Distinguishing Association from Causation in Media Headlines

Milo Schield[1] and Robert Raymond[2]
[1] Augsburg College, Minneapolis, MN
[2] Emeritus, University of Saint Thomas

Abstract

Headlines from 2,000 news stories were analyzed for the presence of 727 keywords indicating an association, a causal connection or something in-between. 71% had such keywords. Of those with such keywords, very few (6%) had keywords clearly indicating causation or association. Most (94%) had "between" keywords: keywords that described an association but had a causal connotation. Between keywords included action verb keywords such as *ups* or *cuts* (61%), comparison keywords such as *more* or *less* (19%), sufficient keywords such as *prevent* or *stop* (8%) and temporal or quasi-causal keywords such as *after* and *due to* (7%). A content analysis of three statistics textbooks indicates that statisticians may use *effect* without implying causation. This data may be useful for both statisticians and journalists in trying to understand how the other group distinguishes association from causation.

1. Introduction

The goal of this paper is to analyze how journalists use ordinary English to describe associations in the headlines of news stories and compare this with how statisticians describe association and causation in their text books.

1.1. Words Used by Journalists to Describe an Association

Headlines were chosen because they must focus on the essentials of a story in a very limited space and what they focus on is what the reader learns – perhaps the only thing. Consider headlines all based on the same journal article: *Health Insurance and Mortality* by Wilper et al (2009). Here are some of the headlines:

- 45,000 deaths *attributable to* uninsurance (PNHP 9/17/09)
- 45,000 American deaths *associated with* lack of insurance (CNN 9/18)
- Study: Uninsured Americans Have 40 Percent *Higher* Death Risk (Ivanhoe, 9/18)
- Study: 45,000 Uninsured Die a Year (CBS News, 9/17/09)
- No health coverage *tied to* 45,000 deaths a year (Rueters MSNBC 9/17)
- Lack of insurance *linked to* 45,000 deaths (White Coat News, 9/17/09)
- Study *links* 45,000 U.S. deaths to lack of insurance (Reuters, 9/17/09)
- Study: 45,000 U.S. Deaths *From* Lack of Insurance (MoneyNews 9/17/09)
- One American dies every 12 minutes *due to* no health insurance (blog DR 9/17)
- 45,000 Americans die … *because of* lack of health insurance (blog MyDD 9/17)
- Lack of Health Insurance *cause* 44789 deaths in United States every year (blog)
- Lack of insurance *to blame for* almost 45,000 deaths: Study (HealthDay 9/17/09)
- Lack of Health Insurance *Kills* 45,000 a Year (Health Insurance com, Inst.)

Look at the variety of words (in italics) that indicate the kind of relationship between being uninsured and dying: *attributable to, associated with, percent higher, tied to, linked to, links, from, due to, because of, kills, cause* and *blame.* Notice the progression of

words from intransitive (attributable to, associated with, tied to, linked to, are from, due to, because of) to transitive active (kills, causes).

Note the use of an action verb (*kills*) to describe the association in the last title above. Seeing that journalists use action verbs to describe an association is a forerunner of what will be a primary result of this study.

Figure 1: Lack of Health Insurance Kills (Uninsurance Kills)

But data from one story is not an adequate basis for generalizing. The first step is to identify those keywords that distinguish association from causation. The second step is to acquire a wide variety of news stories. The third step is to match the selected key-words with the news headlines and analyze the results.

1.2. Keywords that Indicate Association or Causation

Before starting, one needs an idea of the many ways that ordinary English indicates different kinds of relationships by using certain keywords.

Association keywords are taken to include *associate*, *relate* and *predict*:

- Tree-Lined Streets Associated With Lower Childhood Asthma Rates
- Kids' problems related to drinking in pregnancy
- Smiles Predict Marriage Success; Finger Length Predicts Aggression in Men

Causation keywords are taken to include *cause*, *effect* and *result*.

