for data handling in R.	Wickham, H., et al. (2019). Welcome to the Tidyverse. Journal of Open Source Software 4(43), 1686. doi.org/10.21105/joss.01686
Session 13	Posit Cloud, where we run R in the browser.	Posit Software. Posit Cloud. posit.cloud
Session 15, Session 16	Effect size (Cohen's d) as a measure of how big a difference is.	Cohen, J. (1992). A power primer. Psychological Bulletin 112(1), 155-159. doi.org/10.1037/0033-2909.112.1.155

SECTION	CLAIM IN THE TEXT	SOURCE
Session 15, Session 16, Session 17	Controlling the false discovery rate when many tests are run.	Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B: Statistical Methodology 57(1), 289-300. doi.org/10.1111/j.2517-6161.1995.tb02031.x
Session 16, Session 17, Session 18	Pre-registration as a defence against analytical flexibility.	Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. Proceedings of the National Academy of Sciences 115(11), 2600-2606. doi.org/10.1073/pnas.1708274114
Session 17	The t-test for comparing the means of two groups.	Student (1908). The Probable Error of a Mean. Biometrika 6(1), 1. doi.org/10.2307/2331554
Session 17	The Mann–Whitney U test for comparing two independent groups by rank.	Mann, H. B., & Whitney, D. R. (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. The Annals of Mathematical Statistics 18(1), 50-60. doi.org/10.1214/aoms/1177730491
Session 17	Publication bias: positive results are more likely to be published.	Dickersin, K. (1990). The Existence of Publication Bias and Risk Factors for Its Occurrence. JAMA: The Journal of the American Medical Association 263(10), 1385. doi.org/10.1001/jama.1990.03440100097014
Session 19	The garden of forking paths: many possible analyses inflate false positives.	Gelman, A., & Loken, E. (2014). The Statistical Crisis in Science. American Scientist 102(6), 460. doi.org/10.1511/2014.111.460

PROGRAMME 2

Skills cohort المهارات

The nine Skills chapters, the fourteen session pages, the teaching notes and the Skills assignments. 172 claims, 71 sources.

Chapter 0 · Arriving الوصول

SECTION	CLAIM IN THE TEXT	SOURCE
k0.0	What a p-value is and is not.	Wasserstein, R. L., & Lazar, N. A. (2016). The ASA Statement on p - Values: Context, Process, and Purpose. The American Statistician 70(2), 129-133. doi.org/10.1080/00031305.2016.1154108
k0.0	Confounding: a third factor that can explain an apparent link.	Greenland, S., Pearl, J., & Robins, J. M. (1999). Causal Diagrams for Epidemiologic Research. Epidemiology 10(1), 37-48. doi.org/10.1097/00001648-199901000-00008
k0.0, k0.1, k0.2, k0.3, k0.5, k0.7 and more	Posit Cloud, where we run R in the browser.	Posit Software. Posit Cloud. posit.cloud
k0.1	cBioPortal as a way to browse and download cancer genomics data.	Cerami, E., et al. (2012). The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. Cancer Discovery 2(5), 401-404. doi.org/10.1158/2159-8290.cd-12-0095
k0.1	DepMap: genome-wide dependency screens across cancer cell lines.	Tsherniak, A., et al. (2017). Defining a Cancer Dependency Map. Cell 170(3), 564-576.e16. doi.org/10.1016/j.cell.2017.06.010
k0.1, k0.3	The Cancer Genome Atlas: matched tumour, molecular and clinical data.	The Cancer Genome Atlas Research Network, et al. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nature Genetics 45(10), 1113-1120. doi.org/10.1038/ng.2764
k0.3, k0.5	The tidyverse packages (dplyr and friends) used for data handling in R.	Wickham, H., et al. (2019). Welcome to the Tidyverse. Journal of Open Source Software 4(43), 1686. doi.org/10.21105/joss.01686

Chapter 1 · Thinking with data التفكير بالبيانات

SECTION	CLAIM IN THE TEXT	SOURCE
k1.1	What a p-value is and is not.	Wasserstein, R. L., & Lazar, N. A. (2016). The ASA Statement on p - Values: Context, Process, and Purpose. The American Statistician 70(2), 129-133. doi.org/10.1080/00031305.2016.1154108

SECTION	CLAIM IN THE TEXT	SOURCE
k1.1, k1.4	The Cancer Genome Atlas: matched tumour, molecular and clinical data.	The Cancer Genome Atlas Research Network, et al. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nature Genetics 45(10), 1113-1120. doi.org/10.1038/ng.2764
k1.1, k1.4	DepMap: genome-wide dependency screens across cancer cell lines.	Tsherniak, A., et al. (2017). Defining a Cancer Dependency Map. Cell 170(3), 564-576.e16. doi.org/10.1016/j.cell.2017.06.010
k1.3	Posit Cloud, where we run R in the browser.	Posit Software. Posit Cloud. posit.cloud
k1.4	Tidy data: one row per observation, one column per variable.	Wickham, H. (2014). Tidy Data. Journal of Statistical Software 59(10). doi.org/10.18637/jss.v059.i10
k1.4	The tidyverse packages (dplyr and friends) used for data handling in R.	Wickham, H., et al. (2019). Welcome to the Tidyverse. Journal of Open Source Software 4(43), 1686. doi.org/10.21105/joss.01686
k1.4	ggplot2, the plotting package used in the course.	Wickham, H. (2016). ggplot2. Springer International Publishing. doi.org/10.1007/978-3-319-24277-4

Chapter 2 · Cleaning without cheating تنظيف البيانات بدون غش

SECTION	CLAIM IN THE TEXT	SOURCE
k2.1, k2.3	Why the reason data are missing matters (missing completely at random, at random, not at random).	Rubin, D. B. (1976). Inference and missing data. Biometrika 63(3), 581-592. doi.org/10.1093/biomet/63.3.581
k2.2	Many teams analysing the same data reach different answers.	Silberzahn, R., et al. (2018). Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Results. Advances in Methods and Practices in Psychological Science 1(3), 337-356. doi.org/10.1177/2515245917747646
k2.2	Analytical flexibility (p-hacking) can make almost any result look significant.	Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology. Psychological Science 22(11), 1359-1366. doi.org/10.1177/0956797611417632
k2.2	Pre-registration as a defence against analytical flexibility.	Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. Proceedings of the National Academy of Sciences 115(11), 2600-2606. doi.org/10.1073/pnas.1708274114
k2.2, k2.3	What a p-value is and is not.	Wasserstein, R. L., & Lazar, N. A. (2016). The ASA Statement on p -Values: Context, Process, and Purpose. The American Statistician 70(2), 129-133. doi.org/10.1080/00031305.2016.1154108
k2.2, k2.3	The garden of forking paths: many possible analyses inflate false positives.	Gelman, A., & Loken, E. (2014). The Statistical Crisis in Science. American Scientist 102(6), 460. doi.org/10.1511/2014.111.460
k2.2, k2.3	The ICMJE criteria for authorship.	International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals. icmje.org/recommendations

SECTION	CLAIM IN THE TEXT	SOURCE
k2.3	Spearman's rank correlation.	Spearman, C. (1904). The Proof and Measurement of Association between Two Things. The American Journal of Psychology 15(1), 72. doi.org/10.2307/1412159
k2.3	The box plot as a summary of a distribution.	Mcgill, R., Tukey, J. W., & Larsen, W. A. (1978). Variations of Box Plots. The American Statistician 32(1), 12-16. doi.org/10.1080/00031305.1978.10479236
k2.3	The tidyverse packages (dplyr and friends) used for data handling in R.	Wickham, H., et al. (2019). Welcome to the Tidyverse. Journal of Open Source Software 4(43), 1686. doi.org/10.21105/joss.01686
k2.3	ggplot2, the plotting package used in the course.	Wickham, H. (2016). ggplot2. Springer International Publishing. doi.org/10.1007/978-3-319-24277-4
k2.3	The Cancer Genome Atlas: matched tumour, molecular and clinical data.	The Cancer Genome Atlas Research Network, et al. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nature Genetics 45(10), 1113-1120. doi.org/10.1038/ng.2764

Chapter 3 · Looking before testing النظر قبل الاختبار

SECTION	CLAIM IN THE TEXT	SOURCE
k3.1	The t-test for comparing the means of two groups.	Student (1908). The Probable Error of a Mean. Biometrika 6(1), 1. doi.org/10.2307/2331554
k3.1, k3.4	What a p-value is and is not.	Wasserstein, R. L., & Lazar, N. A. (2016). The ASA Statement on p - Values: Context, Process, and Purpose. The American Statistician 70(2), 129-133. doi.org/10.1080/00031305.2016.1154108
k3.1, k3.4	Anscombe's quartet: same summary statistics, very different data.	Anscombe, F. J. (1973). Graphs in Statistical Analysis. The American Statistician 27(1), 17-21. doi.org/10.1080/00031305.1973.10478966
k3.2	ggplot2, the plotting package used in the course.	Wickham, H. (2016). ggplot2. Springer International Publishing. doi.org/10.1007/978-3-319-24277-4
k3.3	The box plot as a summary of a distribution.	Mcgill, R., Tukey, J. W., & Larsen, W. A. (1978). Variations of Box Plots. The American Statistician 32(1), 12-16. doi.org/10.1080/00031305.1978.10479236
k3.4	The Cancer Genome Atlas: matched tumour, molecular and clinical data.	The Cancer Genome Atlas Research Network, et al. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nature Genetics 45(10), 1113-1120. doi.org/10.1038/ng.2764
k3.4	cBioPortal as a way to browse and download cancer genomics data.	Cerami, E., et al. (2012). The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. Cancer Discovery 2(5), 401-404. doi.org/10.1158/2159-8290.cd-12-0095

Chapter 4 · Comparing groups مقارنة المجموعات

SECTION	CLAIM IN THE TEXT	SOURCE
4.1	G6PD is the first enzyme of the pentose phosphate pathway and makes the NADPH cells need for building fats and for coping with oxidative stress.	Stanton, R. C. (2012). Glucose-6-phosphate dehydrogenase, NADPH, and cell survival. IUBMB Life 64(5), 362–369. doi.org/10.1002/iub.1017
4.1	Neyman and Pearson (1933) recast testing as a decision rule with two kinds of error, a significance level fixed in advance and power, so that 'in the long run of experience, we shall not be too often wrong'.	Neyman, J., & Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. Philosophical Transactions of the Royal Society of London, Series A 231(694–706), 289–337. doi.org/10.1098/rsta.1933.0009
4.1	Fisher rejected testing as a decision rule, the two sides argued for decades, and most textbooks now teach a mix of both approaches.	Lehmann, E. L. (1993). The Fisher, Neyman-Pearson theories of testing hypotheses: one theory or two? Journal of the American Statistical Association 88(424), 1242–1249. doi.org/10.1080/01621459.1993.10476404
4.1, 4.2	R's t.test() runs Welch's version, which lets the two groups' spreads differ and adjusts the degrees of freedom, so df need not be a whole number.	Welch, B. L. (1947). The generalization of 'Student's' problem when several different population variances are involved. Biometrika 34(1–2), 28–35. doi.org/10.1093/biomet/34.1-2.28
k4.1, k4.2, k4.3, k4.4	The t-test for comparing the means of two groups.	Student (1908). The Probable Error of a Mean. Biometrika 6(1), 1. doi.org/10.2307/2331554
k4.1, k4.2, k4.3, k4.4	What a p-value is and is not.	Wasserstein, R. L., & Lazar, N. A. (2016). The ASA Statement on p - Values: Context, Process, and Purpose. The American Statistician 70(2), 129-133. doi.org/10.1080/00031305.2016.1154108
4.2	William Gosset worked out the t-test for small samples while at the Guinness brewery in Dublin and published it in 1908 under the pen name 'Student'.	Student. (1908). The probable error of a mean. Biometrika 6(1), 1–25. doi.org/10.2307/2331554
4.2	A 95% confidence interval: if the study were repeated many times, 95% of intervals built this way would contain the true difference.	Neyman, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability. Philosophical Transactions of the Royal Society of London, Series A 236(767), 333–380. doi.org/10.1098/rsta.1937.0005
k4.2, k4.3	The box plot as a summary of a distribution.	Mcgill, R., Tukey, J. W., & Larsen, W. A. (1978). Variations of Box Plots. The American Statistician 32(1), 12-16. doi.org/10.1080/00031305.1978.10479236
k4.2, k4.3, k4.4	Effect size (Cohen's d) as a measure of how big a difference is.	Cohen, J. (1992). A power primer. Psychological Bulletin 112(1), 155-159. doi.org/10.1037/0033-2909.112.1.155

SECTION	CLAIM IN THE TEXT	SOURCE
4.3	Frank Wilcoxon described the two-group rank test in 1945 in a short paper, 'Individual comparisons by ranking methods'.	Wilcoxon, F. (1945). Individual comparisons by ranking methods. Biometrics Bulletin 1(6), 80–83. doi.org/10.2307/3001968
4.3	Henry Mann and D. Ransom Whitney worked out the rank test's mathematics more fully in 1947, hence the names Mann-Whitney U and Wilcoxon-Mann-Whitney.	Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. The Annals of Mathematical Statistics 18(1), 50–60. doi.org/10.1214/aoms/1177730491
4.3	When data really are bell-shaped, the t-test is a little better than the rank test at finding a real difference.	Hodges, J. L., & Lehmann, E. L. (1956). The efficiency of some nonparametric competitors of the t-test. The Annals of Mathematical Statistics 27(2), 324–335. doi.org/10.1214/aoms/1177728261 For normal data the rank test's efficiency relative to the t-test is 3/π, about 0.95.
4.3	The rank test is not a test of medians: it asks whether a value from one group tends to be higher, which concerns medians only if the groups have the same shape.	Divine, G. W., Norton, H. J., Barón, A. E., & Juarez-Colunga, E. (2018). The Wilcoxon–Mann–Whitney procedure fails as a test of medians. The American Statistician 72(3), 278–286. doi.org/10.1080/00031305.2017.1305291
4.3	An effect size for a rank test: the share of cross-group pairs in which one group's value is higher (W divided by the number of pairs).	McGraw, K. O., & Wong, S. P. (1992). A common language effect size statistic. Psychological Bulletin 111(2), 361–365. doi.org/10.1037/0033-2909.111.2.361
4.3	SLC16A1 encodes MCT1, a transporter that moves lactate into and out of cells.	Halestrap, A. P., & Price, N. T. (1999). The proton-linked monocarboxylate transporter (MCT) family: structure, function and regulation. Biochemical Journal 343(2), 281–299. doi.org/10.1042/bj3430281
4.4	The Bonferroni correction divides 0.05 by the number of tests, keeping the chance of even one false positive across all tests below 5%.	Dunn, O. J. (1961). Multiple comparisons among means. Journal of the American Statistical Association 56(293), 52–64. doi.org/10.1080/01621459.1961.10482090 The method rests on Bonferroni's inequality; Dunn's paper set it out as a correction for multiple comparisons.
4.4	Benjamini and Hochberg, at Tel Aviv University, defined the false discovery rate in 1995 (JRSS B) as the expected proportion of errors among rejected hypotheses, with a step-up procedure.	Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, Series B 57(1), 289–300. doi.org/10.1111/j.2517-6161.1995.tb02031.x
4.4	Benjamini and Yekutieli showed in 2001 that the BH procedure still controls the FDR under a common kind of positive dependence between tests.	Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. The Annals of Statistics 29(4), 1165–1188. doi.org/10.1214/aos/1013699998
k4.4	DESeq2 for testing differences in gene expression from raw counts.	Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15(12), 550. doi.org/10.1186/s13059-014-0550-8

Chapter 5 · Relationships and confounding العلاقات والعوامل المربكة

SECTION	CLAIM IN THE TEXT	SOURCE
k5.1	What a p-value is and is not.	Wasserstein, R. L., & Lazar, N. A. (2016). The ASA Statement on p - Values: Context, Process, and Purpose. The American Statistician 70(2), 129-133. doi.org/10.1080/00031305.2016.1154108
k5.1, k5.2, k5.3	Spearman's rank correlation.	Spearman, C. (1904). The Proof and Measurement of Association between Two Things. The American Journal of Psychology 15(1), 72. doi.org/10.2307/1412159
k5.2	The Cancer Genome Atlas: matched tumour, molecular and clinical data.	The Cancer Genome Atlas Research Network, et al. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nature Genetics 45(10), 1113-1120. doi.org/10.1038/ng.2764
k5.2, k5.3	Confounding: a third factor that can explain an apparent link.	Greenland, S., Pearl, J., & Robins, J. M. (1999). Causal Diagrams for Epidemiologic Research. Epidemiology 10(1), 37-48. doi.org/10.1097/00001648-199901000-00008

