STATISTICAL LITERACY: DIFFICULTIES IN DESCRIBING AND COMPARING RATES AND PERCENTAGES

Milo Schield, Augsburg College
Dept. of Business & MIS. 2211 Riverside Drive. Minneapolis, MN 55454

Keywords: Epistemology, Education, Grammar

Abstract: A basic goal of statistical literacy is to construct readily understandable ratio-based *comparisons* that follow directly from data, take into account multiple factors and can support arguments about causation. College students have considerable difficulty constructing such comparisons using rates and percentages. This paper asserts that the main cause of student difficulties is the combination of complexity, subtlety and ambiguity. Complexity is the dominant source of difficulty. Indications of complexity include the unique grammars associated with the three kinds of arithmetic comparisons and four families of named ratios: *ratio, percentage*, *rate* and *chance*. To add to the complexity there are two ways of using *percent*, three ways of using *percentage* and seven ways of using *rate* in descriptions. For each of the 10 ways of using percentage or rate descriptions, there is a corresponding comparison. And there are four ways of using *likely* in comparisons. Examples of subtlety and ambiguity are presented. Subtle differences in syntax (grammar) are shown to cause significant difference in semantics (meaning). The defining and comparing of rates and percentages is also difficult because it includes most of the mathematical difficulty of using English to describe the concepts of variable, function, multivariate function and partial derivative.

1. RATES AND PERCENTAGES

Rates and percentages can be very difficult to describe and compare. The paper analyzes some difficulties in terms of complexity, subtlety and ambiguity.

Rates and percentages are quite common. In the 1997 US Statistical Abstract, about 30% of the tables contain percentages, 10% contain rates and 10% contain other statistics (mean, median, percentile).

Rates and percentages are extremely valuable. They take into account – control for – the *size* of a related group or whole and they can be used in comparisons, thereby controlling for *multiple* factors. By controlling for multiple factors, rates and percentages have the same kind of power and mental complexity, as do partial derivatives. (Schield, 2000)

Describing and comparing rates and percentages is a key element in statistical literacy where a major goal is to construct readily understandable ratio-based *comparisons* that follow directly from data, that take into account multiple factors, and can support arguments about causation. For more detail, see Schield 1999a.

2. COMPLEXITY OF RATIOS

Ratios are interpreted differently than counts. 'The number of men who smoke' is the same as 'the number of men among smokers.' But 'the percentage of men who smoke' is different from 'the percentage of men among smokers.' Ratios are more difficult to parse and interpret than are counts or measures.

Ratios per 100 (percentages) are typically described using *percent*. The percent ratio may be a *part-whole* ratio: "Each of the four boys received 25% of the inheritance" where 25% is *part* divided by *whole*. The ratio may be a *percent change* ratio: "Sales went from 10 million to 12 million: a 20% increase."

Rates typically express the denominator using *per*. Some rates are part-whole rates: "The death *rate* in motor vehicle accidents was 16 per 100,000 population." Others are not: "The death rate in motor vehicle accidents was 1.9 per 100 million vehicle miles." Some rates involve a time-interval: "The *rate* of speed was 100 km per hour; the interest rate was 18% per year." Others are not: "The unemployment *rate* is 5%."

Some ratios may be described using chance or probability. "The *chance* of rolling two sixes is one in 12."

At this level, describing ratios using *percent, rate* and *chance* may not seem very difficult. But difficulties do lurk just beneath the surface.

The irreversibility of the part and whole is somewhat subtle. Students wonder if "X% of A are B," then isn't it true that "X% of B are A?" Certainly if *some* S are P then *some* P are S. But percent descriptions of percentages are irreversible in the same way that *all* statements are irreversible. "All S are P" doesn't imply that "All P are S." (See *conversion* in logic for more on this topic.)