- Damp, Moldy Homes May Cause Depression
- Effects of child head injuries can last for years

Other types of associations are indicated by various keywords as follows:

- Sufficiency keywords: Consider titles that include keywords indicating sufficiency such as *cancel, eliminate, hinder, kill* and *prevent*. Note that most of these involve a negative outcome (e.g., stop rather than start).
  - Milk cancels health benefit of drinking tea: study
  - Simple regimen eliminates chronic bad breath
  - Cream with green tea extract hinders HIV transmission: study
  - WHO: Sun Exposure Kills 60,000 Worldwide Each Year
  - HPV vaccine prevents genital warts in males
  - Diet Plans Produce Similar Results

- Temporal keywords: Consider titles with time-related keywords such as *after*, *follow* and *leads*:
  - Fracture risk doubled after obesity surgery
  - When toddlers point a lot, more words will follow
  - Childhood Obesity Leads to Adolescent Obesity

- Quasi-Causal keywords: Consider titles containing keywords such as *as* and *due to*.
  - As gas prices go up, auto deaths drop;

Child abuse spikes as U.S. economy founders
- Increased Infant Death Rate Due to Rise in Premature Births
  Study: 12 percent of Indian deaths due to alcohol

- Connector keywords: Consider titles with keywords such as *link, factor* or *connect*
  - No Link Between 100% Juice and Kids' Overweight: Study
    Study links breastfeeding to high grades, college entry
    Verbal abuse from teachers linked to risk of early sexual intercourse
  - Father's Age a Factor in Infertility

- Comparative keywords: Consider titles using comparative keywords such as *more, less* or one of the many comparatives ending in "*er*."
  - British poll: Religion does more harm than good
    Dads More Likely Than Moms to Pass on MS
    Babies Born at Night More Likely to Die
  - U.S. Blacks Hear Better Than Whites
    Bigger Tableware Helps Widen Waistlines

- Action verb keywords: In saying "the child hit the ball," we indicate the child caused the motion in the ball. Consider titles using action verb keywords:
  - Gene increases depression risk: study
  - Study: Estratest doubles breast cancer risk; Weddings boost mood: study
  - 45,000 excess deaths annually linked to lack of health insurance: study

- Action noun keywords: If someone says, "Laetrile is a cancer killer" they are saying that Laetrile kills cancer. Consider titles using action noun keywords:
  - bad water remains a killer
  - Staying silent in marital spats a killer for women

These are the kinds of keywords that will be used in computer-match analysis. The value of any computer-text matching study depends critically on the range of keywords that is associated with a given semantic relationship. Unfortunately, there is no agreement on which words indicate association or causation, so this attempt is provisional.

1.3. Analyzing Titles of News Stories

Titles of 2,000 news stories have been collected between 8/2005 and 9/2009. Almost all these stories were selected from the web – primarily from Yahoo Health. The news articles are typically one or two pages. Sports, weather, stock prices or original research studies were excluded. News articles were included if they involved numbers and they had any of the following: had "study," "survey" or "report" in the title, involved a study, poll, survey or report, involve diagnostic tests (medical or otherwise), involved longitudinal data or subject manipulation, involved random assignment or random selection, involved a sample, sample size or margin of error, had "significantly" or "(in)significant" in the text, involved taking into account a confounder, or used statistics as evidence for causation.

2. Methodology and Data

This study involved computer matching 727 keywords with 2,000 news headlines. This combination involves over a million possible matches. For more background, see Schield and Raymond (2008).

Appendix A shows the counts for each of the 727 fields (or their associated lemma) and the categories in which they were summarized. Appendix B presents some of the steps in processing the text data.

Unfortunately, simple keyword matches are not sufficient to identify either association or causation. Consider these examples that don't state either association or causation:

- Homeowners' Association in debt.
- The new link is very handy.
- Saving lives is a great cause.

Nevertheless, a computer analysis of keywords allows a scale of analysis that would be difficult to do manually. This paper uses isolated elements of syntax as a basis for inferring the semantics of the headline in question.

Matching does allow for the presence of multiple keywords in the same title. Consider this title: *Eating nuts may help cure cancer-related risks*. If *help*, *cure* and *related* are counted separately, this title will be counted three times.

3. Association vs. Causation

The following tables present the counts (the number of matches) in the top row. The middle row shows the distribution of these counts among the various categories. Although multiple matches per title are possible, they are unlikely so the bottom row is the percentage of the 2,000 titles that involve matches with a keyword or group of keywords.

The first step is to start with those keywords that clearly identify association and causation. Then consider those keywords that indicate various kinds of relationships that may imply or indicate causation to the non-professional, but do not indicate causation to a professional in the social or physical sciences. Note that these choices are provisional. No group or organization has taken a stand on which words indicate causation and which indicate association.

3.1. Causation Keywords

Consider those words that most definitely indicate causation. In most cases, the term shown stand for all forms of the word as a noun, verb or adjective. Thus, *cause* stands for *cause, causes, caused, causing* and *causal*. These header words are called lemmas. See *Lemma* in Wikipedia for more details.