Chapter 6 · Time to an event الزمن حتى الحدث

SECTION	CLAIM IN THE TEXT	SOURCE
k6.1, k6.2, k6.3, k6.4	The Kaplan–Meier estimate of a survival curve, which handles censored follow-up.	Kaplan, E. L., & Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. Journal of the American Statistical Association 53(282), 457-481. doi.org/10.1080/01621459.1958.10501452
k6.1, k6.3	The survival package in R (survfit, coxph).	Therneau, T. M. survival: Survival Analysis. R package. CRAN.R-project.org/package=survival
k6.2, k6.3, k6.4	Cox proportional hazards regression and the hazard ratio.	Cox, D. R. (1972). Regression Models and Life-Tables. Journal of the Royal Statistical Society Series B: Statistical Methodology 34(2), 187-202. doi.org/10.1111/j.2517-6161.1972.tb00899.x
k6.2, k6.4	The log-rank test for comparing survival curves.	Peto, R., & Peto, J. (1972). Asymptotically Efficient Rank Invariant Test Procedures. Journal of the Royal Statistical Society. Series A (General) 135(2), 185. doi.org/10.2307/2344317
k6.3	Checking the proportional hazards assumption with Schoenfeld residuals (cox.zph).	Grambsch, P. M., & Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. Biometrika 81(3), 515-526. doi.org/10.1093/biomet/81.3.515
k6.3	Confounding: a third factor that can explain an apparent link.	Greenland, S., Pearl, J., & Robins, J. M. (1999). Causal Diagrams for Epidemiologic Research. Epidemiology 10(1), 37-48. doi.org/10.1097/00001648-199901000-00008
k6.3, k6.4	What a p-value is and is not.	Wasserstein, R. L., & Lazar, N. A. (2016). The ASA Statement on p - Values: Context, Process, and Purpose. The American Statistician 70(2), 129-133. doi.org/10.1080/00031305.2016.1154108
k6.4	Controlling the false discovery rate when many tests are run.	Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B: Statistical Methodology 57(1), 289-300. doi.org/10.1111/j.2517-6161.1995.tb02031.x

Chapter 7 · Models and their limits النماذج وحدودها

SECTION	CLAIM IN THE TEXT	SOURCE
7.1	The first human cell line, HeLa, came from Henrietta Lacks's cervical tumour at Johns Hopkins in 1951, grown in George Gey's lab, taken without her or her family's consent.	Nott, R. (2020). HeLa cell line. Embryo Project Encyclopedia. Arizona State University. The Embryo Project dates the sample to 1951, after she first sought care in January; it gives no month for the biopsy.
7.1	Prostate cancer is one of the commonest cancers in men.	Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., & Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians 74(3), 229–263. doi.org/10.3322/caac.21834 In GLOBOCAN 2022 prostate cancer is the second most common cancer in men worldwide, after lung.
7.1	High-glucose DMEM has about 25 mM glucose, roughly five times normal blood glucose (about 5 mM).	Thermo Fisher Scientific. (n.d.). DMEM, high glucose (11965) media formulation. The formulation lists D-glucose at 4,500 mg/L (25.0 mM).
7.1	Cantor and colleagues (2017) built a human plasma-like medium; uric acid, ~10x higher in human than mouse blood, inhibits UMP synthase and lowers cancer-cell sensitivity to 5-fluorouracil.	Cantor, J. R., et al. (2017). Physiologic medium rewires cellular metabolism and reveals uric acid as an endogenous inhibitor of UMP synthase. Cell 169(2), 258–272.e17. doi.org/10.1016/j.cell.2017.03.023
7.1	27 MCF7 stocks from different labs differed genetically; tested against 321 compounds, at least 75% of compounds that strongly inhibited some stocks were inactive in others.	Ben-David, U., et al. (2018). Genetic and transcriptional evolution alters cancer cell line drug response. Nature 560(7718), 325–330. doi.org/10.1038/s41586-018-0409-3
7.1	In 1967 Stanley Gartler reported that HeLa cells had contaminated other human cell lines.	Gartler, S. M. (1967). Genetic markers as tracers in cell culture. National Cancer Institute Monograph 26, 167–195. The better-known journal version is Gartler, S. M. (1968), Nature 217(5130), 750–751, doi.org/10.1038/217750a0.
7.1	MDA-MB-435 was started at MD Anderson in 1976 from a woman with breast cancer.	Cailleau, R., Olivé, M., & Cruciger, Q. V. J. (1978). Long-term human breast carcinoma cell lines of metastatic origin: Preliminary characterization. In Vitro 14(11), 911–915. doi.org/10.1007/BF02616120 The NCI's note on the line gives 1976, from the pleural effusion of a 31-year-old woman, citing this paper.
7.1	A 2000 gene expression study of 60 cancer lines placed MDA-MB-435 among the melanoma lines.	Ross, D. T., et al. (2000). Systematic variation in gene expression patterns in human cancer cell lines. Nature Genetics 24(3), 227–235. doi.org/10.1038/73432
7.1	A 2005 genetic (SNP) analysis showed MDA-MB-435 came from the same person as the melanoma line M14.	Garraway, L. A., et al. (2005). Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. Nature 436(7047), 117–122. doi.org/10.1038/nature03664

SECTION	CLAIM IN THE TEXT	SOURCE
7.1	In 2007 James Rae and colleagues concluded that every available stock of MDA-MB-435 was the melanoma line M14.	Rae, J. M., Creighton, C. J., Meck, J. M., Haddad, B. R., & Johnson, M. D. (2007). MDA-MB-435 cells are derived from M14 melanoma cells —a loss for breast cancer, but a boon for melanoma research. Breast Cancer Research and Treatment 104(1), 13–19. doi.org/10.1007/s10549-006-9392-8
7.1	By one count, more than 650 published breast cancer studies had used MDA-MB-435.	Scudellari, M. (n.d.). A case of mistaken identity. The Scientist.
7.1	ICLAC keeps a public register of misidentified lines; labs check identity by comparing a line's STR profile with the reference.	International Cell Line Authentication Committee. (2026). Register of Misidentified Cell Lines (Version 14). Version 14 (15 February 2026) lists 608 lines; the register began as Capes-Davis, A., et al. (2010), International Journal of Cancer 127(1), 1–8, doi.org/10.1002/ijc.25242.
7.1, 7.2	DepMap, a Broad Institute project, knocks out thousands of genes one at a time in over a thousand cancer cell lines to find which genes each line needs.	Tsherniak, A., et al. (2017). Defining a cancer dependency map. Cell 170(3), 564–576.e16. doi.org/10.1016/j.cell.2017.06.010
7.1, 7.2, 7.3	Our scores come from DepMap release 24Q4: 1,178 cancer cell lines from 30 lineages, including 53 breast lines; releases change scores, so name the one used.	DepMap, Broad. (2024). DepMap 24Q4 Public [Data set]. Figshare+. doi.org/10.25452/figshare.plus.27993248
7.1, 7.3	Lines are grouped by a curated lineage column (OncotreeLineage), the OncoTree cancer classification, not by their names.	Kundra, R., et al. (2021). OncoTree: A cancer classification system for precision oncology. JCO Clinical Cancer Informatics 5, 221–230. doi.org/10.1200/CCI.20.00108
7.2	A pooled CRISPR knockout screen: guide RNAs steer Cas9 to cut each gene, cells grow for weeks, and guides against needed genes drop out when counted by sequencing.	Shalem, O., et al. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. Science 343(6166), 84–87. doi.org/10.1126/science.1247005
7.2	Guides cutting amplified DNA make many cuts, and the damage alone slows cells, so early CRISPR screens read every gene in amplified regions as a dependency.	Aguirre, A. J., et al. (2016). Genomic copy number dictates a gene-independent cell response to CRISPR/Cas9 targeting. Cancer Discovery 6(8), 914–929. doi.org/10.1158/2159-8290.CD-16-0154
k7.2, k7.3	The Cancer Genome Atlas: matched tumour, molecular and clinical data.	The Cancer Genome Atlas Research Network, et al. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nature Genetics 45(10), 1113-1120. doi.org/10.1038/ng.2764
7.2, 7.3	DepMap scores come from the Chronos model, which combines guides per gene and corrects for biases such as copy number; scaled so non-essential genes' median is 0 and common essentials' -1.	Dempster, J. M., et al. (2021). Chronos: A cell population dynamics model of CRISPR experiments that improves inference of gene fitness effects. Genome Biology 22, 343. doi.org/10.1186/s13059-021-02540-7 The Chronos paper recommends shifting and scaling the whole matrix globally, not per cell line.
7.3	PHGDH expression is highest in Basal tumours in the TCGA breast data (PanCancer Atlas, through cBioPortal).	Hoadley, K. A., et al. (2018). Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. Cell 173(2), 291–304.e6. doi.org/10.1016/j.cell.2018.03.022

SECTION	CLAIM IN THE TEXT	SOURCE
7.3	Possemato et al. (2011): PHGDH sits in a region often gained in breast cancer, its protein is raised in ~70% of ER-negative tumours, and knockdown slows only high-PHGDH lines.	Possemato, R., et al. (2011). Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. Nature 476(7360), 346–350. doi.org/10.1038/nature10350
7.3	Chen et al. (2013): shRNA knockdown of PHGDH slowed MDA-MB-468, BT-20 and HCC1806 in colony assays, but established xenografts of these lines kept growing.	Chen, J., et al. (2013). Phosphoglycerate dehydrogenase is dispensable for breast tumor maintenance and growth. Oncotarget 4(12), 2502–2511. doi.org/10.18632/oncotarget.1540
7.3	Ngo et al. (2020): brain metastases depend on PHGDH because the brain has little serine and glycine; blocking it reduced brain metastases in mice but not growth elsewhere.	Ngo, B., et al. (2020). Limited environmental serine and glycine confer brain metastasis sensitivity to PHGDH inhibition. Cancer Discovery 10(9), 1352–1373. doi.org/10.1158/2159-8290.CD-19-1228
7.3	Lung cancer cells in culture fuel the TCA cycle with glutamine, but lung tumours in mice used little glutamine and leaned more on glucose.	Davidson, S. M., et al. (2016). Environment impacts the metabolic dependencies of Ras-driven non-small cell lung cancer. Cell Metabolism 23(3), 517–528. doi.org/10.1016/j.cmet.2016.01.007
7.3	Of 47 ovarian lines, SK-OV-3 and A2780 (about 60% of publications) matched high-grade serous tumours poorly; the twelve best, led by KURAMOCHI, had about 1% of publications.	Domcke, S., Sinha, R., Levine, D. A., Sander, C., & Schultz, N. (2013). Evaluating cell lines as tumour models by comparison of genomic profiles. Nature Communications 4, 2126. doi.org/10.1038/ncomms3126

Chapter 8 · Trusting an analysis الثقة في التحليل

SECTION	CLAIM IN THE TEXT	SOURCE
k8.1	Posit Cloud, where we run R in the browser.	Posit Software. Posit Cloud. posit.cloud
k8.1	Retractions: why published papers get withdrawn.	Fang, F. C., Steen, R. G., & Casadevall, A. (2012). Misconduct accounts for the majority of retracted scientific publications. Proceedings of the National Academy of Sciences 109(42), 17028-17033. doi.org/10.1073/pnas.1212247109
k8.2	The t-test for comparing the means of two groups.	Student (1908). The Probable Error of a Mean. Biometrika 6(1), 1. doi.org/10.2307/2331554
k8.2, k8.3, k8.4	The Wilcoxon rank-based test, used when the data are not normally distributed.	Wilcoxon, F. (1945). Individual Comparisons by Ranking Methods. Biometrics Bulletin 1(6), 80. doi.org/10.2307/3001968
k8.2, k8.3, k8.4	What a p-value is and is not.	Wasserstein, R. L., & Lazar, N. A. (2016). The ASA Statement on p -Values: Context, Process, and Purpose. The American Statistician 70(2), 129-133. doi.org/10.1080/00031305.2016.1154108

SECTION	CLAIM IN THE TEXT	SOURCE
k8.2, k8.3, k8.4	DepMap: genome-wide dependency screens across cancer cell lines.	Tsherniak, A., et al. (2017). Defining a Cancer Dependency Map. Cell 170(3), 564-576.e16. doi.org/10.1016/j.cell.2017.06.010
k8.3	Cox proportional hazards regression and the hazard ratio.	Cox, D. R. (1972). Regression Models and Life-Tables. Journal of the Royal Statistical Society Series B: Statistical Methodology 34(2), 187-202. doi.org/10.1111/j.2517-6161.1972.tb00899.x
k8.3	Many teams analysing the same data reach different answers.	Silberzahn, R., et al. (2018). Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Results. Advances in Methods and Practices in Psychological Science 1(3), 337-356. doi.org/10.1177/2515245917747646
k8.3	Pre-registration as a defence against analytical flexibility.	Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. Proceedings of the National Academy of Sciences 115(11), 2600-2606. doi.org/10.1073/pnas.1708274114
k8.3	The Cancer Genome Atlas: matched tumour, molecular and clinical data.	The Cancer Genome Atlas Research Network, et al. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nature Genetics 45(10), 1113-1120. doi.org/10.1038/ng.2764
k8.3	Survival data in TCGA and its curated survival endpoints.	Liu, J., et al. (2018). An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. Cell 173(2), 400-416.e11. doi.org/10.1016/j.cell.2018.02.052
k8.3, k8.4	The garden of forking paths: many possible analyses inflate false positives.	Gelman, A., & Loken, E. (2014). The Statistical Crisis in Science. American Scientist 102(6), 460. doi.org/10.1511/2014.111.460
k8.4	Effect size (Cohen's d) as a measure of how big a difference is.	Cohen, J. (1992). A power primer. Psychological Bulletin 112(1), 155-159. doi.org/10.1037/0033-2909.112.1.155
k8.4	HARKing: presenting a hypothesis made after seeing the results as if it came first.	Kerr, N. L. (1998). HARKing: Hypothesizing After the Results are Known. Personality and Social Psychology Review 2(3), 196-217. doi.org/10.1207/s15327957pspr0203_4
k8.4	CONSORT for reporting randomised trials.	Schulz, K. F., Altman, D. G., Moher, D., & for the CONSORT Group (2010). CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMJ 340(mar23 1), c332-c332. doi.org/10.1136/bmj.c332
k8.4	The ICMJE criteria for authorship.	International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals. icmje.org/recommendations

Skills sessions

25 CLAIMS

SECTION	CLAIM IN THE TEXT	SOURCE
Session 1	R, the free language for statistics used in the course.	R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. r-project.org

SECTION	CLAIM IN THE TEXT	SOURCE
Session 1, Session 2, Session 3, Session 4, Session 5, Session 6, Session 7, Session 8, Session 9, Session 10, Session 11, Session 12	The tidyverse packages (dplyr and friends) used for data handling in R.	Wickham, H., et al. (2019). Welcome to the Tidyverse. Journal of Open Source Software 4(43), 1686. doi.org/10.21105/joss.01686
Session 1, Session 4, Session 5, Session 6, Session 7, Session 8, Session 11, Session 12	ggplot2, the plotting package used in the course.	Wickham, H. (2016). ggplot2. Springer International Publishing. doi.org/10.1007/978-3-319-24277-4
Session 1, Session 12, Session 13, Session 14, Session plan	Posit Cloud, where we run R in the browser.	Posit Software. Posit Cloud. posit.cloud

SECTION	CLAIM IN THE TEXT	SOURCE
Session 2, Session 3, Session 4, Session 5, Session 8, Session 13, Session plan	The Cancer Genome Atlas: matched tumour, molecular and clinical data.	The Cancer Genome Atlas Research Network, et al. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nature Genetics 45(10), 1113-1120. doi.org/10.1038/ng.2764
Session 2, Session 13, Session plan	cBioPortal as a way to browse and download cancer genomics data.	Cerami, E., et al. (2012). The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. Cancer Discovery 2(5), 401-404. doi.org/10.1158/2159-8290.cd-12-0095
Session 4, Session 5, Session 6, Session 11	The box plot as a summary of a distribution.	Mcgill, R., Tukey, J. W., & Larsen, W. A. (1978). Variations of Box Plots. The American Statistician 32(1), 12-16. doi.org/10.1080/00031305.1978.10479236
Session 5, Session 6, Session 7, Session 13, Session plan	The t-test for comparing the means of two groups.	Student (1908). The Probable Error of a Mean. Biometrika 6(1), 1. doi.org/10.2307/2331554
Session 6	The Mann–Whitney U test for comparing two independent groups by rank.	Mann, H. B., & Whitney, D. R. (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. The Annals of Mathematical Statistics 18(1), 50-60. doi.org/10.1214/aoms/1177730491
Session 6, Session 7, Session 8, Session 10	What a p-value is and is not.	Wasserstein, R. L., & Lazar, N. A. (2016). The ASA Statement on p -Values: Context, Process, and Purpose. The American Statistician 70(2), 129-133. doi.org/10.1080/00031305.2016.1154108