The relation between *part* and *whole* when naming categories of things is somewhat subtle. Suppose that "60% of SMOKERS are males." *Smokers* is the whole and *males* is the part. In arithmetic, a part-whole ratio always has the smaller number on top. So how can *males* be the part (on top) if *males* is a bigger group than *smokers*. To say we want just the "*part* of the *part* within the *whole*" is even subtler – if not confusing.

The concept of a *rate* is extremely ambiguous. Rates can be time-only rates (e.g., velocity), time-independent rates (e.g., prevalence) or time-scaled rates (e.g., incidence). This ambiguity in *rate* can allow opposing claims to appear valid. Consider these examples.

- Suppose someone asserts, "Crime is up." Someone else says, "Crime is down." The first means the number of crimes per year is increasing (a time-only rate). This increase in crimes may be due to a large increase in population. The second speaker means the crime rate (the number of crimes per year per 1,000 population) is decreasing (a time-scaled rate). Both statements are true due to an ambiguity – they are measuring *crime* using different kinds of rates.
- Suppose someone says "the marriage rate has recently doubled" meaning the number of marriages (or even the *incidence* of marriages) in the past month has doubled. Someone else says, "the marriage rate is fairly stable," meaning the *prevalence* of married people in the population is fairly stable. Again, both statements are true due to an ambiguity in *rate*.

3. COMPARISONS

Arithmetic comparisons are somewhat complex. There are three different kinds of comparisons between two values: a test value and a base value of comparison. (Schield, 1999b)

COMPARISON	GRAMMATICAL FORM
Simple Difference:	Test is X *more than* Base.
Simple Ratio:	Test is X *times as much as* Base.
	Test is X% [*times*] *as much as* Base.
Relative Difference:	Test is X*% more/less than* Base.
	Test is X *times more/less than* Base.

The use of *times* in arithmetic comparisons is subtly equivocal. The keyword *times* is found in both the simple ratio and the relative difference comparisons. Students presume that *times* always indicates the simple ratio while the phrase *more than (less than)* merely indicates whether the ratio is larger or smaller than one. But this presumption leads them to view relative differences as simple ratios – a fatal error.

Percentage points are fairly subtle. The difference between two percentages is not measured in *percent* but in *percentage points*. This is an exception to the rule that units remain unchanged in addition and subtraction.

4. PERCENT VERSUS PERCENTAGE

The distinction between *percent* and *percentage* is extremely subtle! *Percentage* indicates a part-whole ratio. *Percents* are the units of the ratio. *Percents* are to *percentage* as *volts* are to *voltage*, as *inches* are to *height* or *pounds* are to *weight*. [1] However percent language can describe a part-whole ratio as can *percentage*.

Part-whole ratios are described using clauses in percent language while using phrases in percentage language.

[1] Ohm's law is improperly stated as *volts* = *amps* x *ohms*. Properly stated, Ohms Law is *voltage* = *amperage* x *ohmage.*

1. X% of *whole* are *part*.[2,3]
2. Among *whole,* X% are *part*.

P1. The *percentage* of *whole* who are *part* is X%.
P2 Among *whole* the *percentage* who are *part* is X%.
P3. Among *whole* the *percentage* of *part* is X%.

The complexity increases here. Now we have two ways to use *percent* and three ways to use *percentage*.

We can say, "26% (unit) of very low-IQ adults live in poverty." (E.g., Some very low-IQ adults live in poverty.) We cannot say, "26 *percentage* (ratio) of very low-IQ adults live in poverty."

We can say, "The *percentage* (the ratio) of very low IQ adults who live in poverty is 26%." We cannot say, "The *percent* (the unit of measure) of very low-IQ adults who live in poverty is 26%." We cannot say, "Very low-IQ adults who live in poverty are 26%."

A subtle equivocation is in using the preposition *of* to indicate either whole or part depending on the context. E.g., If the percentage *of* low-IQ adults <u>who live in poverty</u> is 26%, then the percentage *of* <u>poverty</u> AMONG low-IQ adults is 26%. In the first case, *of* indicates the whole; in the second, *of* indicates the part. A small change in syntax produces a big change in semantics.