Table 1: Causation Keywords: Cause, Effect Result

Keyword	ALL	Cause	Effect	Result
Counts	43	23	13	7
% of Group	100%	53%	30%	18%
% of Titles	2%	1%	1%	0%

Very few of the titles (2%) have keywords indicating causation. This may reflect the extreme caution taken by journalists to avoid claims that might trigger a legal liability.

3.2. Sufficient-Action Keywords

One criterion for a deterministic cause is sufficiency. There are numerous verbs in English that indicate sufficiency. The sufficiency of these words may depend on the context. To *ban*, may mean to stop something, but the policy may not be enforced or honored. E.g., the city banned smoking in apartments.

Table 2: Sufficient-Action Keywords: Prevent, Stop, End, etc.

Keyword	ALL (19)	Prevent, Stop	End, Start	Kill, Cure	Avoid, Ban	Quit, Block	Ward off, Stave off	Others (7)
Counts	110	**41**	**12**	**15**	**14**	**10**	**7**	**11**
% of Group	100%	37%	11%	14%	13%	9%	6%	10%
% of Titles	6%	2%	1%	1%	1%	1%	0%	1%

Of the 53 lemmas tested for matches, 19 had matches. Very few of the titles (6%) included any of these words. So even if causal keywords were broadened to include these sufficient action keywords, only 8% of the titles would contain causation keywords.

3.3. Association Keywords

Consider four lemmas that might indicate an association: *associate, relate, correlate and predict*. As mentioned before, this choice is provisional. Statisticians have taken no stand on which words clearly indicate association aside from *associate* itself.

Table 3: Association Keywords: Predict, Relate and Associate

Keyword	ALL (4 lemmas)	Predicts Predict	Related Relate	Associated Associate	Correlated Correlate
Counts	46	**30**	**12**	**4**	**0**
% of Group	100%	63%	28%	9%	0%
% of Titles	2%	2%	1%	0%	0%

Of these titles, very few (2%) have words indicating association. In titles containing any of these four keywords, *predict* (63%) and *relate* (28%) are most common, while *associate* (9%) and *correlate* (0%) appear less often.

Associations can be indicated by comparisons (e.g., Women live *longer* than men). But comparisons can indicate causation (e.g., Students who study more get better grades).

Superlatives indicate an implicit comparison (e.g., the *best* team is the team that is at least as good as – if not better than – all the rest.

3.4. Comparison Keywords

Comparison keywords involved four different types: compare adjectives (e.g., *more, better*), superlatives (e.g., *most, best*), compare adverbs (e.g., *largely, mostly*) and words that indicate quantity without involving a number (e.g., *often, much* and *many*).

Comparative adjectives: The following table gives the most common comparative adjectives. These are the exact word matches – not the lemma forms. Note that *lower* was not included since it is also an action verb.

Table 4: Comparative-Adjective Keywords: More, Better, Higher, etc.

Keyword	ALL	more	better	higher	less	earlier	longer	greater	faster	Other
Counts	208	**97**	**24**	**23**	**23**	**13**	**9**	**4**	**3**	**12**
% of Group	100%	47%	12%	11%	11%	6%	4%	2%	1%	6%
% of Titles	10%	5%	1%	1%	1%	1%	0%	0%	0%	1%

Among those titles involving comparative adjectives, *more* is most common (47%), with *better, higher* and *less* accounting for almost all (81%) of the matches.

Superlatives usually involve an implicit comparison.

Table 5: Superlative Keywords: Most, First, Last, etc.

Keyword	ALL	most	first	last	best	highest	lowest	Other
Counts	52	**28**	**9**	**5**	**2**	**2**	**2**	**4**
% of Group	100%	54%	17%	10%	4%	4%	4%	8%
% of Titles	3%	1%	0%	0%	0%	0%	0%	1%

Superlatives are rare (3%). Among the titles with superlatives, over half (54%) involved *most*.

Only two of the titles (0.1%) included compare adverbs (*largely* and *mostly*).

Schield (2005) identified 545 words that indicated a quantity without indicating a specific number. Of these, 24 were matched against the 2,000 headlines. While only 4% of these titles included any of the 24 keywords presented, the big three are definitely *many*, *often* and *common*. Since these quantity words don't give any context for a comparison, they have been omitted in summarizing the comparative keywords.