SECTION	CLAIM IN THE TEXT	SOURCE
Session 6, Session 7, Session 11, Session 13, Session 14, Session plan	Effect size (Cohen's d) as a measure of how big a difference is.	Cohen, J. (1992). A power primer. Psychological Bulletin 112(1), 155-159. doi.org/10.1037/0033-2909.112.1.155
Session 6, Session 7, Session 13	Welch's t-test, which does not assume the two groups have equal variances.	Welch, B. L. (1947). The Generalization OF 'Student'S' Problem When Several Different Population Varlances Are Involved. Biometrika 34(1-2), 28-35. doi.org/10.1093/biomet/34.1-2.28
Session 6, Session 11, Session 13	The Wilcoxon rank-based test, used when the data are not normally distributed.	Wilcoxon, F. (1945). Individual Comparisons by Ranking Methods. Biometrics Bulletin 1(6), 80. doi.org/10.2307/3001968
Session 7, Session plan	Controlling the false discovery rate when many tests are run.	Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B: Statistical Methodology 57(1), 289-300. doi.org/10.1111/j.2517-6161.1995.tb02031.x
Session 7	The Bonferroni correction for multiple comparisons.	Dunn, O. J. (1961). Multiple Comparisons among Means. Journal of the American Statistical Association 56(293), 52-64. doi.org/10.1080/01621459.1961.10482090
Session 8	Spearman's rank correlation.	Spearman, C. (1904). The Proof and Measurement of Association between Two Things. The American Journal of Psychology 15(1), 72. doi.org/10.2307/1412159
Session 8, Session 11	Confounding: a third factor that can explain an apparent link.	Greenland, S., Pearl, J., & Robins, J. M. (1999). Causal Diagrams for Epidemiologic Research. Epidemiology 10(1), 37-48. doi.org/10.1097/00001648-199901000-00008
Session 9	The chi-squared test for an association between two categorical variables.	Pearson, K. (1900). X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 50(302), 157-175. doi.org/10.1080/14786440009463897
Session 9	The log-rank test for comparing survival curves.	Peto, R., & Peto, J. (1972). Asymptotically Efficient Rank Invariant Test Procedures. Journal of the Royal Statistical Society. Series A (General) 135(2), 185. doi.org/10.2307/2344317

SECTION	CLAIM IN THE TEXT	SOURCE
Session 9, Session 10, Session plan	The Kaplan–Meier estimate of a survival curve, which handles censored follow-up.	Kaplan, E. L., & Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. Journal of the American Statistical Association 53(282), 457-481. doi.org/10.1080/01621459.1958.10501452
Session 9, Session 10, Session plan	Cox proportional hazards regression and the hazard ratio.	Cox, D. R. (1972). Regression Models and Life-Tables. Journal of the Royal Statistical Society Series B: Statistical Methodology 34(2), 187-202. doi.org/10.1111/j.2517-6161.1972.tb00899.x
Session 9, Session 10	The survival package in R (survfit, coxph).	Therneau, T. M. survival: Survival Analysis. R package. CRAN.R-project.org/package=survival
Session 10	Checking the proportional hazards assumption with Schoenfeld residuals (cox.zph).	Grambsch, P. M., & Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. Biometrika 81(3), 515-526. doi.org/10.1093/biomet/81.3.515
Session 11, Session 12, Session plan	DepMap: genome-wide dependency screens across cancer cell lines.	Tsherniak, A., et al. (2017). Defining a Cancer Dependency Map. Cell 170(3), 564-576.e16. doi.org/10.1016/j.cell.2017.06.010
Session 13	Peer review: what it checks and what it misses.	Smith, R. (2006). Peer Review: A Flawed Process at the Heart of Science and Journals. Journal of the Royal Society of Medicine 99(4), 178-182. doi.org/10.1177/014107680609900414

Session teaching notes

22 CLAIMS

SECTION	CLAIM IN THE TEXT	SOURCE
Session 1	The tidyverse packages (dplyr and friends) used for data handling in R.	Wickham, H., et al. (2019). Welcome to the Tidyverse. Journal of Open Source Software 4(43), 1686. doi.org/10.21105/joss.01686
Session 1, Session 13, Session 14	Posit Cloud, where we run R in the browser.	Posit Software. Posit Cloud. posit.cloud
Session 2, Session 3, Session 8, Session 13	The Cancer Genome Atlas: matched tumour, molecular and clinical data.	The Cancer Genome Atlas Research Network, et al. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nature Genetics 45(10), 1113-1120. doi.org/10.1038/ng.2764

SECTION	CLAIM IN THE TEXT	SOURCE
Session 3	Why the reason data are missing matters (missing completely at random, at random, not at random).	Rubin, D. B. (1976). Inference and missing data. Biometrika 63(3), 581-592. doi.org/10.1093/biomet/63.3.581
Session 3, Session 4, Session 6, Session 7, Session 8, Session 10 and more	What a p-value is and is not.	Wasserstein, R. L., & Lazar, N. A. (2016). The ASA Statement on p -Values: Context, Process, and Purpose. The American Statistician 70(2), 129-133. doi.org/10.1080/00031305.2016.1154108
Session 3, Session 8	Spearman's rank correlation.	Spearman, C. (1904). The Proof and Measurement of Association between Two Things. The American Journal of Psychology 15(1), 72. doi.org/10.2307/1412159
Session 4	Anscombe's quartet: same summary statistics, very different data.	Anscombe, F. J. (1973). Graphs in Statistical Analysis. The American Statistician 27(1), 17-21. doi.org/10.1080/00031305.1973.10478966
Session 4	ggplot2, the plotting package used in the course.	Wickham, H. (2016). ggplot2. Springer International Publishing. doi.org/10.1007/978-3-319-24277-4
Session 5	The box plot as a summary of a distribution.	Mcgill, R., Tukey, J. W., & Larsen, W. A. (1978). Variations of Box Plots. The American Statistician 32(1), 12-16. doi.org/10.1080/00031305.1978.10479236
Session 6, Session 7, Session 11, Session 13, Session 14	Effect size (Cohen's d) as a measure of how big a difference is.	Cohen, J. (1992). A power primer. Psychological Bulletin 112(1), 155-159. doi.org/10.1037/0033-2909.112.1.155
Session 6, Session 7, Session 13	The t-test for comparing the means of two groups.	Student (1908). The Probable Error of a Mean. Biometrika 6(1), 1. doi.org/10.2307/2331554
Session 6, Session 13	Welch's t-test, which does not assume the two groups have equal variances.	Welch, B. L. (1947). The Generalization OF 'Student'S' Problem When Several Different Population Varlances Are Involved. Biometrika 34(1-2), 28-35. doi.org/10.1093/biomet/34.1-2.28
Session 7	The Bonferroni correction for multiple comparisons.	Dunn, O. J. (1961). Multiple Comparisons among Means. Journal of the American Statistical Association 56(293), 52-64. doi.org/10.1080/01621459.1961.10482090

SECTION	CLAIM IN THE TEXT	SOURCE
Session 7	Controlling the false discovery rate when many tests are run.	Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B: Statistical Methodology 57(1), 289-300. doi.org/10.1111/j.2517-6161.1995.tb02031.x
Session 8	Confounding: a third factor that can explain an apparent link.	Greenland, S., Pearl, J., & Robins, J. M. (1999). Causal Diagrams for Epidemiologic Research. Epidemiology 10(1), 37-48. doi.org/10.1097/00001648-199901000-00008
Session 9, Session 10	The Kaplan–Meier estimate of a survival curve, which handles censored follow-up.	Kaplan, E. L., & Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. Journal of the American Statistical Association 53(282), 457-481. doi.org/10.1080/01621459.1958.10501452
Session 9, Session 10	The log-rank test for comparing survival curves.	Peto, R., & Peto, J. (1972). Asymptotically Efficient Rank Invariant Test Procedures. Journal of the Royal Statistical Society. Series A (General) 135(2), 185. doi.org/10.2307/2344317
Session 10, Session 11	Cox proportional hazards regression and the hazard ratio.	Cox, D. R. (1972). Regression Models and Life-Tables. Journal of the Royal Statistical Society Series B: Statistical Methodology 34(2), 187-202. doi.org/10.1111/j.2517-6161.1972.tb00899.x
Session 11, Session 12, Session 14	DepMap: genome-wide dependency screens across cancer cell lines.	Tsherniak, A., et al. (2017). Defining a Cancer Dependency Map. Cell 170(3), 564-576.e16. doi.org/10.1016/j.cell.2017.06.010
Session 13	Many teams analysing the same data reach different answers.	Silberzahn, R., et al. (2018). Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Results. Advances in Methods and Practices in Psychological Science 1(3), 337-356. doi.org/10.1177/2515245917747646
Session 14	HARKing: presenting a hypothesis made after seeing the results as if it came first.	Kerr, N. L. (1998). HARKing: Hypothesizing After the Results are Known. Personality and Social Psychology Review 2(3), 196-217. doi.org/10.1207/s15327957pspr0203_4
Session 14	The Wilcoxon rank-based test, used when the data are not normally distributed.	Wilcoxon, F. (1945). Individual Comparisons by Ranking Methods. Biometrics Bulletin 1(6), 80. doi.org/10.2307/3001968

Skills assignments

19 CLAIMS

SECTION	CLAIM IN THE TEXT	SOURCE
Assignments	The tidyverse packages (dplyr and friends) used for data handling in R.	Wickham, H., et al. (2019). Welcome to the Tidyverse. Journal of Open Source Software 4(43), 1686. doi.org/10.21105/joss.01686
Assignments	The Cancer Genome Atlas: matched tumour, molecular and clinical data.	The Cancer Genome Atlas Research Network, et al. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nature Genetics 45(10), 1113-1120. doi.org/10.1038/ng.2764
Assignments	Posit Cloud, where we run R in the browser.	Posit Software. Posit Cloud. posit.cloud

SECTION	CLAIM IN THE TEXT	SOURCE
Assignments	The t-test for comparing the means of two groups.	Student (1908). The Probable Error of a Mean. Biometrika 6(1), 1. doi.org/10.2307/2331554
Assignments	Welch's t-test, which does not assume the two groups have equal variances.	Welch, B. L. (1947). The Generalization OF 'Student'S' Problem When Several Different Population Varlances Are Involved. Biometrika 34(1-2), 28-35. doi.org/10.1093/biomet/34.1-2.28
Assignments	The Wilcoxon rank-based test, used when the data are not normally distributed.	Wilcoxon, F. (1945). Individual Comparisons by Ranking Methods. Biometrics Bulletin 1(6), 80. doi.org/10.2307/3001968
Assignments	Effect size (Cohen's d) as a measure of how big a difference is.	Cohen, J. (1992). A power primer. Psychological Bulletin 112(1), 155-159. doi.org/10.1037/0033-2909.112.1.155
Assignments	What a p-value is and is not.	Wasserstein, R. L., & Lazar, N. A. (2016). The ASA Statement on p -Values: Context, Process, and Purpose. The American Statistician 70(2), 129-133. doi.org/10.1080/00031305.2016.1154108
Assignments	Spearman's rank correlation.	Spearman, C. (1904). The Proof and Measurement of Association between Two Things. The American Journal of Psychology 15(1), 72. doi.org/10.2307/1412159
Assignments	Confounding: a third factor that can explain an apparent link.	Greenland, S., Pearl, J., & Robins, J. M. (1999). Causal Diagrams for Epidemiologic Research. Epidemiology 10(1), 37-48. doi.org/10.1097/00001648-199901000-00008
Assignments	The Kaplan–Meier estimate of a survival curve, which handles censored follow-up.	Kaplan, E. L., & Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. Journal of the American Statistical Association 53(282), 457-481. doi.org/10.1080/01621459.1958.10501452
Assignments	The log-rank test for comparing survival curves.	Peto, R., & Peto, J. (1972). Asymptotically Efficient Rank Invariant Test Procedures. Journal of the Royal Statistical Society. Series A (General) 135(2), 185. doi.org/10.2307/2344317
Assignments	Cox proportional hazards regression and the hazard ratio.	Cox, D. R. (1972). Regression Models and Life-Tables. Journal of the Royal Statistical Society Series B: Statistical Methodology 34(2), 187-202. doi.org/10.1111/j.2517-6161.1972.tb00899.x
Assignments	Checking the proportional hazards assumption with Schoenfeld residuals (cox.zph).	Grambsch, P. M., & Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. Biometrika 81(3), 515-526. doi.org/10.1093/biomet/81.3.515
Assignments	The chi-squared test for an association between two categorical variables.	Pearson, K. (1900). X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 50(302), 157-175. doi.org/10.1080/14786440009463897
Assignments	DepMap: genome-wide dependency screens across cancer cell lines.	Tsherniak, A., et al. (2017). Defining a Cancer Dependency Map. Cell 170(3), 564-576.e16. doi.org/10.1016/j.cell.2017.06.010

SECTION	CLAIM IN THE TEXT	SOURCE
Assignments	Survival data in TCGA and its curated survival endpoints.	Liu, J., et al. (2018). An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. Cell 173(2), 400-416.e11. doi.org/10.1016/j.cell.2018.02.052
Assignments	Controlling the false discovery rate when many tests are run.	Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B: Statistical Methodology 57(1), 289-300. doi.org/10.1111/j.2517-6161.1995.tb02031.x
Assignments	Peer review: what it checks and what it misses.	Smith, R. (2006). Peer Review: A Flawed Process at the Heart of Science and Journals. Journal of the Royal Society of Medicine 99(4), 178-182. doi.org/10.1177/014107680609900414

PROGRAMME 3

Principles of Research مبادئ البحث

The nine Principles guides, the session plans and teaching notes, and the Principles assignments. 140 claims, 64 sources.

Guide 0 · Arriving الوصول

SECTION	CLAIM IN THE TEXT	SOURCE
g0.0, g0.3, g0.4, g0.5, g0.7	Zotero, the free reference manager.	Corporation for Digital Scholarship. Zotero. zotero.org
g0.0, g0.3, g0.4, g0.5, g0.7	PubMed, the free search engine for the biomedical literature.	National Library of Medicine. PubMed. pubmed.ncbi.nlm.nih.gov
g0.3, g0.4	Google Scholar for searching across disciplines.	Google. Google Scholar. scholar.google.com

Guide 1 · How research works كيف يشتغل البحث

SECTION	CLAIM IN THE TEXT	SOURCE
g1.1	Peer review: what it checks and what it misses.	Smith, R. (2006). Peer Review: A Flawed Process at the Heart of Science and Journals. Journal of the Royal Society of Medicine 99(4), 178-182. doi.org/10.1177/014107680609900414
g1.1, g1.7	The Warburg effect: tumours take up glucose and make lactate even with oxygen present.	Vander Heiden, M. G., Cantley, L. C., & Thompson, C. B. (2009). Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation. Science 324(5930), 1029-1033. doi.org/10.1126/science.1160809
g1.2	A hypothesis earns credibility by surviving serious attempts to refute it.	Popper, K. (2002). The Logic of Scientific Discovery. Routledge (first English edition 1959). ISBN 9780415278447
g1.3, g1.4, g1.6	Confounding: a third factor that can explain an apparent link.	Greenland, S., Pearl, J., & Robins, J. M. (1999). Causal Diagrams for Epidemiologic Research. Epidemiology 10(1), 37-48. doi.org/10.1097/00001648-199901000-00008
g1.4	The Cancer Genome Atlas: matched tumour, molecular and clinical data.	The Cancer Genome Atlas Research Network, et al. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nature Genetics 45(10), 1113-1120. doi.org/10.1038/ng.2764

SECTION	CLAIM IN THE TEXT	SOURCE
g1.5	Publication bias: positive results are more likely to be published.	Dickersin, K. (1990). The Existence of Publication Bias and Risk Factors for Its Occurrence. JAMA: The Journal of the American Medical Association 263(10), 1385. doi.org/10.1001/jama.1990.03440100097014
g1.5, g1.6	Effect size (Cohen's d) as a measure of how big a difference is.	Cohen, J. (1992). A power primer. Psychological Bulletin 112(1), 155-159. doi.org/10.1037/0033-2909.112.1.155
g1.5, g1.6	Immortal time bias in survival comparisons.	Suissa, S. (2008). Immortal Time Bias in Pharmacoepidemiology. American Journal of Epidemiology 167(4), 492-499. doi.org/10.1093/aje/kwm324
g1.5, g1.7	Most landmark preclinical cancer findings could not be confirmed.	Begley, C. G., & Ellis, L. M. (2012). Raise standards for preclinical cancer research. Nature 483(7391), 531-533. doi.org/10.1038/483531a
g1.6, g1.7	PubMed, the free search engine for the biomedical literature.	National Library of Medicine. PubMed. pubmed.ncbi.nlm.nih.gov
g1.6, g1.7	Google Scholar for searching across disciplines.	Google. Google Scholar. scholar.google.com