Since *percentage* language is more complex than *percent* language, one might wonder why it is needed. The language of *percentage* is needed to form comparisons of percents – a key goal of statistical literacy.

There are three matching common-part comparisons.

CP1. The *percentage* of *test-whole* who are *part* is <compare> as/than that of *base-whole*.
CP2 The *percentage* who are *part* is <compare> among *test-whole* as/than [among] *base-whole*.
CP3. The *percentage* of *part* is <compare> among *test-whole* as/than [among] *base-whole*.

Using *percent*, we cannot generally compare two percentages described in words.[4] E.g., In the US in 1995, 48% of pregnancies for unmarried women were aborted (8% for married women). Yes, "48% is six times as much as 8%." But we can't say, "The *percent* (unit) of pregnancies that were aborted was six times as great for unmarried women as for married women."

At this point, we have three ways to form arithmetic comparisons, five ways to describe percents and three ways to compare percents. The subtlety, the ambiguity and the complexity are just beginning to build.

[2] An equivalent (half) may be used instead of a percent (50%).
[3] In saying, "The X% of W who are Y are P," Y is a whole or whole delimiter. The relative clause has no separate status.
[4] A valid but uncommon percent-based comparison is, "A warship <u>hits</u> with 30% more of its shots than a submarine."

5. PHRASE-BASED RATE GRAMMAR

It might seem that descriptions using *rate* are fairly straightforward. But in fact, *rate* has more ways of describing a ratio than does *percentage*. Rate-based descriptions have three simple phrase-based forms.[5]

R1. The rate of *part* is N per D *whole* [among *whole*].
R2. The *part* rate is N per D *whole* [among *whole*.]
R3. The *part* rate of *whole* is N per D [among *whole*.]

For example consider these US rates from 1995:
R1. "The rate of death was 9.1 per 1,000 MALES."
R2. "The death rate was 9.1 per 1,000 MALES."
R3. "The death rate of MALES was 9.1 per 1,000."

In R1, the phrase *rate of* introduces the part whereas in R3, the phrase *rate of* introduces the whole. Adding just one word modifying *rate* changes '*rate of*' from introducing the part to introducing the whole. Here is another example of subtlety: a small change in syntax (grammar) creates a big change in semantics (meaning).

At this point, *rate* descriptions have the same complexity, as *percentage* descriptions. The preposition *of* can introduce either the part or the whole.

One additional subtlety is modifying *rate* with a countable, but intending the modifier to delimit the whole as an implicit possessive. E.g., "The teen rate of death."

Now consider forming "readily understandable comparisons." Let's compare the death rates of US adults age 25-34 in 1995. The death rate was 205 per 100,000 US males (78 per 100,000 US females). Now 205 per 100,000 is 2.6 times as much as 78 per 100,000.

Matching common-part comparisons are:
CR1 The rate of *part* is {compare} among *test-whole* as/than among *base-whole*.
CR2 The *part* rate is {compare} among *test-whole* as/than among *base-whole*.
CR3 The *part* rate of *test-whole* is {compare} that of *base-whole*.

The first difficulty is recognizing that common terms cannot be divided out (cancelled) in a ratio comparison. We can cancel the common numbers 'per 100,000' and say 205 is 2.6 times as much as 78. But we cannot cancel the common terms such as 'US age 25-34' anymore than we can cancel 'death'.

News stories sometimes omit a common whole and that can be 'fatal.' In 1996, the accident rate *per 1,000 miles* of road was 35 in Hawaii and 7 in Arkansas. Thus it seems that the accident rate is 5 times as high in Hawaii as in Arkansas. But in 1996, the accident rate *per 100,000* vehicles was 18 in Hawaii and 36 in Arkansas. Thus it seems that the accident rate was 2 times as high in Arkansas as in Hawaii. So is the accident rate higher or lower in Hawaii than in Arkansas? The answer is, "We can't say." The appropriate whole was omitted so both comparisons are ambiguous.