These three sources of comparison keywords are summarized as follows:

Table 6: Comparison Keywords: More, Most and Mostly

Keyword	ALL	Compare adjectives	Superlatives	Compare adverbs
Counts	263	**208**	**53**	**2**
% of Group	100%	79%	20%	1%
% of Titles	13%	10%	3%	0%

Roughly a tenth (13%) of the headlines contains one of these comparative keywords. Comparative adjectives (79%) are most common.

3.5. Temporal Keywords

Some kinds of causation clearly involve a temporal relation. In a game of pool, hitting the cue ball with the cue occurs first, the motion of the cue ball second, and its contact with another ball occurs thereafter. But temporal words don't always indicate causation. In deductive logic, this is called the *post hoc* fallacy: the fallacy of assuming that what comes before is the cause of what follows.

Table 7: Temporal Keywords: Leads, After, etc.

Keyword	ALL (6 lemmas)	Lead, leads, leading	After	Before	Follow, following	Precede, Subsequent
Counts	65	**30**	**23**	**6**	**6**	**0**
% of Group	100%	46%	35%	9%	9%	0%
% of Titles	3%	2%	1%	0%	0%	0%

About 3% of the titles included temporal keywords. Among those titles having temporal words, the majority (81%) involve leads (46%) and after (35%).

3.6. Quasi-Causal Keywords: As, Due to

Quasi causal keywords may indicate causation but this depends on the context. Association: As he walked down the street, he waved to the crowd. Causation: As children grow older, they grow taller. The pressure increased due to the increasing temperature.

Table 8: Quasi-Causal Keywords: As, Factor, Due to, Given, Because

Keyword	ALL	As	Factor	Due to	Given	Because of	Others
Counts	28	**16**	**7**	**3**	**2**	**0**	**0**
% of Group	100%	57%	25%	11%	7%	0%	**0**
% of Titles	3%	2%	1%	0%	0%	0%	0%

Only 3% of these titles involved any of these keywords. Of those titles involving these keywords, the majority involved *as* (57%). It seems surprising that none of the titles involved *because of*, but perhaps that is a result of its having more letters than *due to*.

3.7. Action Verb Keywords

Action verbs often indicate causation (e.g., He helps around the house, she cut the bread) In this study, 72 lemmas for action verbs were tested for matches. Of these, 16 had no match. This table shows the top 26 action verbs in order by descending prevalence.

Table 9: Action Verb Keywords: Links, Helps, Ups, Cuts, etc.

Keyword	ALL	Link	Helps	Ups	Cuts	Raises	Boosts	Other
Counts	858	**106**	**91**	**49**	**44**	**41**	**38**	**489**
% of Group	100%	12%	11%	6%	5%	5%	4%	57%
% of Titles	43%	5%	5%	2%	2%	2%	2%	24%

Of these 2,000 headlines, 43% involved action verbs from 55 lemmas. Among the 858 matches involving action verbs, *link* (12%) and *help* (11%) were the most common lemmas. Other matching action verbs (in order by decreasing prevalence) not shown include *gets, increases, lowers, affect, fights, does, changes, reduces, makes, protects, drops, treats, eases, doubles, grows, improves, keeps, triggers, controls* and *delays*.

Note that the lemma *link* (106) included *linked* (71), *link* (20) and *links* (15).

3.8. Action Noun Keywords

Action nouns are a way of indicating an action without using the verb. E.g., Nuts are a powerful cancer-preventer. In this study, 48 action nouns were tested for matches. Of these only nine matched. This may indicate a lack of creativity in generating the original list or it may indicate the paucity of such uses by those who write headlines for articles.

Table 10: Action Noun Keywords: Killer, Booster, etc.

Keyword	ALL	Killer	Booster	Inhibitor	Reducer
Counts	9	**6**	**1**	**1**	**1**
% of Group	100%	67%	11%	11%	11%
% of Titles	0%	0%	0%	0%	0%

Less than 1% of the titles involved an action noun keyword. Of those titles that contained an action noun, *killer* dominated (67%).

3.9. Summarizing Association-Between-Causation Keywords

The data in these individual tables can now be summarized into a single table with associate (A) on the left and cause (C) on the right. Everything between these is considered "between" (B). Note that the quantity category has been excluded

Table 11: A-B-C Summary of Types of Keywords in Headlines

	ALL	**Associate**	Action verb, noun	Temporal	Quasi-causal	Compare	Sufficient	**Cause**
Count	1,420	**46**	**865**	**65**	**28**	**263**	**110**	**43**
% of ABC	100%	3%	61%	5%	2%	19%	8%	3%
% of All	71%	2%	43%	3%	1%	13%	6%	2%

While the majority (71%) of these 2,000 titles included at least one of the keywords, very few (4%) of those included keywords specified either association (2%) or causation (2%). Of the titles containing one of these keywords, almost all (94%) involved *between* words: words that may indicate an association to a professional, but are often taken to imply causation by non-professionals.