Guide 2 · Reading a paper كيف تقرأ ورقة علمية

SECTION	CLAIM IN THE TEXT	SOURCE
g2.1, g2.8	Reading a paper in three passes.	Keshav, S. (2007). How to read a paper. ACM SIGCOMM Computer Communication Review 37(3), 83-84. doi.org/10.1145/1273445.1273458
g2.2	The Gene Expression Omnibus, a public repository of expression data.	Edgar, R. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Research 30(1), 207-210. doi.org/10.1093/nar/30.1.207
g2.3	What a p-value is and is not.	Wasserstein, R. L., & Lazar, N. A. (2016). The ASA Statement on p - Values: Context, Process, and Purpose. The American Statistician 70(2), 129-133. doi.org/10.1080/00031305.2016.1154108
g2.3	Spin: reporting that makes results look better than they are.	Boutron, I. (2010). Reporting and Interpretation of Randomized Controlled Trials With Statistically Nonsignificant Results for Primary Outcomes. JAMA 303(20), 2058. doi.org/10.1001/jama.2010.651
g2.4	Confounding: a third factor that can explain an apparent link.	Greenland, S., Pearl, J., & Robins, J. M. (1999). Causal Diagrams for Epidemiologic Research. Epidemiology 10(1), 37-48. doi.org/10.1097/00001648-199901000-00008
g2.6, g2.8	Retractions: why published papers get withdrawn.	Fang, F. C., Steen, R. G., & Casadevall, A. (2012). Misconduct accounts for the majority of retracted scientific publications. Proceedings of the National Academy of Sciences 109(42), 17028-17033. doi.org/10.1073/pnas.1212247109
g2.6, g2.8	Retraction Watch and its database of retracted papers.	Retraction Watch Database. The Center for Scientific Integrity. retractiondatabase.org

SECTION	CLAIM IN THE TEXT	SOURCE
g2.6, g2.8	Zotero, the free reference manager.	Corporation for Digital Scholarship. Zotero. zotero.org

Guide 3 · Finding sources, and knowing which to trust

كيف تلاقي المصادر، وتعرف أي مصدر تثق فيه

SECTION	CLAIM IN THE TEXT	SOURCE
g3.1	Combining search terms with AND, OR and NOT.	National Library of Medicine. PubMed User Guide. pubmed.ncbi.nlm.nih.gov/help
g3.1, g3.2, g3.3, g3.5, g3.6, g3.7	PubMed, the free search engine for the biomedical literature.	National Library of Medicine. PubMed. pubmed.ncbi.nlm.nih.gov
g3.2	MeSH, the controlled vocabulary used to index PubMed.	National Library of Medicine. Medical Subject Headings (MeSH). nlm.nih.gov/mesh
g3.2, g3.5, g3.6, g3.7	Retractions: why published papers get withdrawn.	Fang, F. C., Steen, R. G., & Casadevall, A. (2012). Misconduct accounts for the majority of retracted scientific publications. Proceedings of the National Academy of Sciences 109(42), 17028-17033. doi.org/10.1073/pnas.1212247109
g3.3, g3.4, g3.5, g3.6, g3.7	Peer review: what it checks and what it misses.	Smith, R. (2006). Peer Review: A Flawed Process at the Heart of Science and Journals. Journal of the Royal Society of Medicine 99(4), 178-182. doi.org/10.1177/014107680609900414
g3.3, g3.4, g3.5, g3.6, g3.7	Google Scholar for searching across disciplines.	Google. Google Scholar. scholar.google.com
g3.4, g3.5	Fabrication, falsification and plagiarism as research misconduct.	US Office of Research Integrity. Definition of Research Misconduct, 42 CFR Part 93. ori.hhs.gov/definition-research-misconduct
g3.5, g3.6, g3.7	Retraction Watch and its database of retracted papers.	Retraction Watch Database. The Center for Scientific Integrity. retractiondatabase.org
g3.5, g3.6	Zotero, the free reference manager.	Corporation for Digital Scholarship. Zotero. zotero.org
g3.5, g3.6, g3.7	COPE guidance on publication ethics and corrections.	Committee on Publication Ethics (COPE). Core Practices. publicationethics.org/core-practices
g3.6, g3.7	Predatory journals and how to spot them.	Grudniewicz, A., et al. (2019). Predatory journals: no definition, no defence. Nature 576(7786), 210-212. doi.org/10.1038/d41586-019-03759-y

Guide 4 · Using AI well in science كيف تستخدم الذكاء الاصطناعي صح في العلم

SECTION	CLAIM IN THE TEXT	SOURCE
4.1, 4.2	Chat tools are built on large language models trained on huge amounts of text, which write by predicting the most likely next word.	Brown, T. B., et al. (2020). Language Models are Few-Shot Learners. arXiv 2005.14165. doi.org/10.48550/arXiv.2005.14165 Also published in Advances in Neural Information Processing Systems 33 (2020).
4.1, 4.2	Nothing in how a language model writes separates true sentences from ones that only sound true, so it states wrong details as fluently as right ones.	Ji, Z., et al. (2023). Survey of Hallucination in Natural Language Generation. ACM Computing Surveys 55(12), 1–38. doi.org/10.1145/3571730
g4.1, g4.3, g4.6, g4.7	PubMed, the free search engine for the biomedical literature.	National Library of Medicine. PubMed. pubmed.ncbi.nlm.nih.gov
4.2	AI tools tend to go along with the way a question is framed, giving reasons for a premise whether or not it is true.	Sharma, M., et al. (2023). Towards Understanding Sycophancy in Language Models. arXiv 2310.13548. doi.org/10.48550/arXiv.2310.13548 Presented at ICLR 2024. The text describes the behaviour without naming it; the research term is sycophancy.
4.2	Elsevier's author policy warns that AI-generated references can be incorrect or fabricated, and asks authors to check AI output, including its sources.	Elsevier. (n.d.). Generative AI policies for journals. Accessed 26 October 2026.
4.2	The terms for TCGA data forbid any attempt to identify the people the data came from.	National Cancer Institute Genomic Data Commons. (n.d.). Data access policies. Accessed 26 October 2026. The condition comes from the NIH Genomic Data Sharing Policy, which the GDC applies to open-access TCGA data.
g4.2	The Cancer Genome Atlas: matched tumour, molecular and clinical data.	The Cancer Genome Atlas Research Network, et al. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nature Genetics 45(10), 1113-1120. doi.org/10.1038/ng.2764
g4.2	Retractions: why published papers get withdrawn.	Fang, F. C., Steen, R. G., & Casadevall, A. (2012). Misconduct accounts for the majority of retracted scientific publications. Proceedings of the National Academy of Sciences 109(42), 17028-17033. doi.org/10.1073/pnas.1212247109
4.2, 4.4, 4.7	Walters and Wilder had ChatGPT write 84 short reviews on 42 topics (636 references): 55% (GPT-3.5) and 18% (GPT-4) were invented; 43% and 24% of real ones had substantive errors.	Walters, W. H., & Wilder, E. I. (2023). Fabrication and errors in the bibliographic citations generated by ChatGPT. Scientific Reports 13(1), 14045. doi.org/10.1038/s41598-023-41032-5
4.2, 4.4	Elsevier: AI is never an author; use goes in a published AI declaration (basic grammar checks exempt), research use in Methods; no AI-made data images; reviewers never upload manuscripts.	Elsevier. (n.d.). Generative AI policies for journals. Accessed 26 October 2026.

SECTION	CLAIM IN THE TEXT	SOURCE
4.2, 4.4	Nature Portfolio: no AI authorship; disclose AI use; language polishing allowed, extensive writing help needs checking and disclosure; no photorealistic AI images; reviewers avoid public AI tools.	Nature Portfolio. (n.d.). Artificial Intelligence (AI) editorial policy. Accessed 26 October 2026.
4.3	Worked check: Warburg, O. (1956), On the origin of cancer cells, Science 123, 309–314, DOI 10.1126/science.123.3191.309.	Warburg, O. (1956). On the Origin of Cancer Cells. Science 123(3191), 309–314. doi.org/10.1126/science.123.3191.309
g4.3	The Warburg effect: tumours take up glucose and make lactate even with oxygen present.	Vander Heiden, M. G., Cantley, L. C., & Thompson, C. B. (2009). Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation. Science 324(5930), 1029-1033. doi.org/10.1126/science.1160809
4.3, 4.6	A real DOI typed after https://doi.org/ resolves to the paper; a made-up one fails or opens a different record.	DOI Foundation. (n.d.). DOI Handbook. doi.org/10.1000/182
4.3, 4.5	mRNA and protein levels for the same gene don't always agree, so reporting mRNA as protein changes what a result means.	Liu, Y., Beyer, A., & Aebersold, R. (2016). On the Dependency of Cellular Protein Levels on mRNA Abundance. Cell 165(3), 535–550. doi.org/10.1016/j.cell.2016.03.014
g4.3, g4.6, g4.7	Google Scholar for searching across disciplines.	Google. Google Scholar. scholar.google.com
g4.3, g4.4, g4.6, g4.7	AI tools cannot be listed as authors; their use must be disclosed.	International Committee of Medical Journal Editors. Recommendations: Defining the Role of Authors and Contributors (AI section). icmje.org/recommendations
4.4	ICMJE: chatbots can't be authors, as they can't be responsible for the work; AI use is disclosed in the cover letter and paper; nondisclosure may count as misconduct.	International Committee of Medical Journal Editors. (n.d.). Artificial Intelligence (AI) Use by Authors. Accessed 26 October 2026.
4.4	ICMJE: humans are responsible for AI-assisted material and must review it, as output can be incorrect, incomplete or biased; AI-generated material can't be cited as a primary source.	International Committee of Medical Journal Editors. (n.d.). Artificial Intelligence (AI) Use by Authors. Accessed 26 October 2026.
4.4	COPE: AI tools can't be authors because they can't take responsibility or declare conflicts of interest; use is described in the Methods; authors are fully responsible, even for AI-produced parts.	Committee on Publication Ethics. (2023). Authorship and AI tools: COPE position statement. Accessed 26 October 2026.
g4.4	Peer review: what it checks and what it misses.	Smith, R. (2006). Peer Review: A Flawed Process at the Heart of Science and Journals. Journal of the Royal Society of Medicine 99(4), 178-182. doi.org/10.1177/014107680609900414
g4.4, g4.7	The ICMJE criteria for authorship.	International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals. icmje.org/recommendations

SECTION	CLAIM IN THE TEXT	SOURCE
g4.4, g4.7	COPE guidance on publication ethics and corrections.	Committee on Publication Ethics (COPE). Core Practices. publicationethics.org/core-practices
4.6, 4.7	Reading-list item A is real: Possemato et al. (2011), Nature 476(7360), 346–350.	Possemato, R., et al. (2011). Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. Nature 476(7360), 346–350. doi.org/10.1038/nature10350
4.6, 4.7	Reading-list item C is real: Locasale et al. (2011), Nature Genetics 43(9), 869–874.	Locasale, J. W., et al. (2011). Phosphoglycerate dehydrogenase diverts glycolytic flux and contributes to oncogenesis. Nature Genetics 43(9), 869–874. doi.org/10.1038/ng.890

Guide 5 · Judging evidence كيف تحكم على الدليل

SECTION	CLAIM IN THE TEXT	SOURCE
5.1	By convention p below 0.05 is called significant; with no real difference about 1 test in 20 still falls under 0.05, so one significant result is weak evidence.	Wasserstein, R. L., & Lazar, N. A. (2016). The ASA statement on p-values: Context, process, and purpose. The American Statistician 70(2), 129–133. doi.org/10.1080/00031305.2016.1154108
5.1, 5.3	A 95% confidence interval gives the range of effects the data are reasonably compatible with; small studies give wide intervals.	Neyman, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability. Philosophical Transactions of the Royal Society A 236(767), 333–380. doi.org/10.1098/rsta.1937.0005
5.1, 5.3, 5.5, 5.13	The guide's PHGDH examples use TCGA breast tumours from the PanCancer Atlas (499 luminal A, 197 luminal B), drawn through cBioPortal.	Hoadley, K. A., et al. (2018). Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. Cell 173(2), 291–304.e6. doi.org/10.1016/j.cell.2018.03.022 The subtype counts and fold differences come from Al Basira's own extract (29 September 2026), not from the paper.
g5.1, g5.3, g5.5, g5.6, g5.7, g5.8 and more	The Cancer Genome Atlas: matched tumour, molecular and clinical data.	The Cancer Genome Atlas Research Network, et al. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nature Genetics 45(10), 1113-1120. doi.org/10.1038/ng.2764
5.2	If all tumours are processed at one time and normal tissue at another, a batch effect can look like biology.	Leek, J. T., et al. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. Nature Reviews Genetics 11(10), 733–739. doi.org/10.1038/nrg2825
5.2	Turner and colleagues: of 74 FDA-registered antidepressant trials, 51% were positive per the FDA but 94% looked positive in journals; 22 negative or questionable trials went unpublished and 11 were spun positive.	Turner, E. H., Matthews, A. M., Linardatos, E., Tell, R. A., & Rosenthal, R. (2008). Selective publication of antidepressant trials and its influence on apparent efficacy. New England Journal of Medicine 358(3), 252–260. doi.org/10.1056/NEJMsa065779

SECTION	CLAIM IN THE TEXT	SOURCE
5.2	The Reproducibility Project: Cancer Biology repeated 50 experiments from 23 high-impact papers; for positive effects the median replication effect was 85% smaller.	Errington, T. M., Mathur, M., Soderberg, C. K., Denis, A., Perfito, N., Iorns, E., & Nosek, B. A. (2021). Investigating the replicability of preclinical cancer biology. eLife 10, e71601. doi.org/10.7554/eLife.71601
g5.2, g5.13	Publication bias: positive results are more likely to be published.	Dickersin, K. (1990). The Existence of Publication Bias and Risk Factors for Its Occurrence. JAMA: The Journal of the American Medical Association 263(10), 1385. doi.org/10.1001/jama.1990.03440100097014
5.3	Chocolate consumption per person correlated with Nobel laureates per person across 23 countries (r = 0.79).	Messerli, F. H. (2012). Chocolate consumption, cognitive function, and Nobel laureates. New England Journal of Medicine 367(16), 1562–1564. doi.org/10.1056/NEJMon1211064
5.3	In 1965 Bradford Hill listed features that make a causal reading more believable, including strength, consistency, temporality and a dose-response gradient; none is proof alone.	Hill, A. B. (1965). The environment and disease: Association or causation? Proceedings of the Royal Society of Medicine 58(5), 295–300. doi.org/10.1177/003591576505800503 Hill listed nine viewpoints in all; the text names four of them.
g5.3	Confounding: a third factor that can explain an apparent link.	Greenland, S., Pearl, J., & Robins, J. M. (1999). Causal Diagrams for Epidemiologic Research. Epidemiology 10(1), 37-48. doi.org/10.1097/00001648-199901000-00008
g5.3, g5.6	Survival data in TCGA and its curated survival endpoints.	Liu, J., et al. (2018). An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. Cell 173(2), 400-416.e11. doi.org/10.1016/j.cell.2018.02.052
5.4	In 2015 IARC classified processed meat as carcinogenic to humans (bowel cancer); each 50 g eaten daily raises bowel cancer risk by 18%.	Bouvard, V., Loomis, D., Guyton, K. Z., Grosse, Y., El Ghissassi, F., Benbrahim-Tallaa, L., et al. (2015). Carcinogenicity of consumption of red and processed meat. The Lancet Oncology 16(16), 1599–1600. doi.org/10.1016/S1470-2045(15)00444-1 The guide links IARC's press release 240 (26 October 2015), which reports the same evaluation.
5.4	The number needed to treat is 100 divided by the absolute risk reduction in percentage points: halving a 4% risk means treating 50 for one fewer event.	Laupacis, A., Sackett, D. L., & Roberts, R. S. (1988). An assessment of clinically useful measures of the consequences of treatment. New England Journal of Medicine 318(26), 1728–1733. doi.org/10.1056/NEJM198806303182605
5.4	Hazard ratios, the most common ratio in cancer survival papers, are relative measures and don't give how many patients had the event.	Cox, D. R. (1972). Regression models and life-tables. Journal of the Royal Statistical Society: Series B 34(2), 187–202. doi.org/10.1111/j.2517-6161.1972.tb00899.x
5.5	Error bars can be SD, SEM or confidence intervals and look alike; Cumming and colleagues' first rule is that the legend must say which.	Cumming, G., Fidler, F., & Vaux, D. L. (2007). Error bars in experimental biology. Journal of Cell Biology 177(1), 7–11. doi.org/10.1083/jcb.200611141
5.5	Weissgerber and colleagues reviewed 703 papers in top physiology journals; most showed continuous data as bar and line graphs, which can hide very different distributions.	Weissgerber, T. L., Milic, N. M., Winham, S. J., & Garovic, V. D. (2015). Beyond bar and line graphs: Time for a new data presentation paradigm. PLOS Biology 13(4), e1002128. doi.org/10.1371/journal.pbio.1002128