Proper comparisons would include the missing whole. The accident rate *per mile of road* was 5 times as high in Hawaii as in Arkansas, but was only half as high *per vehicle* in Hawaii as in Arkansas.

There are two more phrase-based descriptions using *rate*. These locate the part inside a subordinate clause.

R4. The rate at which *part* occurs among *whole* was ...
R5. The rate at which *whole* were *part* was ...

There are matching common-part comparisons.
CR4. The rate at which part occurs is {compare} among TEST-WHOLE as/than [that] among BASE-WHOLE.
CR5. The rate at which TEST-WHOLE is *part* is {compare} as/than that at which BASE-WHOLE is *part*.

In summary, there are five phrase-based descriptions using *rate*, *rate of* can introduce either the part or the whole, common amounts can be cancelled but not common parts or wholes, and there are five rate-based comparisons corresponding to the five descriptions.

6. CLAUSE-BASED RATE GRAMMAR

Rate-based descriptions have two clause-based forms. The keyword *rate* is located in a prepositional phrase.

R6. *Part* (noun) occurred at a rate of N per D *whole*.
R7. *Whole* (noun) *part* (verb) at a rate of N per D.

Consider the 1995 death rate for US adults age 25-34.[6]
R6. Death occurs for MEN at a rate of 205 per 100,000.
R7. WOMEN died at a rate of 78 per 100,000.

There are corresponding common-part comparisons.
CR6. Death occurred at 2.6 times the rate among MEN as [[that] among] WOMEN.
CR7. Men died at a rate which was 2.6 times [as much as] the rate at which women died.[7] MEN died at 2.6 times the rate at which WOMEN died.[8]

Rate language is complex! Now we have seven different ways of describing ratios using rate. Each description has a corresponding comparison.

5 The 'rate of N parts per D wholes' reduces to R1 when the numeric values of N and D are removed.

6 Table 132 in the 1998 US Statistical Abstract
7 The base locates the whole-part subject-predicate in a relative clause (R5) to avoid having two main clauses in a sentence without a conjunction.
8 The base indicator is unstated, 6 is 3 times [as much as] 2.

7. CHANCE

Chance, risk and probability are most commonly used in statistics texts to describe rates and percentages. Grammatically, the syntax of chance is quite straightforward. Chance may be followed by either a subordinate clause (the chance *that a black male will* die), the preposition 'of' followed by a noun (e.g., chance *of* *death*), or the preposition 'of' followed by a clause or gerund (chance of *a man* *dying*, chance of *dying*, etc.)

Chance-based comparisons are fairly straightforward if the whole is contained in a separate phrase. In the US in 1995, for females age 30-34, the chance of being killed was 5 times as great for blacks as for whites.

But in using chance language we have a serious inferential problem. Chance language originated in games of chance where the underlying probabilities were supposedly constant – time independent. The chance of two heads in two flips of a fair coin **IS** 25%.

The problem comes in applying time-independent chance language to time-dependent ratios. We might say, "In the US for females age 30-34, the chance of being killed **IS** 5 times as great for blacks as for whites." But using the indefinite present (is) opens the door to an inferential question. How did we get from past facts to inferences about the present and/or future?

If we use a time-dependent form, we must use the past tense: "In the US in 1995 among black females age 30 - 34, the chance of being killed **WAS** 20.6 per 1,000." But describing historical facts using chance is counter-intuitive for readers who presume chance is always future oriented.[9] Recall that a very basic goal of statistical literacy is "to construct readily understandable comparisons of ratios that follow directly from data and that can support arguments about causation."

Given this goal, the use of chance language is generally disallowed in dealing with time-dependent ratios. Statements of chance do not follow *directly* from the data. They require an assumption of time independence: a Bernoulli trial. That assumption of time-independence may be true, but it requires justification.