Of the titles containing any of these A-B-C keywords, a majority (61%) involved action verbs or nouns. The second largest category was comparatives (19%). Sufficient, temporal and quasi-casual keywords were involved in the rest (15%) of these titles.

If the sufficient keywords (most of which are verbs) were added to the action verb and noun keywords, they would be in 49% of all the headlines.

4. Analysis

One item to review is non-exclusivity: a title being classified under multiple headings. The number of matches is higher because it includes the quantity words which were not included under Compare previously.

Table 12: Matches by Fields and by Titles

	All	Time	Compare	Sufficient	Action Verb	Other
Fields	1870	65	263	110	752	682
Titles	**1780**	**63**	**251**	**108**	**676**	682
% diff	**5%**	**3%**	**4%**	**2%**	**11%**	0%

In most of the categories, there is no difference in the totals by fields (second row) and by titles (third row). Overall, the difference is small (5%). Only one category, action verbs (11%) is larger. This use of *help* with another action verb may explain this.

5. Words Used by Statisticians

Statistical educators recognize the importance of this distinction between association and causation whenever they tell their students, "*Association is not causation*." For more on the importance of the distinction between association and causation, see McKenzie (2004). A content analysis of several statistics textbooks found these matches:

Textbook	Associated with	Related To	Correlated with	Linked	Predict
Multivariate Analysis	101	52	23	1	86
Statistics by Howell	77	86	39	0	400
Making Sense of Data	65	12	7	4	60

Table 13: Association Keywords: Usage by Statisticians

Note the use of intransitive verbs like *associated with, related to* or *correlated with*, but little use of *linked to*. Note the use of *predict* – generally used as a transitive verb.

Textbook	Effect	Effect of	Resulted in	Results in	Result of	Caus*
Multivariate Analysis	87	28	0	6	0	6
Statistics by Howell	264	89	3	10	14	10
Making Sense of Data	303	92	1	12	12	259

Table 14: Causation Keywords: Usage by Statisticians

Note the three-to-one ratio between *effect* and *effect of.* Some of the common *effect* phrases used by statisticians include

- *Effect* as a noun being modified by *confounding, fixed, direct, total, specific, stronger, spurious, synergistic, joint, combined, main, real, postulated, multiplicative, modifying, appreciable, cohort, generation, undesirable, preventative, beneficial* and *protective*.
- *Effect* as an adjective: *effect size, effect modification* and *effect modifier.*

It seems that statisticians often use *effect* in a formal or mathematical sense (e.g., the effect of increasing X was to increase Y) without implying causation. This would make *effect* a *between* word for statisticians. It may be difficult for statisticians to complain about journalists' use of action verbs to describe associations when statisticians use *effect* without implying causation.

Note the much higher usage of *cause* words in *Making Sense of Data* than in the other two books. The first two focus more on formal statistics while the latter focuses on epidemiology – the search for causes.

Textbook	Due to	Because of	Follow*	After	*crease*	Helps
Multivariate Analysis	32	14	108	62	161	4
Statistics by Howell	52	36	184	100	202	5
Making Sense of Data	53	24	136	88	171	0

Table 15: Between Keywords: Usage by Statisticians

While *helps* was the most common action word used by journalists, it is seldom if ever used by statisticians. Statisticians use comparatives such as *increase* or *decrease.* It is interesting to note that statisticians use temporal words (follow, after) much more than logical connects (*due to, because of*). Appendix C contains more detail on the choice of keywords.

6. Conclusion

To be statistically literate, one must be able to read and interpret the titles and headlines of stories in the everyday media. Doing so is not easy when the headlines use words that statistically illiterate readers read as asserting causation.

Table 16: A-B-C Distribution in Headlines

	ALL	Associate	Between	Cause
Count	1,420	**46**	**1,329**	**43**
% of ABC	100%	3%	94%	3%
% of All	71%	2%	67%	2%

Assuming that the keywords over-estimate the true prevalence of the semantic conditions, three conclusions seem solid. First, less than 5% of titles have keywords clearly indicating association or causation. Second, over half the titles involve keywords that indicate a relationship that is somewhere "between" association and causation. Third, action verbs are the most common means journalists use to describe an association.