SECTION	CLAIM IN THE TEXT	SOURCE
g5.5	The box plot as a summary of a distribution.	Mcgill, R., Tukey, J. W., & Larsen, W. A. (1978). Variations of Box Plots. The American Statistician 32(1), 12-16. doi.org/10.1080/00031305.1978.10479236
g5.6, g5.7, g5.8, g5.10, g5.11, g5.13	cBioPortal as a way to browse and download cancer genomics data.	Cerami, E., et al. (2012). The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. Cancer Discovery 2(5), 401-404. doi.org/10.1158/2159-8290.cd-12-0095
g5.6, g5.7, g5.8, g5.11, g5.13	The Gene Expression Omnibus, a public repository of expression data.	Edgar, R. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Research 30(1), 207-210. doi.org/10.1093/nar/30.1.207
g5.6, g5.7, g5.8, g5.10, g5.11, g5.13	DepMap: genome-wide dependency screens across cancer cell lines.	Tsherniak, A., et al. (2017). Defining a Cancer Dependency Map. Cell 170(3), 564-576.e16. doi.org/10.1016/j.cell.2017.06.010
g5.10	DESeq2 for testing differences in gene expression from raw counts.	Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15(12), 550. doi.org/10.1186/s13059-014-0550-8

Guide 6 · From papers to a story من الأوراق إلى قصة

SECTION	CLAIM IN THE TEXT	SOURCE
g6.2	Confounding: a third factor that can explain an apparent link.	Greenland, S., Pearl, J., & Robins, J. M. (1999). Causal Diagrams for Epidemiologic Research. Epidemiology 10(1), 37-48. doi.org/10.1097/00001648-199901000-00008
g6.2	The Cancer Genome Atlas: matched tumour, molecular and clinical data.	The Cancer Genome Atlas Research Network, et al. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nature Genetics 45(10), 1113-1120. doi.org/10.1038/ng.2764

Guide 7 · Writing it up honestly كيف تكتب بأمانة

SECTION	CLAIM IN THE TEXT	SOURCE
g7.2	Zotero, the free reference manager.	Corporation for Digital Scholarship. Zotero. zotero.org
g7.2	Google Scholar for searching across disciplines.	Google. Google Scholar. scholar.google.com
g7.2	AI chatbots can produce references that do not exist.	Walters, W. H., & Wilder, E. I. (2023). Fabrication and errors in the bibliographic citations generated by ChatGPT. Scientific Reports 13(1), 14045. doi.org/10.1038/s41598-023-41032-5

SECTION	CLAIM IN THE TEXT	SOURCE
g7.2, g7.3, g7.4	Retractions: why published papers get withdrawn.	Fang, F. C., Steen, R. G., & Casadevall, A. (2012). Misconduct accounts for the majority of retracted scientific publications. Proceedings of the National Academy of Sciences 109(42), 17028-17033. doi.org/10.1073/pnas.1212247109
g7.2, g7.5	PubMed, the free search engine for the biomedical literature.	National Library of Medicine. PubMed. pubmed.ncbi.nlm.nih.gov
g7.2, g7.4	The ICMJE criteria for authorship.	International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals. icmje.org/recommendations
g7.3, g7.4, g7.5	Fabrication, falsification and plagiarism as research misconduct.	US Office of Research Integrity. Definition of Research Misconduct, 42 CFR Part 93. ori.hhs.gov/definition-research-misconduct
g7.3, g7.4	A hypothesis earns credibility by surviving serious attempts to refute it.	Popper, K. (2002). The Logic of Scientific Discovery. Routledge (first English edition 1959). ISBN 9780415278447
g7.4, g7.5	AI tools cannot be listed as authors; their use must be disclosed.	International Committee of Medical Journal Editors. Recommendations: Defining the Role of Authors and Contributors (AI section). icmje.org/recommendations

Guide 8 · Review, and what comes next المراجعة، وإيش بعدها

SECTION	CLAIM IN THE TEXT	SOURCE
g8.1	PubMed, the free search engine for the biomedical literature.	National Library of Medicine. PubMed. pubmed.ncbi.nlm.nih.gov
g8.1	Google Scholar for searching across disciplines.	Google. Google Scholar. scholar.google.com

Principles sessions

6 CLAIMS

SECTION	CLAIM IN THE TEXT	SOURCE
Session plan	Reading a paper in three passes.	Keshav, S. (2007). How to read a paper. ACM SIGCOMM Computer Communication Review 37(3), 83-84. doi.org/10.1145/1273445.1273458
Session plan	Zotero, the free reference manager.	Corporation for Digital Scholarship. Zotero. zotero.org
Session plan	PubMed, the free search engine for the biomedical literature.	National Library of Medicine. PubMed. pubmed.ncbi.nlm.nih.gov
Session plan	Google Scholar for searching across disciplines.	Google. Google Scholar. scholar.google.com

SECTION	CLAIM IN THE TEXT	SOURCE
Session plan	AI chatbots can produce references that do not exist.	Walters, W. H., & Wilder, E. I. (2023). Fabrication and errors in the bibliographic citations generated by ChatGPT. Scientific Reports 13(1), 14045. doi.org/10.1038/s41598-023-41032-5
Session plan	Fabrication, falsification and plagiarism as research misconduct.	US Office of Research Integrity. Definition of Research Misconduct, 42 CFR Part 93. ori.hhs.gov/definition-research-misconduct

Principles assignments

12 CLAIMS

SECTION	CLAIM IN THE TEXT	SOURCE
Assignments	The Warburg effect: tumours take up glucose and make lactate even with oxygen present.	Vander Heiden, M. G., Cantley, L. C., & Thompson, C. B. (2009). Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation. Science 324(5930), 1029-1033. doi.org/10.1126/science.1160809
Assignments	PubMed, the free search engine for the biomedical literature.	National Library of Medicine. PubMed. pubmed.ncbi.nlm.nih.gov
Assignments	Google Scholar for searching across disciplines.	Google. Google Scholar. scholar.google.com
Assignments	Reading a paper in three passes.	Keshav, S. (2007). How to read a paper. ACM SIGCOMM Computer Communication Review 37(3), 83-84. doi.org/10.1145/1273445.1273458
Assignments	Retractions: why published papers get withdrawn.	Fang, F. C., Steen, R. G., & Casadevall, A. (2012). Misconduct accounts for the majority of retracted scientific publications. Proceedings of the National Academy of Sciences 109(42), 17028-17033. doi.org/10.1073/pnas.1212247109
Assignments	Retraction Watch and its database of retracted papers.	Retraction Watch Database. The Center for Scientific Integrity. retractiondatabase.org
Assignments	Peer review: what it checks and what it misses.	Smith, R. (2006). Peer Review: A Flawed Process at the Heart of Science and Journals. Journal of the Royal Society of Medicine 99(4), 178-182. doi.org/10.1177/014107680609900414
Assignments	MeSH, the controlled vocabulary used to index PubMed.	National Library of Medicine. Medical Subject Headings (MeSH). nlm.nih.gov/mesh
Assignments	AI tools cannot be listed as authors; their use must be disclosed.	International Committee of Medical Journal Editors. Recommendations: Defining the Role of Authors and Contributors (AI section). icmje.org/recommendations
Assignments	The Cancer Genome Atlas: matched tumour, molecular and clinical data.	The Cancer Genome Atlas Research Network, et al. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nature Genetics 45(10), 1113-1120. doi.org/10.1038/ng.2764
Assignments	Fabrication, falsification and plagiarism as research misconduct.	US Office of Research Integrity. Definition of Research Misconduct, 42 CFR Part 93. ori.hhs.gov/definition-research-misconduct

SECTION	CLAIM IN THE TEXT	SOURCE
Assignments	The ICMJE criteria for authorship.	International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals. icmje.org/recommendations

Session teaching notes

29 CLAIMS

SECTION	CLAIM IN THE TEXT	SOURCE
Session 1	Immortal time bias in survival comparisons.	Suissa, S. (2008). Immortal Time Bias in Pharmacoepidemiology. American Journal of Epidemiology 167(4), 492-499. doi.org/10.1093/aje/kwm324
Session 1	In the 1920s Warburg found tumour tissue turns far more glucose into lactate than normal tissue, even with plenty of oxygen.	Warburg, O., Wind, F., & Negelein, E. (1927). The metabolism of tumors in the body. Journal of General Physiology 8(6), 519–530. doi.org/10.1085/jgp.8.6.519
Session 1	The Warburg observation has held up, but his explanation (damaged respiration) has not: many cancers show the effect with normally working mitochondria.	Koppenol, W. H., Bounds, P. L., & Dang, C. V. (2011). Otto Warburg's contributions to current concepts of cancer metabolism. Nature Reviews Cancer 11(5), 325–337. doi.org/10.1038/nrc3038
Session 1	Suissa and Azoulay found 13 observational metformin–cancer studies with immortal time bias; three studies that avoided the time-related biases found no effect.	Suissa, S., & Azoulay, L. (2012). Metformin and the Risk of Cancer: Time-related biases in observational studies. Diabetes Care 35(12), 2665–2673. doi.org/10.2337/dc12-0788
Session 1	The MA.32 randomised trial found metformin made no detectable difference to outcomes in breast cancer.	Goodwin, P. J., et al. (2022). Effect of Metformin vs Placebo on Invasive Disease–Free Survival in Patients With Breast Cancer: The MA.32 Randomized Clinical Trial. JAMA 327(20), 1963–1973. doi.org/10.1001/jama.2022.6147 3,649 patients without diabetes; no improvement in invasive disease-free survival.
Session 1, Session 2, Session 3, Session 4, Session 8	PubMed, the free search engine for the biomedical literature.	National Library of Medicine. PubMed. pubmed.ncbi.nlm.nih.gov
Session 1, Session 2, Session 3, Session 4, Session 8	Google Scholar for searching across disciplines.	Google. Google Scholar. scholar.google.com

SECTION	CLAIM IN THE TEXT	SOURCE
Session 1, Session 4	The Warburg effect: tumours take up glucose and make lactate even with oxygen present.	Vander Heiden, M. G., Cantley, L. C., & Thompson, C. B. (2009). Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation. Science 324(5930), 1029-1033. doi.org/10.1126/science.1160809
Session 1, Session 4	Warburg's 1956 Science paper 'On the origin of cancer cells' (volume 123, pages 309–314) argued cancer starts with damage to respiration; used as the reference-checking example.	Warburg, O. (1956). On the Origin of Cancer Cells. Science 123(3191), 309–314. doi.org/10.1126/science.123.3191.309
Session 1, Session 5	Confounding: a third factor that can explain an apparent link.	Greenland, S., Pearl, J., & Robins, J. M. (1999). Causal Diagrams for Epidemiologic Research. Epidemiology 10(1), 37-48. doi.org/10.1097/00001648-199901000-00008
Session 2	Keshav's three-pass method: a five-to-ten-minute first pass, a second pass on the figures, and a third pass of hours for the papers that matter most.	Keshav, S. (2007). How to read a paper. ACM SIGCOMM Computer Communication Review 37(3), 83–84. doi.org/10.1145/1273445.1273458
Session 2	In 42 of 72 trials with a non-significant main result, the abstract still leaned favourable (spin).	Boutron, I., Dutton, S., Ravaud, P., & Altman, D. G. (2010). Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. JAMA 303(20), 2058–2064. doi.org/10.1001/jama.2010.651 The 42 (58.3%) refers to spin in the abstract's Conclusions section.
Session 2	About one citation in four doesn't properly support the sentence it is attached to.	Jergas, H., & Baethge, C. (2015). Quotation accuracy in medical journal articles: a systematic review and meta-analysis. PeerJ 3, e1364. doi.org/10.7717/peerj.1364 Total quotation error rate 25.4%; major errors 11.9%.
Session 2	Fact-checkers who left a page to see what others said about it (lateral reading) judged sources better and faster than careful readers taken in by a professional look.	Wineburg, S., & McGrew, S. (2019). Lateral Reading and the Nature of Expertise: Reading Less and Learning More When Evaluating Digital Information. Teachers College Record 121(11), 1–40. doi.org/10.1177/016146811912101102
Session 2, Session 3	Retractions: why published papers get withdrawn.	Fang, F. C., Steen, R. G., & Casadevall, A. (2012). Misconduct accounts for the majority of retracted scientific publications. Proceedings of the National Academy of Sciences 109(42), 17028-17033. doi.org/10.1073/pnas.1212247109
Session 3	Predatory journals and how to spot them.	Grudniewicz, A., et al. (2019). Predatory journals: no definition, no defence. Nature 576(7786), 210-212. doi.org/10.1038/d41586-019-03759-y
Session 3	MeSH, the controlled vocabulary used to index PubMed.	National Library of Medicine. Medical Subject Headings (MeSH). nlm.nih.gov/mesh
Session 3	Retraction Watch and its database of retracted papers.	Retraction Watch Database. The Center for Scientific Integrity. retractiondatabase.org
Session 3	BMJ peer reviewers found on average 2.58 of nine deliberately inserted major errors.	Schroter, S., Black, N., Evans, S., Godlee, F., Osorio, L., & Smith, R. (2008). What errors do peer reviewers detect, and does training improve their ability to detect them? Journal of the Royal Society of Medicine 101(10), 507–514. doi.org/10.1258/jrsm.2008.080062

SECTION	CLAIM IN THE TEXT	SOURCE
Session 3	Retracted papers keep being cited: of 13,252 citations after retraction, only 5.4% mentioned the retraction.	Hsiao, T.-K., & Schneider, J. (2021). Continued use of retracted papers: Temporal trends in citations and (lack of) awareness of retractions shown in citation contexts in biomedicine. Quantitative Science Studies 2(4), 1144–1169. doi.org/10.1162/qss_a_00155 The 13,252 are post-retraction citation contexts in PubMed Central open-access papers; 722 acknowledged the retraction.
Session 3, Session 7	Fabrication, falsification and plagiarism as research misconduct.	US Office of Research Integrity. Definition of Research Misconduct, 42 CFR Part 93. ori.hhs.gov/definition-research-misconduct
Session 3, Session 8	Peer review: what it checks and what it misses.	Smith, R. (2006). Peer Review: A Flawed Process at the Heart of Science and Journals. Journal of the Royal Society of Medicine 99(4), 178-182. doi.org/10.1177/014107680609900414
Session 4	AI tools cannot be listed as authors; their use must be disclosed.	International Committee of Medical Journal Editors. Recommendations: Defining the Role of Authors and Contributors (AI section). icmje.org/recommendations
Session 4	COPE guidance on publication ethics and corrections.	Committee on Publication Ethics (COPE). Core Practices. publicationethics.org/core-practices
Session 4	AI chatbots can produce references that do not exist.	Walters, W. H., & Wilder, E. I. (2023). Fabrication and errors in the bibliographic citations generated by ChatGPT. Scientific Reports 13(1), 14045. doi.org/10.1038/s41598-023-41032-5
Session 4, Session 7	The ICMJE criteria for authorship.	International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals. icmje.org/recommendations
Session 5	Publication bias: positive results are more likely to be published.	Dickersin, K. (1990). The Existence of Publication Bias and Risk Factors for Its Occurrence. JAMA: The Journal of the American Medical Association 263(10), 1385. doi.org/10.1001/jama.1990.03440100097014
Session 6	The Cancer Genome Atlas: matched tumour, molecular and clinical data.	The Cancer Genome Atlas Research Network, et al. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nature Genetics 45(10), 1113-1120. doi.org/10.1038/ng.2764
Session 7	A hypothesis earns credibility by surviving serious attempts to refute it.	Popper, K. (2002). The Logic of Scientific Discovery. Routledge (first English edition 1959). ISBN 9780415278447

PROGRAMME 4

Workshops الورش

The scripts, slides, handouts and presenter notes for our workshops, free and paid. 133 claims, 91 sources.