If citizens are to become statistically literate, they must be able to read and interpret statements that use these "between" words to indicate the type of association.

Statistically literate readers should recognize that (1) action verbs, comparatives and temporal, connection and logical keywords in news stories do not assert causation, and (2) the word *effect* when used by statisticians does not assert causation.

Gal (2004) noted that, *Unfortunately, no comparative analysis has so far systematically mapped the types and relative prevalence of statistical and probabilistic concepts and topics across the full range of statistically-related messages or situations that adults may encounter and have to manage in any particular society. Hence, no consensus exists on a basis for determining the statistical demands of common media-based messages*.

Since Gal identified the need for a research base from which statistical educators could establish the statistical literacy requirements for adults in various social situations, it seems appropriate to use his name to describe this project as the Iddo Gal Statistical Literacy Research Base Project: to build a research base that identifies the prevalence of statistical terms and thinking in the full range of domains and environments where adults function. This paper and that by Raymond and Schield (2008) are a contribution toward that research base.

7. Recommendations

Recommendations are of two kinds: first, what teachers might do based on this study and second, how this study might be expanded.

Teachers who want their students to understand and appreciate the difference between association and causation might use news headlines to help students deal with this issue in specific contexts. Helping students see what is most disputable in a given statement is a skill that requires practice to acquire and maintain. But without such exposure, students may be able to repeat the claim that "association is not causation" without being able to apply it in their everyday life.

This study might be expanded in many ways. It could have a greater range of news stories and relevant terms. *Forecast* might be included as a lemma for association. The action verbs should be extended to include generatives such as *generate, yield* and *produce*. In examining modal-auxiliaries, *might help* and *could help* should be included.

An expanded study would

- Identify whether the headline was a question since questions don't assert. Consider this headline: *Could lice prevent asthma?* It includes the keyword *prevent* so the rules in this paper would classify it as *sufficient*. But since the title is a question, counting this as an instance of a sufficient claim inflates that category.
- Identify how often modals are used to qualify a claim. Consider this title: *Anger really can kill you: study*. While the verb *kill* indicates sufficiency, the modifier *can* pulls some – if not most – of the assertiveness from the claim.
- Minimize – if not eliminate – one title falling under multiple categories. Consider this title: *Medicare Spending Caps Cause Seniors to Stop Meds.* The two keywords *cause* and *stops* allow this title to be double counted.
- Focus on words having identical noun and verb forms such as *lead* (e.g., *Lead kills*) or identical verb and adjective forms such as *leading* (e.g., *Leading killer*).
- Focus on the nature of the terms surrounding the main verb to distinguish natural causation (e.g., *Eating nuts cuts health risks*) from human causation (e.g., *Lower wages cuts worker productivity*.)
- Compare the words used to describe the association in the original source document or journal article with those in the media summary, the press release, the news story and the news story headline.

Some of these might be done with better computer programming. But human assessment may be required in order to make significant improvements.

Acknowledgments

This paper is part of the W. M. Keck Statistical Literacy project: a project "to support the development of statistical literacy as an interdisciplinary curriculum in the liberal arts."

References

Gal, Iddo (2002). Adult's Statistical Literacy: Meanings, Components, Responsibilities. International Statistical Review 2002 70, 1, p 1-51. International Statistical Institute (ISI).13. Copy at www.stat.auckland.ac.nz/~iase/cblumberg/gal.pdf

Goldin, Rebecca (2009). Spinning Heads and Spinning News: How a Lack of Statistical Proficiency Affects Media Coverage. 2009 *ASA Proceedings of the Section on Statistical Education*. See www.StatLit.org/pdf/2009GoldineASA.pdf.

McKenzie, John D., Jr. (2004). Conveying the Core Concepts. 2004 *American Statistical Association Proceedings of the Section on Statistical Education.* [CD-ROM] P. 2755-2757. See www.StatLit.org/pdf/2004McKenzieASA.pdf.

Raymond, Robert and Milo Schield (2008). Numbers in the News: A Survey, 2008 *American Statistical Association Proceedings of the Section on Statistical Education.* [CD-ROM] P. 2848-2855. See www.StatLit.org/pdf/2008RaymondSchieldASA.pdf.

Schield, Milo (2005). Quantity Words without Numbers. 2005 QR Conference at Carleton College. See www.StatLit.org/pdf/2005SchieldCarleton.pdf.