Workshop 1 · Is research for me? هل البحث ينا سبني؟

SECTION	CLAIM IN THE TEXT	SOURCE
Handout, Notes, Slides	cBioPortal as a way to browse and download cancer genomics data.	Cerami, E., et al. (2012). The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. Cancer Discovery 2(5), 401-404. doi.org/10.1158/2159-8290.cd-12-0095
Notes, Slides	Most landmark preclinical cancer findings could not be confirmed.	Begley, C. G., & Ellis, L. M. (2012). Raise standards for preclinical cancer research. Nature 483(7391), 531-533. doi.org/10.1038/483531a
Notes, Slides	The Cancer Genome Atlas: matched tumour, molecular and clinical data.	The Cancer Genome Atlas Research Network, et al. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nature Genetics 45(10), 1113-1120. doi.org/10.1038/ng.2764
Notes, Slides	The Warburg effect: tumours take up glucose and make lactate even with oxygen present.	Vander Heiden, M. G., Cantley, L. C., & Thompson, C. B. (2009). Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation. Science 324(5930), 1029-1033. doi.org/10.1126/science.1160809

Workshop 2 · Your first look at cancer data أول نظرة على بيانات السرطان

SECTION	CLAIM IN THE TEXT	SOURCE
Handout, Notes, Script, Slides	The Cancer Genome Atlas: matched tumour, molecular and clinical data.	The Cancer Genome Atlas Research Network, et al. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nature Genetics 45(10), 1113-1120. doi.org/10.1038/ng.2764
Handout, Notes, Script, Slides	cBioPortal as a way to browse and download cancer genomics data.	Cerami, E., et al. (2012). The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. Cancer Discovery 2(5), 401-404. doi.org/10.1158/2159-8290.cd-12-0095
Handout	The Gene Expression Omnibus, a public repository of expression data.	Edgar, R. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Research 30(1), 207-210. doi.org/10.1093/nar/30.1.207
Handout	DepMap: genome-wide dependency screens across cancer cell lines.	Tsherniak, A., et al. (2017). Defining a Cancer Dependency Map. Cell 170(3), 564-576.e16. doi.org/10.1016/j.cell.2017.06.010

Workshop 3 · Make a scientific poster تصميم بوستر علمي

SECTION	CLAIM IN THE TEXT	SOURCE
Slides 6–8, 12, 37–38, 40, 55 (and notes); Handout p1, p3	Erren and Bourne: posters are viewed from a distance but use your presence; Rule 2 is to sell your work in ten seconds; never use text below 24 pt.	Erren, T. C., & Bourne, P. E. (2007). Ten simple rules for a good poster presentation. PLoS Computational Biology 3(5), e102. doi.org/10.1371/journal.pcbi.0030102 The 24 pt floor is in Rule 7; the quoted sentence is in the introduction.
Slides 8–9, 12, 37, 39–40 (and notes); Handout p1, p3	ICSE 2024 asks for posters between A1 and A0, portrait, easily readable at two metres, and recommends 24 to 32 pt text.	ICSE 2024, 46th International Conference on Software Engineering. (2024). Guidelines for creating a poster. conf.researchr.org ICSE states 24–32 pt as a recommendation, not a minimum.
Slides 10–11 (and notes); Handout p3	A0 is 841 × 1189 mm, one square metre; each A size halves the one before, so sixteen A4 sheets make one A0.	International Organization for Standardization. (2007). ISO 216:2007 Writing paper and certain classes of printed matter — Trimmed sizes — A and B series, and indication of machine direction. ISO.
Slides 12, 29 (and notes); Handout p3	EAHP Congress 2026 posters must be A0 portrait, with title letters at least 2.5 cm and text at least 8 mm high, read left to right, top to bottom.	European Association of Hospital Pharmacists. (2026). Poster guidelines, EAHP Congress 2026. eahp.eu
Slides 15–16, 19, 32, 41, 44 (and notes); Handout p2	The worked example uses 981 primary breast tumours with a PAM50 subtype from the TCGA PanCancer Atlas.	Hoadley, K. A., et al. (2018). Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. Cell 173(2), 291–304.e6. doi.org/10.1016/j.cell.2018.03.022
Slide 16 (and notes)	A real TCGA breast result: basal-like tumours have the highest hypoxia scores.	Bhandari, V., et al. (2019). Molecular landmarks of tumor hypoxia across cancer types. Nature Genetics 51(2), 308–318. doi.org/10.1038/s41588-018-0318-2 True for the Buffa and Winter hypoxia scores in the PanCancer Atlas data; on the Ragnum score HER2-enriched tumours are slightly higher.
Slides 19, 32, 41, 44 (and notes); Handout p2	Accessed via cBioPortal: median PHGDH mRNA 3,899 in basal-like (n 171) against 523 in luminal A (n 499), a 7.5-fold difference; log2 medians 11.9 and 9.0.	Cerami, E., et al. (2012). The cBio Cancer Genomics Portal: An open platform for exploring multidimensional cancer genomics data. Cancer Discovery 2(5), 401–404. doi.org/10.1158/2159-8290.CD-12-0095; Gao, J., et al. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Science Signaling 6(269), pl1. doi.org/10.1126/scisignal.2004088 The cBioPortal TCGA PanCancer Atlas breast data give a ratio of 7.45 and subtype n of 171, 78, 36, 197 and 499 (total 981), matching the slides.

SECTION	CLAIM IN THE TEXT	SOURCE
Slide 29 (and notes); Handout p2	Mike Morrison, a psychology PhD student at Michigan State, popularised the #betterposter in 2019: the main finding in plain English in the middle, plus a QR code.	Young, J. R. (2019, June 21). Research posters are a staple of academic conferences. Could a new design speed discovery? EdSurge.
Slides 30–31, 39 (and notes); Handout p2	Purrington's poster guide: aim for 1,000 words or fewer, about 200 per main section; graph details should be readable from 6 feet.	Purrington, C. (n.d.). Designing conference posters. Colin Purrington's blog. The 1,000 words include about 300 for all figure legends; the 200-word Results section excludes legends.
Slide 32; Handout p2	PHGDH catalyses the first step of serine synthesis, and its protein is raised in about 70% of ER-negative breast cancers.	Possemato, R., et al. (2011). Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. Nature 476(7360), 346–350. doi.org/10.1038/nature10350
Slides 41, 44	PHGDH mRNA is measured as RSEM values, plotted as log2(RSEM + 1).	Li, B., & Dewey, C. N. (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12, 323. doi.org/10.1186/1471-2105-12-323
Slides 42–43 (and notes); Handout p3	About 8% of men of European descent and about 5% of Asian men are red-green colour blind; Okabe and Ito's eight-colour palette (#000000 … #CC79A7) is safe.	Okabe, M., & Ito, K. (2002, revised 2008). Color Universal Design (CUD): How to make figures and presentations that are friendly to colorblind people. J*Fly. The page gives the palette as RGB 0–255 values; the hex codes on the slide convert them exactly.
Slide 43	Wong's Nature Methods column recommends the same colour-blind-safe palette.	Wong, B. (2011). Points of view: Color blindness. Nature Methods 8(6), 441. doi.org/10.1038/nmeth.1618
Slide 44; Handout p2	Tumours are grouped by PAM50 intrinsic subtype (basal-like, HER2-enriched, luminal A, luminal B, normal-like).	Parker, J. S., et al. (2009). Supervised risk predictor of breast cancer based on intrinsic subtypes. Journal of Clinical Oncology 27(8), 1160–1167. doi.org/10.1200/JCO.2008.18.1370
Slide 44; Handout p2	Basal-like and luminal A tumours are compared with a Mann–Whitney test.	Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. The Annals of Mathematical Statistics 18(1), 50–60. doi.org/10.1214/aoms/1177730491
Handout, Notes, Slides	The Cancer Genome Atlas: matched tumour, molecular and clinical data.	The Cancer Genome Atlas Research Network, et al. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nature Genetics 45(10), 1113-1120. doi.org/10.1038/ng.2764
Slides	The box plot as a summary of a distribution.	Mcgill, R., Tukey, J. W., & Larsen, W. A. (1978). Variations of Box Plots. The American Statistician 32(1), 12-16. doi.org/10.1080/00031305.1978.10479236

Workshop 4 · Write a literature review كتابة مراجعة أدبيات

SECTION	CLAIM IN THE TEXT	SOURCE
Handout, Notes, Slides	Retraction Watch and its database of retracted papers.	Retraction Watch Database. The Center for Scientific Integrity. retractiondatabase.org
Handout, Notes, Script, Slides	PubMed, the free search engine for the biomedical literature.	National Library of Medicine. PubMed. pubmed.ncbi.nlm.nih.gov
Handout, Notes, Slides	MeSH, the controlled vocabulary used to index PubMed.	National Library of Medicine. Medical Subject Headings (MeSH). nlm.nih.gov/mesh
Notes, Slides	The Cancer Genome Atlas: matched tumour, molecular and clinical data.	The Cancer Genome Atlas Research Network, et al. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nature Genetics 45(10), 1113-1120. doi.org/10.1038/ng.2764
Notes, Slides	Retractions: why published papers get withdrawn.	Fang, F. C., Steen, R. G., & Casadevall, A. (2012). Misconduct accounts for the majority of retracted scientific publications. Proceedings of the National Academy of Sciences 109(42), 17028-17033. doi.org/10.1073/pnas.1212247109
Notes, Script, Slides	PRISMA 2020 for reporting systematic reviews.	Page, M. J., et al. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ, n71. doi.org/10.1136/bmj.n71
Notes	Publication bias: positive results are more likely to be published.	Dickersin, K. (1990). The Existence of Publication Bias and Risk Factors for Its Occurrence. JAMA: The Journal of the American Medical Association 263(10), 1385. doi.org/10.1001/jama.1990.03440100097014
Notes	AI chatbots can produce references that do not exist.	Walters, W. H., & Wilder, E. I. (2023). Fabrication and errors in the bibliographic citations generated by ChatGPT. Scientific Reports 13(1), 14045. doi.org/10.1038/s41598-023-41032-5
Notes, Script, Slides	The ICMJE criteria for authorship.	International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals. icmje.org/recommendations
Notes, Slides	COPE guidance on publication ethics and corrections.	Committee on Publication Ethics (COPE). Core Practices. publicationethics.org/core-practices
Notes	Fabrication, falsification and plagiarism as research misconduct.	US Office of Research Integrity. Definition of Research Misconduct, 42 CFR Part 93. ori.hhs.gov/definition-research-misconduct
Slides	cBioPortal as a way to browse and download cancer genomics data.	Cerami, E., et al. (2012). The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. Cancer Discovery 2(5), 401-404. doi.org/10.1158/2159-8290.cd-12-0095

Workshop 5 · Cancer metabolism in 90 minutes أيض السرطان في تسعين دقيقة

SECTION	CLAIM IN THE TEXT	SOURCE
Handout, Notes, Script, Slides	The Cancer Genome Atlas: matched tumour, molecular and clinical data.	The Cancer Genome Atlas Research Network, et al. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nature Genetics 45(10), 1113-1120. doi.org/10.1038/ng.2764
Handout, Notes, Script, Slides	DepMap: genome-wide dependency screens across cancer cell lines.	Tsherniak, A., et al. (2017). Defining a Cancer Dependency Map. Cell 170(3), 564-576.e16. doi.org/10.1016/j.cell.2017.06.010
Handout, Notes, Script, Slides	The Warburg effect: tumours take up glucose and make lactate even with oxygen present.	Vander Heiden, M. G., Cantley, L. C., & Thompson, C. B. (2009). Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation. Science 324(5930), 1029-1033. doi.org/10.1126/science.1160809
Handout	PubMed, the free search engine for the biomedical literature.	National Library of Medicine. PubMed. pubmed.ncbi.nlm.nih.gov
Script, Slides	The hallmarks of cancer.	Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. Cancer Discovery 12(1), 31-46. doi.org/10.1158/2159-8290.cd-21-1059

Workshop 6 · Figures that tell the truth رسوم بيانية صادقة

SECTION	CLAIM IN THE TEXT	SOURCE
Handout p1	Cividis is a colour map recommended in place of a rainbow scale.	Nuñez, J. R., Anderton, C. R., & Renslow, R. S. (2018). Optimizing colormaps with consideration for color vision deficiency to enable accurate interpretation of scientific data. PLOS ONE 13(7), e0199239. doi.org/10.1371/journal.pone.0199239
Slides 3, 7–8 (and notes)	Anscombe built four 11-point datasets with the same means, spreads and correlation but very different scatter plots; in set IV one point at x = 19 creates r = 0.82.	Anscombe, F. J. (1973). Graphs in statistical analysis. The American Statistician 27(1), 17–21. doi.org/10.1080/00031305.1973.10478966 Recomputed from Anscombe's published values: r = 0.816–0.817 for all four sets; set IV has ten x values of 8 and one of 19 (y = 12.50).
Slides 9, 28–29, 34–35; Handout p4	Of 703 top physiology papers, 85.6% had a bar graph (mostly mean ± SE), 13.4% a dot plot, 8.0% a histogram, 5.3% a box plot; authors supply Excel templates.	Weissgerber, T. L., Milic, N. M., Winham, S. J., & Garovic, V. D. (2015). Beyond bar and line graphs: Time for a new data presentation paradigm. PLOS Biology 13(4), e1002128. doi.org/10.1371/journal.pbio.1002128 77.6% of the bar-graph papers showed mean ± SE; the notes round box plots to 'about 5' (paper: 5.3%).
Slides 10, 49–50; Handout p1	People judge positions along a common scale more accurately than lengths, angles or areas, which is why pies and 3D charts are hard to read.	Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. Journal of the American Statistical Association 79(387), 531–554. doi.org/10.1080/01621459.1984.10478080

SECTION	CLAIM IN THE TEXT	SOURCE
Slides 17–18	Online experiments found a truncated y-axis still makes differences feel bigger, even when the chart marks the cut; the quoted sentence is from the paper.	Correll, M., Bertini, E., & Franconeri, S. (2020). Truncating the y-axis: Threat or menace? Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 1–12. doi.org/10.1145/3313831.3376222 Quote checked word for word against the paper's abstract (arXiv 1907.02035).
Slides 30–32; Handout p4	Eight error-bar rules: Rule 1, say in the legend what they are; Rule 2, state n; Rule 3, no error bars on replicates of one experiment.	Cumming, G., Fidler, F., & Vaux, D. L. (2007). Error bars in experimental biology. Journal of Cell Biology 177(1), 7–11. doi.org/10.1083/jcb.200611141 Rule 1 quotation and Rules 2–3 checked against the full text (PMC2064100); the paper has eight rules.
Slide 33	473 authors in psychology, behavioural neuroscience and medicine adjusted two means until they looked just significant; many misjudged error bars and confused SE bars with confidence intervals.	Belia, S., Fidler, F., Williams, J., & Cumming, G. (2005). Researchers misunderstand confidence intervals and standard error bars. Psychological Methods 10(4), 389–396. doi.org/10.1037/1082-989X.10.4.389
Slide 34	Drummond and Vowler's article title sums up the advice: 'show the data, don't conceal them'.	Drummond, G. B., & Vowler, S. L. (2011). Show the data, don't conceal them. Journal of Physiology 589(8), 1861–1863. doi.org/10.1113/jphysiol.2011.205062
Slide 36 (notes); Script	The tumours are grouped by PAM50 intrinsic subtype: basal-like, HER2-enriched, luminal A, luminal B and normal-like.	Parker, J. S., et al. (2009). Supervised risk predictor of breast cancer based on intrinsic subtypes. Journal of Clinical Oncology 27(8), 1160–1167. doi.org/10.1200/JCO.2008.18.1370
Slide 36; Handout p2	PHGDH values are RSEM expression estimates, plotted as log2(RSEM + 1).	Li, B., & Dewey, C. N. (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12, 323. doi.org/10.1186/1471-2105-12-323
Slide 36 (notes); Handout p2	PHGDH is the first enzyme in the pathway cells use to make serine.	Possemato, R., et al. (2011). Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. Nature 476(7360), 346–350. doi.org/10.1038/nature10350 The abstract says 'PHGDH catalyses the first step in the serine biosynthesis pathway'.
Slides 36, 53–54; Handout pp2, 4; Script	The practice file is real TCGA PanCancer Atlas breast tumour data (75 tumours, 15 per subtype); the full TCGA breast cohort has 1,082 tumours.	Hoadley, K. A., et al. (2018). Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. Cell 173(2), 291–304.e6. doi.org/10.1016/j.cell.2018.03.022 cBioPortal's Breast Invasive Carcinoma (TCGA, PanCancer Atlas) study has 1,084 samples, 1,082 of them with RNA-seq. Means and medians on the slides match ws06-practice-data.csv.
Slides 36, 54; Handout p4	The practice data were downloaded through cBioPortal.	Cerami, E., et al. (2012). The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. Cancer Discovery 2(5), 401–404. doi.org/10.1158/2159-8290.CD-12-0095 cBioPortal asks users to cite Gao, J., et al. (2013), Science Signaling 6(269), pl1, doi.org/10.1126/scisignal.2004088, as well.