Wilper A.P., S. Woodlander, K. E. Lasser, D. McCormick, D. H. Bor and D. U. Himmelstein (2009). Health Insurance and Mortality in US adults. *American Journal of Public Health* 10.2105. See www.ajph.org/cgi/content/abstract/AJPH.2008.157685v1

Appendix A: Query Results – Summary

A total of 727 fields were matched against the 2,000 titles. This appendix contains the counts for groups of related fields along with the individual total for each field including those words that had no matches in any of the 2,000 titles.

Query1Overview Subtotals [240 fields]:

- Study subtotal (350): Study (338), studies (12)
- Report-Survey-Poll subtotal (57): Report (30), survey (17), poll (5), reports (3), surveys (2), polls (0)
- Science-Research subtotal (39): Scientists (16), researchers (10), research (7), science (6)
- Ratios subtotal (333): Risk (207), rate (31), risks (27), rates (25), likely (23), percent (8), odds (6), chances (3), share (2), chance (1), fraction (0), incidence (0), likelihood (0), percentage (0), prevalence (0), prevalent (0), probability (0), probable (0) Note: *percentage* was inadvertently omitted but had no matches.
- Cause-Effect-Result subtotal (43): Cause (13), effects (8), results (7), causes (5), effect (5), caused (3), causal (1), causing (1), effected (0), effecting (0), result (0), resulted (0), resulting (0)
- Association-Prediction-Relation Subtotal (46). Predict (18), related (12), predicts (11), associated (4), predicting (1), associate (0), associates (0), association (0), correlate (0), correlated (0), correlates (0), correlating (0), correlation (0), predicted (0), relate (0), relates (0), relating (0), relation (0)
- Link-Connect subtotal (106): Linked (71), link (20), links (15), connect (0), connected (0), connecting (0), connects (0), linking (0)

- Factor subtotal (7): factor (4), factors (3)
- Temporal Relation subtotal (65): After (23), lead (19), leads (9), before (6), follow (4), following (2), leading (2), followed (0), follows (0), led (0), precede (0), preceded (0), precedes (0), preceding (0), subsequent (0), subsequently (0)
- Logical Relation Subtotal (21): As (16), due to (3), given (2), because (0), for lack of (0), for want of (0), on account of (0), owing to (0), responsible (0), thus (0), whereas (0)
- Action Noun Subtotal (7): Killer (4), booster (1), inhibitor (1), reducer (1), accelerator (0), activator (0), affecter (0), beater (0), bender (0), blocker (0), changer (0), contributor (0), controller (0), cutter (0), deactivator (0), delayer (0), doer (0), dropper (0), easer (0), ender (0), energizer (0), enhancer (0), enlarger (0), extender (0), fighter (0), generator (0), getter (0), grower (0), helper (0), hitter (0), hurter (0), improver (0), increaser (0), influencer (0), keeper (0), linker (0), maker (0), modifier (0), preventer (0), producer (0), raiser (0), repressor (0), speeder (0), stimulator (0), stopper (0), suppressor (0), thwarter (0), upper (0)
- Modals Subtotal (373): May (263), can (51), could (36), might (15), will (6), should (2), shall (0), would (0)
- Modal-Auxiliary Subtotal (33): May help (27), can help (6)
- Reporting Subtotal (30): Say (14), find (5), hope (5), think (4), believe (2), know (0)
- Comparative adjectives Subtotal (208): More (97), better (24), higher (23), less (23), earlier (13), longer (9), greater (4), faster (3), bigger (2), shorter (2), taller (2), thinner (2), wider (2), fatter (1), larger (1), quicker (0), slower (0), smaller (0), sooner (0)
- Comparative-Than Subtotal (7):More Than (5), less than (2)
- Superlative Subtotal (53): First (9), last (5), most (28), least (0), best (2), worst (1), largest (0), highest (2), lowest (2), maximum (0), minimum (0), biggest (1), smallest (0), greatest (0), most popular (0), most likely (0), least likely (0), most common (1), top priority (1), first rate (0), tallest (1), shortest (0), littlest (0)
- Comparative Adverb subtotal (2): Largely (1), mostly (1), generally (0), mainly (0)
- Majority-Minority subtotal (2): Majority (2), minority (1)
- Many-Often subtotal (79): Often (23), many (22), common (18), much (6), major (4), significant (2), a lot (1), excess (1), huge (1), massive (1), abundance (0), additional (0), amazing (0), awesome (0), extensive (0), frequently (0), giant (0), incredible (0), minor (0), numerous (0), plenty (0), substantial (0), surplus (0), vast (0)