SECTION	CLAIM IN THE TEXT	SOURCE
Slide 45	Okabe and Ito, both colour blind, estimate red-green colour blindness at 1 in 12 men of European descent, 1 in 20 Asian men and 1 in 25 African men.	Okabe, M., & Ito, K. (2008). Color Universal Design (CUD): How to make figures and presentations that are friendly to colorblind people. J*FLY Data Depository for Drosophila Researchers. The page says 'Caucasian' (8%), 'Asian' (5%) and 'African' (4%) males, and that both authors are strong protanopes. The page carries no single dated version; 2008 is its commonly cited revision.
Slide 46	Nature's figure guide points authors to Bang Wong's 2011 Nature Methods column on colour blindness, which uses the same palette.	Wong, B. (2011). Points of view: Color blindness. Nature Methods 8(6), 441. doi.org/10.1038/nmeth.1618
Slide 46; Handout p3	Okabe and Ito's eight-colour palette: black #000000, orange #E69F00, sky blue #56B4E9, bluish green #009E73, yellow #F0E442, blue #0072B2, vermillion #D55E00, reddish purple #CC79A7.	Okabe, M., & Ito, K. (2008). Color Universal Design (CUD): How to make figures and presentations that are friendly to colorblind people. J*FLY Data Depository for Drosophila Researchers. The hex codes match R's grDevices definition of the Okabe-Ito palette and the colour table in Nature's figure guide; on the Okabe–Ito page itself they appear only in an image.
Slide 46; Handout p3	In R 4.0 or later, palette.colors(palette = "Okabe-Ito") returns the Okabe–Ito colours.	R Core Team. (2020). R 4.0.0 NEWS: palette.colors() and new built-in palettes. R Foundation for Statistical Computing. It returns nine colours: the eight on the slide plus grey (#999999).
Slides 46, 51, 61; Handout p4	Nature asks for sans-serif text of 5 pt or more, axis lines and ticks, units in brackets, an accessible palette; no gridlines, shadows, patterns or red-green pairs.	Nature. (n.d.). Preparing figures: our specifications. Nature Research Figure Guide.
Slide 47	Rainbow (jet) colour maps create false boundaries in smoothly changing data.	Borland, D., & Taylor, R. M., II. (2007). Rainbow color map (still) considered harmful. IEEE Computer Graphics and Applications 27(2), 14–17. doi.org/10.1109/MCG.2007.323435
Slide 47; Handout pp1, 3	Viridis is a perceptually even colour map that also reads in greyscale; ggplot2 offers it as scale_fill_viridis_c().	Garnier, S., Ross, N., Rudis, R., Camargo, A. P., Sciaini, M., & Scherer, C. (2024). viridis(Lite): Colorblind-friendly color maps for R (R package version 0.6.5). doi.org/10.5281/zenodo.4679423 Citation taken from the package's CRAN citation page; scale_fill_viridis_c() is in ggplot2's reference.
Slide 47; Handout p4	Rainbow colour maps distort data and mislead readers, including those with colour-vision deficiency; Crameri and colleagues' 2020 guide sets this out.	Crameri, F., Shephard, G. E., & Heron, P. J. (2020). The misuse of colour in science communication. Nature Communications 11(1), 5444. doi.org/10.1038/s41467-020-19160-7
Slides 47, 60	Nature prints figures 89 mm wide for one column and 183 mm for two, with a 5 pt minimum text size, and advises avoiding rainbow scales.	Nature. (n.d.). Building and exporting figure panels. Nature Research Figure Guide.
Slide 56; Handout p3	The R recipe draws the dot plot with ggplot2 (geom_jitter plus a median crossbar).	Wickham, H. (2016). ggplot2: Elegant graphics for data analysis (2nd ed.). Springer. doi.org/10.1007/978-3-319-24277-4

SECTION	CLAIM IN THE TEXT	SOURCE
Slide 61	Nature's reporting checklist asks for the exact n per group as a number, whether samples were distinct or measured repeatedly, and the variation shown (e.g. SD or CI).	Nature Portfolio. (n.d.). Reporting summary.

Workshop 7 · Read a paper in an hour اقرأ ورقة علمية في ساعة

SECTION	CLAIM IN THE TEXT	SOURCE
Handout p2; Slide 31	The Retraction Watch Database can be searched by DOI or title and has been free to use since 2023.	Hendricks, G., Lammey, R., Ofiesh, L., Bilder, G., & Pentz, E. (2023, September 12). News: Crossref and Retraction Watch. Crossref Blog.
Slides 4, 25; Handout p2; Script	Tonight's paper: Domcke, Sinha, Levine, Sander and Schultz, Nature Communications 4, 2126, 9 July 2013; Memorial Sloan Kettering; open access under CC BY 3.0; four figures.	Domcke, S., Sinha, R., Levine, D. A., Sander, C., & Schultz, N. (2013). Evaluating cell lines as tumour models by comparison of genomic profiles. Nature Communications 4, 2126. doi.org/10.1038/ncomms3126 Licence per Crossref and the article: CC BY 3.0. The first author was also at TU München.
Slides 7–8, 14; Handout pp1–2	Keshav's three passes: 5–10 minutes first, up to an hour second, a deep third; after the first pass you should answer the five Cs.	Keshav, S. (2007). How to read a paper. ACM SIGCOMM Computer Communication Review 37(3), 83–84. doi.org/10.1145/1273445.1273458 Keshav's first pass reads the whole introduction; the handout's 'last paragraph of the introduction and of the discussion' is Al Basira's adaptation.
Slide 12	SK-OV-3 and A2780 had about 60% of publications but match poorly, lacking the TP53 mutation in 95% of tumours; the twelve best matches had about 1%.	Domcke, S., Sinha, R., Levine, D. A., Sander, C., & Schultz, N. (2013). Evaluating cell lines as tumour models by comparison of genomic profiles. Nature Communications 4, 2126. doi.org/10.1038/ncomms3126
Slides 13 (notes), 15	Tumour data come from The Cancer Genome Atlas (TCGA) high-grade serous ovarian cancer project.	The Cancer Genome Atlas Research Network. (2011). Integrated genomic analyses of ovarian carcinoma. Nature 474(7353), 609–615. doi.org/10.1038/nature10166
Slides 13 (notes), 15	Cell line data come from the Cancer Cell Line Encyclopedia (CCLE).	Barretina, J., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature 483(7391), 603–607. doi.org/10.1038/nature11003
Slides 15, 23	The study compared 47 CCLE ovarian cancer cell lines with 316 TCGA high-grade serous tumours on copy number, mutations and mRNA, with no drug testing.	Domcke, S., Sinha, R., Levine, D. A., Sander, C., & Schultz, N. (2013). Evaluating cell lines as tumour models by comparison of genomic profiles. Nature Communications 4, 2126. doi.org/10.1038/ncomms3126
Slides 19–20, 33	Figure 3 values (abstracts, r): KURAMOCHI 6, 0.58; SKOV3 2,101, −0.05; A2780 1,363, 0.20; NIHOVCAR3 1,082, 0.30; CAOV3 278, 0.33; IGROV1 255, 0.23.	Domcke, S., Sinha, R., Levine, D. A., Sander, C., & Schultz, N. (2013). Evaluating cell lines as tumour models by comparison of genomic profiles. Nature Communications 4, 2126. doi.org/10.1038/ncomms3126

SECTION	CLAIM IN THE TEXT	SOURCE
Slide 21	KURAMOCHI's original publication called its tumour undifferentiated carcinoma, a classification the paper calls ambiguous; its genome looks high-grade serous.	Domcke, S., Sinha, R., Levine, D. A., Sander, C., & Schultz, N. (2013). Evaluating cell lines as tumour models by comparison of genomic profiles. Nature Communications 4, 2126. doi.org/10.1038/ncomms3126 The paper's wording is 'rather ambiguously classified as undifferentiated carcinoma'.
Slides 24–26	The abstract claims the gap can be bridged 'for all tumour types'; the introduction calls it a pilot study; the discussion says better matches needn't change results.	Domcke, S., Sinha, R., Levine, D. A., Sander, C., & Schultz, N. (2013). Evaluating cell lines as tumour models by comparison of genomic profiles. Nature Communications 4, 2126. doi.org/10.1038/ncomms3126 All four quoted abstract phrases and the last sentence match the published abstract.
Slide 29	A meta-analysis of 28 studies of medical journal citations found 25.4% did not properly support their sentence and 11.9% were major errors.	Jergas, H., & Baethge, C. (2015). Quotation accuracy in medical journal articles: A systematic review and meta-analysis. PeerJ 3, e1364. doi.org/10.7717/peerj.1364
Slide 30; Handout p2	Professional fact-checkers judge a source by leaving it and reading what others say about it (lateral reading).	Wineburg, S., & McGrew, S. (2019). Lateral reading and the nature of expertise: Reading less and learning more when evaluating digital information. Teachers College Record 121(11), 1–40. doi.org/10.1177/016146811912101102
Slides 30–31	The work was funded by the US National Cancer Institute through the TCGA grant and by Stand Up To Cancer, among others.	Domcke, S., Sinha, R., Levine, D. A., Sander, C., & Schultz, N. (2013). Evaluating cell lines as tumour models by comparison of genomic profiles. Nature Communications 4, 2126. doi.org/10.1038/ncomms3126 Acknowledgements: NCI TCGA Genome Data Analysis Center grant U24CA143840 and an SU2C Dream Team grant; D.A.L. was funded by the DoD.
Slide 31; Script	About 1,300 works cite the paper by Crossref's count (1,299 on 14 October 2026), and no correction or retraction is registered.	Crossref. (2026). Metadata record for doi:10.1038/ncomms3126 (is-referenced-by-count). Crossref REST API. On 26 October 2026 the count was 1,304, and Crossref listed no updates (corrections or retractions).
Slide 31 (notes)	The authors work at Memorial Sloan Kettering, where cBioPortal was built.	Cerami, E., et al. (2012). The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. Cancer Discovery 2(5), 401–404. doi.org/10.1158/2159-8290.CD-12-0095 Cerami and colleagues are at the Computational Biology Center, Memorial Sloan-Kettering Cancer Center, the same centre as Domcke and colleagues.
Slide 32	The suitability score is the authors' own weighting, the order isn't meant as a fine ranking, and altered-genome cut-offs were 0.2 for tumours and 0.3 for cell lines.	Domcke, S., Sinha, R., Levine, D. A., Sander, C., & Schultz, N. (2013). Evaluating cell lines as tumour models by comparison of genomic profiles. Nature Communications 4, 2126. doi.org/10.1038/ncomms3126 The paper says the order 'is not meaningful as a finely graduated ranking'.

Workshop 8 · Present your research اعرض بحثك

SECTION	CLAIM IN THE TEXT	SOURCE
Handout p1; Slide 6; Notes	A listener can hold only a few ideas at once, so a talk should carry one message.	Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. Behavioral and Brain Sciences 24(1), 87–114. doi.org/10.1017/S0140525X01003922 The text attributes this to Research chapter 2; Cowan is the standard evidence for a working-memory limit of about four chunks.
Handout p1; Slides 10, 15, 18; Notes	Bourne's rules say no more than one visual per minute of speaking, so about ten slides for a ten-minute talk.	Bourne, P. E. (2007). Ten simple rules for making good oral presentations. PLoS Computational Biology 3(4), e77. doi.org/10.1371/journal.pcbi.0030077 The one-visual-per-minute rule of thumb is in Bourne's Rule 8.
Handout p2; Slides 23, 24, 26; Notes	110 engineering students heard the same 8-minute MRI narration; sentence-headline-and-picture slides gave essay scores of 9.4 vs 6.7 of 15, and 7.8 vs 5.5 of 10 a week later.	Garner, J. K., & Alley, M. P. (2013). How the design of presentation slides affects audience comprehension: A case for the assertion-evidence approach. International Journal of Engineering Education 29(6), 1564–1579.
Handout p2; Slides 25, 28, 43; Notes	In 981 TCGA breast tumours, basal-like tumours have about 7.5 times the median PHGDH mRNA of luminal A (3,899 vs 523; n per subtype 171, 78, 36, 197, 499).	Hoadley, K. A., Yau, C., Hinoue, T., Wolf, D. M., Lazar, A. J., Drill, E., et al. (2018). Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. Cell 173(2), 291–304.e6. doi.org/10.1016/j.cell.2018.03.022 Re-run from cBioPortal (brca_tcga_pan_can_atlas_2018, RSEM, PanCancer SUBTYPE): medians and n match exactly; ratio 7.45. Notes round it to 'about seven times'.
Handout p2; Slides 25, 28	The breast cancer data shown were taken from the TCGA PanCancer Atlas through cBioPortal.	Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., et al. (2012). The cBio Cancer Genomics Portal: An open platform for exploring multidimensional cancer genomics data. Cancer Discovery 2(5), 401–404. doi.org/10.1158/2159-8290.CD-12-0095 cBioPortal also asks users to cite Gao, J., et al. (2013). Science Signaling 6(269), pl1. doi.org/10.1126/scisignal.2004088.
Handout p2; Slides 25, 28	Breast tumours are grouped into basal-like, HER2-enriched, normal-like, luminal A and luminal B subtypes.	Perou, C. M., Sørlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., et al. (2000). Molecular portraits of human breast tumours. Nature 406(6797), 747–752. doi.org/10.1038/35021093 TCGA's subtype calls use the PAM50 classifier: Parker, J. S., et al. (2009). J Clin Oncol 27(8), 1160–1167. doi.org/10.1200/JCO.2008.18.1370.
Handout p3; Slides 32, 34; Notes	Bourne's rule 9 is to review audio or video recordings of your own talks; rule 7 is to practise, trying the talk on friendly colleagues first.	Bourne, P. E. (2007). Ten simple rules for making good oral presentations. PLoS Computational Biology 3(4), e77. doi.org/10.1371/journal.pcbi.0030077 Rule 7 ('Practice and time your presentation') suggests giving it first to research collaborators or a lab group.
Handout p3; Slide 41; Notes	3MT rules: three minutes strictly (over means disqualified), one static slide with no animation, no props, and spoken word only (no poems, raps or songs).	University of New England. (n.d.). Three Minute Thesis (3MT): Rules and judging criteria. UQ's 3MT competitor guide (threeminutethesis.uq.edu.au/resources/3mt-competitor-guide) also bans slide transitions, animation and props.

SECTION	CLAIM IN THE TEXT	SOURCE
Slides 8, 55; Notes	Audiences trust a speaker who raises the weaknesses before anyone asks.	Williams, K. D., Bourgeois, M. J., & Croyle, R. T. (1993). The effects of stealing thunder in criminal and civil trials. Law and Human Behavior 17(6), 597–609. doi.org/10.1007/BF01044684 The text attributes this to Research chapter 2. The evidence ('stealing thunder') comes from mock-trial studies, where disclosing first raised credibility.
Slide 9; Notes	Bourne: the audience's time is precious and should not be abused with uninteresting preliminary material.	Bourne, P. E. (2007). Ten simple rules for making good oral presentations. PLoS Computational Biology 3(4), e77. doi.org/10.1371/journal.pcbi.0030077 Quoted word for word from Rule 3.
Slide 14; Notes	Bourne's rule 5: be logical, with a clear beginning, middle and end, the same order as a paper.	Bourne, P. E. (2007). Ten simple rules for making good oral presentations. PLoS Computational Biology 3(4), e77. doi.org/10.1371/journal.pcbi.0030077
Slide 28	PHGDH expression is shown as mRNA quantified with RSEM.	Li, B., & Dewey, C. N. (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12, 323. doi.org/10.1186/1471-2105-12-323
Slide 28; Notes (model talk)	PHGDH, the first enzyme of serine synthesis, matters in some breast cancers, which is why its level by subtype is worth showing.	Possemato, R., Marks, K. M., Shaul, Y. D., Pacold, M. E., Kim, D., Birsoy, K., et al. (2011). Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. Nature 476(7360), 346–350. doi.org/10.1038/nature10350 Possemato found PHGDH protein raised in 70% of ER-negative breast cancers, and needed by cell lines with high PHGDH.
Slide 40; Notes	The Three Minute Thesis began at the University of Queensland in 2008 with 160 research students, inspired by drought-era shower timers; now 900+ universities in 85+ countries.	University of Queensland. (n.d.). About 3MT. Three Minute Thesis. The UQ page says three-minute egg timers on bathroom walls during a Queensland drought.
Slide 42; Notes	3MT talks are judged on comprehension and content, and engagement and communication, weighted equally.	University of New England. (n.d.). Three Minute Thesis (3MT): Rules and judging criteria. UNE lists six criteria, three in each group, each equally weighted.

Workshop 9 · Getting into research as a student

كيف تبدأ في البحث العلمي وأنت طالب

SECTION	CLAIM IN THE TEXT	SOURCE
Handout p2; Slides 7, 23, 25, 26; Notes; Run sheet	KAUST VSRP: third year up, GPA 3.5, 2–6 months any time, USD 1,000 a month, flights and a room; transcript, CV, statement, one recommender; IELTS 6.5 on acceptance.	King Abdullah University of Science and Technology. (n.d.). Visiting Student Research Program (VSRP) internships. KAUST Admissions. Matches the page on 26 Oct 2026. IELTS 6.5 needs at least 6 in each section; some degree holders are exempt.

SECTION	CLAIM IN THE TEXT	SOURCE
Handout p2; Slides 7, 23, 25; Notes; Run sheet	VBC Summer School, Vienna: open worldwide to students with two years of life sciences; nine weeks; €1,400 after tax plus housing; call opens 15 November.	Vienna BioCenter. (n.d.). VBC Summer School. VBC Training.
Handout p2; Slides 7, 25; Notes; Run sheet	KAUST SSI is for Saudi seniors and recent graduates in applied maths, computer science, electrical and computer engineering and statistics; GPA 3.5; 6–10 weeks.	King Abdullah University of Science and Technology, CEMSE Division. (n.d.). Saudi Summer Internship (SSI). The page also admits current master's students; 2026 dates were 15 June to 6 August.
Slide 9; Notes	Search your question in PubMed to find recent authors working on it.	National Center for Biotechnology Information, U.S. National Library of Medicine. (n.d.). PubMed. NCBI says PubMed holds more than 40 million citations and abstracts.
Slide 9; Notes	Google Scholar can also be used to find recent authors on a question.	Google. (n.d.). About Google Scholar.
Slides 23, 24; Notes; Run sheet	RISE Germany is only for undergraduates at North American, British and Irish universities.	DAAD. (n.d.). RISE Germany. German Academic Exchange Service. The quoted eligibility line matches the page.
Slide 26; Notes; Run sheet	On 14 October 2026 KAUST listed 154 VSRP projects; you apply to one project at a time.	King Abdullah University of Science and Technology. (n.d.). Internship search (VSRP project list). KAUST Admissions. A dated snapshot: the list showed 157 projects on 26 October 2026. Divisions filter includes Biological and Environmental Sciences and Engineering, and Biomedical Sciences.
Slide 31; Notes	Most medical journals use the ICMJE's four criteria to decide who counts as an author.	International Committee of Medical Journal Editors. (2026). Defining the role of authors and contributors. Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals. All four must be met: substantial contribution; drafting or critical revision; final approval; agreement to be accountable.

Workshop 10 · Academic writing, from abstract to discussion

الكتابة الأكاديمية من الملخّص للمناقشة

SECTION	CLAIM IN THE TEXT	SOURCE
Handout p1; Slides 6, 7; Notes	The ICMJE calls IMRAD not an arbitrary format but a reflection of the process of scientific discovery.	International Committee of Medical Journal Editors. (2026). Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals: Preparing a manuscript for submission to a medical journal. icmje.org Quoted word for word.

SECTION	CLAIM IN THE TEXT	SOURCE
Handout p1; Slide 8; Notes	In BMJ, JAMA, The Lancet and NEJM (1935–1985), IMRAD appeared in the 1940s, reached 80% in the 1970s and was the only pattern by the 1980s.	Sollaci, L. B., & Pereira, M. G. (2004). The introduction, methods, results, and discussion (IMRAD) structure: A fifty-year survey. Journal of the Medical Library Association 92(3), 364–367. PMID 15243643; no DOI.
Handout p1; Slides 9, 10, 55; Notes; Run sheet	Scientific Reports: abstract up to 200 words, unstructured, no references; title up to 20 words, no puns; suggested order Introduction, Results, Discussion, Methods.	Scientific Reports. (n.d.). Submission guidelines. Nature Portfolio.
Handout p1; Slides 10, 48; Notes; Run sheet	The BMJ wants a structured abstract of 250–300 words under eight headings from Objectives to Conclusions, plus a box on what is known and what the study adds.	The BMJ. (n.d.). Research: Article types and preparation. Resources for authors. The page lists eight headings Objectives–Conclusions, then Trial registration; Interventions can be dropped.
Handout p1; Slide 13; Notes	John Swales (1990) described three moves in research introductions: establish the territory, find the niche, occupy it.	Swales, J. M. (1990). Genre analysis: English in academic and research settings. Cambridge University Press. Known as the CARS (Create a Research Space) model; the slide also cites the UMass Amherst Writing Center summary.
Handout p2; Slides 21, 22, 23, 24; Notes	ICMJE Methods: clarity about how and why, enough for others with the data to reproduce results, an ethics approval statement; anything learned during the study belongs in Results.	International Committee of Medical Journal Editors. (2026). Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals: Preparing a manuscript for submission to a medical journal. icmje.org Wording otherwise matches the ICMJE page.
Handout p2; Slide 25; Notes	The EQUATOR Network keeps the reporting guidelines for each study design.	EQUATOR Network. (n.d.). Enhancing the QUAlity and Transparency Of health Research.
Handout p2; Slide 25; Notes	CONSORT is the reporting checklist for randomised trials.	Hopewell, S., Chan, A.-W., Collins, G. S., Hróbjartsson, A., Moher, D., Schulz, K. F., et al. (2025). CONSORT 2025 statement: Updated guideline for reporting randomised trials. BMJ 389, e081123. doi.org/10.1136/bmj-2024-081123 CONSORT 2025 replaced CONSORT 2010; the text names no version.
Handout p2; Slide 25; Notes	STROBE is the reporting checklist for observational studies, like the made-up cohort.	von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., & Vandenbroucke, J. P. (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. The Lancet 370(9596), 1453–1457. doi.org/10.1016/S0140-6736(07)61602-X
Handout p2; Slide 25; Notes	PRISMA is the reporting checklist for systematic reviews.	Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ 372, n71. doi.org/10.1136/bmj.n71

SECTION	CLAIM IN THE TEXT	SOURCE
Handout p2; Slide 25; Notes	STARD is the reporting checklist for diagnostic accuracy studies.	Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L., et al. (2015). STARD 2015: An updated list of essential items for reporting diagnostic accuracy studies. BMJ 351, h5527. doi.org/10.1136/bmj.h5527
Handout p2; Slide 25; Notes	ARRIVE is the reporting checklist for animal studies.	Percie du Sert, N., Hurst, V., Ahluwalia, A., Alam, S., Avey, M. T., Baker, M., et al. (2020). The ARRIVE guidelines 2.0: Updated guidelines for reporting animal research. PLOS Biology 18(7), e3000410. doi.org/10.1371/journal.pbio.3000410
Handout p2; Slides 29, 30, 32; Notes	ICMJE Results: main findings first, don't repeat all table data, give counts behind percentages, report every outcome named in Methods.	International Committee of Medical Journal Editors. (2026). Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals: Preparing a manuscript for submission to a medical journal. icmje.org
Handout p3, p4; Notes	Cell studies suggest some breast cancer cells depend on PHGDH, the first enzyme of serine synthesis; the made-up result 'fits lab work on the serine pathway'.	Possemato, R., Marks, K. M., Shaul, Y. D., Pacold, M. E., Kim, D., Birsoy, K., et al. (2011). Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. Nature 476(7360), 346–350. doi.org/10.1038/nature10350 The PHGDH study itself is made up; this real background claim is not cited in the text.
Handout p3, p4; Slides 31, 32, 39, 40, 41; Notes	The made-up study reports an adjusted hazard ratio from a Cox model adjusted for stage.	Cox, D. R. (1972). Regression models and life-tables. Journal of the Royal Statistical Society: Series B (Methodological) 34(2), 187–202. doi.org/10.1111/j.2517-6161.1972.tb00899.x The numbers are invented; only the method is real.
Handout p3; Slides 37, 38, 43; Notes	ICMJE Discussion: summarise the main findings briefly, set them in context, explore explanations, state limitations, give implications; link conclusions to goals without unsupported claims.	International Committee of Medical Journal Editors. (2026). Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals: Preparing a manuscript for submission to a medical journal. icmje.org The slide 38 quote matches the ICMJE wording.
Handout p4; Slides 46, 47; Notes	Of 72 trials with a non-significant main result, 42 abstracts (58%) still had conclusions leaning towards the treatment.	Boutron, I., Dutton, S., Ravaud, P., & Altman, D. G. (2010). Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. JAMA 303(20), 2058–2064. doi.org/10.1001/jama.2010.651
Slide 14; Notes	The ICMJE asks an introduction for context, the specific purpose, only directly pertinent references, and no data or conclusions from the work itself.	International Committee of Medical Journal Editors. (2026). Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals: Preparing a manuscript for submission to a medical journal. icmje.org
Slide 24	A redoable Methods sentence names the software and version, e.g. R 4.4.	R Core Team. (2024). R: A language and environment for statistical computing (Version 4.4). R Foundation for Statistical Computing, Vienna, Austria. Current release on 26 Oct 2026 is R 4.6.1; 4.4 is a valid earlier version.
Slide 26; Notes	Scientific Reports asks for any use of a large language model to be documented in the Methods.	Scientific Reports. (n.d.). Submission guidelines. Nature Portfolio.

SECTION	CLAIM IN THE TEXT	SOURCE
Slide 26; Notes	COPE holds authors fully responsible for everything in a paper, including what an AI tool produced.	Committee on Publication Ethics. (2023). Authorship and AI tools: COPE position statement. COPE also asks for AI use to be disclosed in the Methods or similar section.
Slide 26; Notes	The ICMJE also holds authors responsible for any AI-assisted material and asks them to disclose AI use.	International Committee of Medical Journal Editors. (2026). Artificial intelligence (AI)-assisted technology: AI use by authors. Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals.
Slide 48; Notes	The ICMJE asks for structured abstracts for original research, noting important limitations and not overinterpreting findings.	International Committee of Medical Journal Editors. (2026). Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals: Preparing a manuscript for submission to a medical journal. icmje.org
Notes	Audiences, and readers, trust an author who brings up the weaknesses before anyone asks.	Williams, K. D., Bourgeois, M. J., & Croyle, R. T. (1993). The effects of stealing thunder in criminal and civil trials. Law and Human Behavior 17(6), 597–609. doi.org/10.1007/BF01044684 The text attributes this to Research chapter 2. The evidence comes from mock-trial studies, not from readers of papers.

Al Basira × Zahra · 22 October 2026 البصيرة × زهرة: معًا لمجتمع واعٍ

SECTION	CLAIM IN THE TEXT	SOURCE
Handout, Slides	About 37 in 100 new cancers in women in Saudi Arabia are breast cancer; about 1,022 deaths from breast cancer in 2024.	International Agency for Research on Cancer. Global Cancer Observatory, GLOBOCAN 2024 fact sheet: Saudi Arabia. gco.iarc.who.int/media/globocan/factsheets/populations/682-saudi-arabia-fact-sheet.pdf The fact sheet gives 36.6% of new cancers in women.
Handout, Notes	4,079 Saudi women were diagnosed with breast cancer in 2023; median age 52, youngest 15.	Saudi Health Council, National Cancer Center. Cancer Incidence Report, Saudi Arabia 2023. shc.gov.sa/ar/Centers/NCC/AnnualReports/Cancer%20Incidence%20Report%202023.pdf The report gives a range of 15 to 105 years.
Handout, Notes, Slides	Screening from 40 every two years, from 30 with a risk factor; most breast cancers are painless and most lumps benign; deodorants are not linked to breast cancer; booking through Sehhaty.	Ministry of Health, Saudi Arabia (2024). Breast cancer awareness guide. moh.gov.sa/awarenessplateform/ChronicDisease/Documents/BreastCancer.pdf

SECTION	CLAIM IN THE TEXT	SOURCE
Notes, Slides	The changes worth seeing a doctor about, the main risk factors, and that men can get breast cancer too.	World Health Organization. Breast cancer: fact sheet. who.int/news-room/fact-sheets/detail/breast-cancer

PROGRAMME 5

Free guides الأدلة المجانية

The free guides on the website, open to everyone. 15 claims, 15 sources.

The free guides

15 CLAIMS

SECTION	CLAIM IN THE TEXT	SOURCE
Guides page	What a p-value is and is not.	Wasserstein, R. L., & Lazar, N. A. (2016). The ASA Statement on p -Values: Context, Process, and Purpose. The American Statistician 70(2), 129-133. doi.org/10.1080/00031305.2016.1154108
Guides page	DESeq2 for testing differences in gene expression from raw counts.	Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15(12), 550. doi.org/10.1186/s13059-014-0550-8
Guides page	The Cancer Genome Atlas: matched tumour, molecular and clinical data.	The Cancer Genome Atlas Research Network, et al. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nature Genetics 45(10), 1113-1120. doi.org/10.1038/ng.2764
Guides page	Survival data in TCGA and its curated survival endpoints.	Liu, J., et al. (2018). An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. Cell 173(2), 400-416.e11. doi.org/10.1016/j.cell.2018.02.052
Guides page	cBioPortal as a way to browse and download cancer genomics data.	Cerami, E., et al. (2012). The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. Cancer Discovery 2(5), 401-404. doi.org/10.1158/2159-8290.cd-12-0095
Guides page	The Gene Expression Omnibus, a public repository of expression data.	Edgar, R. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Research 30(1), 207-210. doi.org/10.1093/nar/30.1.207
Guides page	DepMap: genome-wide dependency screens across cancer cell lines.	Tsherniak, A., et al. (2017). Defining a Cancer Dependency Map. Cell 170(3), 564-576.e16. doi.org/10.1016/j.cell.2017.06.010
Guides page	Retractions: why published papers get withdrawn.	Fang, F. C., Steen, R. G., & Casadevall, A. (2012). Misconduct accounts for the majority of retracted scientific publications. Proceedings of the National Academy of Sciences 109(42), 17028-17033. doi.org/10.1073/pnas.1212247109
Guides page	Retraction Watch and its database of retracted papers.	Retraction Watch Database. The Center for Scientific Integrity. retractiondatabase.org
Guides page	Reading a paper in three passes.	Keshav, S. (2007). How to read a paper. ACM SIGCOMM Computer Communication Review 37(3), 83-84. doi.org/10.1145/1273445.1273458
Guides page	Spin: reporting that makes results look better than they are.	Boutron, I. (2010). Reporting and Interpretation of Randomized Controlled Trials With Statistically Nonsignificant Results for Primary Outcomes. JAMA 303(20), 2058. doi.org/10.1001/jama.2010.651
Guides page	Zotero, the free reference manager.	Corporation for Digital Scholarship. Zotero. zotero.org

SECTION	CLAIM IN THE TEXT	SOURCE
Guides page	PubMed, the free search engine for the biomedical literature.	National Library of Medicine. PubMed. pubmed.ncbi.nlm.nih.gov
Guides page	Google Scholar for searching across disciplines.	Google. Google Scholar. scholar.google.com
Guides page	Combining search terms with AND, OR and NOT.	National Library of Medicine. PubMed User Guide. pubmed.ncbi.nlm.nih.gov/help

What changed in this edition

This edition covers every programme, not just the Research chapters: the Research cohort (Chapters 0 to 6, session plans, teaching notes and assignments), the Skills cohort, Principles of Research, our workshops and the free guides.

Every source was checked again against its record: DOIs against Crossref, and standards, guidelines, reports and software against their official pages.

Four corrections to the text came out of the check. Chapter 2 now gives the ICMJE's four authorship criteria, and no longer calls gift and ghost authorship misconduct in the formal sense. A session reading is now labelled as the full Crash Course Biology series. Workshop 10's notes now say only what the ICMJE actually asks for in a Methods section.

The federal definition of research misconduct cited for Chapter 2 now sits at 42 CFR §93.234, after the 2024 revision.

Figures

We drew the diagrams, schematics and decision guides for Al Basira, and they come under the same CC BY 4.0 licence as the rest of the work. Any figure taken from elsewhere gives its source and licence where it appears.

If you find a claim in our material that isn't listed here and should be, or a source here that doesn't support the claim it's attached to, write to contact@researchbasira.com. That is a correction we want.