Query1Sufficient Subtotals (110 matches of 53 lemmas representing 211 fields):
Prevent (28), stop (13), end (9), kill (9), avoid (7), ban (7), quit (7), cure (6), ward off (5), block (3), hinder (3), start (3), counter (2), remove (2), stave off (2), cancel (1), eliminate (1), interfere (1), quash (1), abolish (0), abort (0), annul (0), avert (0), counteract (0), countermand (0), defeat (0), delete (0), deter (0), discontinue (0), eradicate (0), expunge (0), fend off (0), forestall (0), frustrate (0), impede (0), interpose (0), interrupt (0), intervene (0), invalidate (0), negate (0), neutralize (0), nullify (0), obliterate (0), obviate (0), overcome (0), override (0), overrule (0), preclude (0), prohibit (0), restrain (0), rule out (0), subdue (0), terminate (0).

Note: the words shown here are lemmas each of which stands for all forms of the verb.

Query2: Action Verbs Subtotals (752 matches of 71 lemmas representing 276 fields):
Help (91), up (49), cut (44), raise (41), boost (38), get (33), increase (28), lower (24), affect (23), fight (23), do (22), change (21), reduce (20), make (19), protect (17), drop (16), treat (16), ease (15), double (14), grow (14), improve (14), keep (12), trigger (11), control (10), delay (10), curb (9), hit (9), slow (9), extend (8), gain (8), speed (8), beat (6), hurt (6), shield (6), supplement (6), add (5), fuel (5), influence (5), spread (5), spur (5), impact (3), produce (3), decrease (2), enhance (2), lessen (2), prolong (2), shorten (2), shrink (2), suppress (2), widen (2), bend (1), contribute (1), expand (1), stimulate (1), thwart (1), accelerate (0), activate (0), amplify (0), augment (0), deactivate (0), energize (0), enlarge (0), generate (0), inhibit (0), modify (0), multiply (0), narrow (0), repress (0), soar (0), subtract (0), wards (0).
Note: the words shown here are lemmas each of which stands for all forms of the verb.

Appendix B: Data Processing Details

All of the data obtained are based on computerized matches. Titles were entered into an Excel file and manipulated in an Access database. The key to matching is the quality of the matching criteria. Here is the SQL syntax for the select criteria for the word "study":

Study: IIf([Title] Like "*[!A-Z]study[!A-Z]*",1,0)

This syntax instructs the program to look for a match in the Title field. It requires a non-alphabetic character to precede and follow the word or phrase being selected. It allows for optional additional characters on either end.

To avoid problems with words appearing at the beginning or end of the title field, a period and blank space were inserted before the first character of the title and a blank space and period were added after the last character of the title. This Excel command was used: = ". "&<title>&" ."

To document the query used, open the query in design view. Right mouse and select SQL view. Copy the SQL commands to a MS Word document. To show one query per line, do a global replace of "IIF" with "^pIIF.

To document the fields used in a query, select the top row of a query output. In an Excel spreadsheet, paste the results in the top two lines (the headers come along automatically). Re-select the area just pasted and then Edit/Special onto a cell below this area. Select *Values* and *Transpose*.

The key to good matching depends on the quality and quantity of the words presented for matching. For example if the verbs *cut* and *up* had been omitted from the list of action verbs, this would eliminate over a tenth of the matches found in this category. Unfortunately, we are unaware of any list of action verbs, so we have no way to measure completeness in these lists.

For more details on data processing, contact the first author.

Appendix C: Selection Details for Statistics Textbooks

Note that *follow* with no space afterward includes *follows, followed* and *following*. Note that *creas* with no space before or after includes *increase, increases, increased* and *increasing* as well as *decrease, decreases, decreased* and *decreasing*. Note that *help* with no space afterwards includes *helps, helped, helping* and *helpful*.

The word *effect* was found by adding a space at the end to avoid counting *effective* or *effectiveness*. The phrases *results in* and *results of* were chosen to avoid including *results* as a noun. The phrase *related to* was found by adding a space at the beginning to avoid counting *correlated to.* The phrase *effect of* was used to avoid counting *effect* as a noun as in *effect size*, or *effective*. The phrase *Caus** with a space at the beginning to eliminate *because* includes *cause, causes, caused, causing, causation, causal* and *causally*. Note: in *Making Sense of Data* only pages 160 on were counted; 18 headings with "Causes and Effects" were ignored.