[9] Frequentist statisticians don't like to use probability to refer to things that could not be otherwise. Thus they don't consider the probability that a given confidence interval contains the population parameter: it either does or does not. Nor do they consider the probability that the alternate hypothesis is true: it either is or is not. Some frequentist statisticians are flexible when dealing with probabilities of winning when the winner is determined but unknown. In the "three-door" problem, they say there is one chance in three of selecting the door with the prize. Although this outcome has the semblance of a future event (discovery), in fact it is a past event. To use chance about the past with a frequentist notion of probability, one must introduce resampling which some find unintuitive.

8. MACRO CONFUSION

Suppose that students understand the grammar *within* each of the four families: percent, percentage, rate and chance. There is still some confusion involving similarities and differences *between* these four different grammars. Consider the following phrases. Are they grammatically appropriate?

Adjective modifier: Unemployment *rate*? Yes but not the unemployment *percent*, unemployment percentage or unemployment *chance*.

Relative clause: 'The *percentage* of the workforce who are unemployed'? Yes, but not 'the *percent* of the workforce who are unemployed', 'the *rate* of the workforce who are unemployed', or 'the *chance* of a worker who is unemployed.'[10] Only in *percentage* language does the relative clause take on a different part-whole role than what it modifies. This is one of the most surprising aspects of the grammar used to describe ratios.

Using *of* for the part: The *rate* of unemployment? The *chance* of unemployment? Yes. The *percentage* of unemployment? Perhaps, but not 'the *percent* of unemployment' or 'the unemployment *rate* of males.'

Each of these little syntax idiosyncrasies increases the complexity involved in this enterprise.

9. READING TABLES

At this point, all the problems have involved decoding statements about rates and percentages. There are other problems in reading the titles, subtitles, and the column and row headings for the ratio in question. Determining whether the whole for a given ratio is a column, row or entire table is not always obvious.

Teaching students how to decode tables takes additional time. That is the subject of another paper. The point here is that decoding tables is another difficulty in reading and comparing rates and percentages. Again, the combination of subtlety and complexity is increasing.

Students need to appreciate the existence of two kinds of comparisons. This need exists in both decoding and creating comparisons. It is more easily presented in terms of decoding. Consider this table.

1990 US Accidental Deaths: Percentage by Race

	---------- RACE ----------		
SEX	White	Non-Whites	All
Males	71%	77%	72%
Females	29%	23%	28%
Total	100%	100%	100%

Mathematically, there is nothing significant about which two percentages we choose compare. They

10 We can say, 'the rate at which the workforce is unemployed' or 'the chance of a worker being unemployed.'

might be in the same row – or in the same column. But in terms of value, there is a big difference. Comparing two percentages in the same row means they have a *common part* and different wholes. Comparing two percentages in the same column means they have a *common whole* with different parts.

A *common part* comparison is much preferred over a *common whole* comparison. The purpose of creating a ratio in the first place was to take into account (control for) different sized groups. Thus a *common-part* comparison is consistent with that goal: it controls for the size of two different groups. A *common-whole* comparison is not consistent: it involves only one whole, so it does little more than provide a ratio of the counts involved in the two parts.

The difficulty here is that students can't readily identify whether a comparison in a table or in words is a *common whole* or *common part* comparison.

10. COMPARISONS USING LIKELY

At this point, students have a very full plate. Two of the bigger problems are (1) the ambiguity of *of* in rate and percentage language, and (2) the ambiguity of *times* in describing either a simple ratio comparison (times as much as) or a relative difference comparison (times more than). (Schield, 1999b)

One solution to the first of the problems is to find a new kind of comparison that totally avoids using *of* to indicate either part or whole. This comparison exists. It is an adverbial comparison that uses words such as *likely, risky/riskier, probable and prevalent,* This adverbial comparison may be the most common technique for comparing rates or percentages. The rules for adverbial comparisons of part-whole ratios are simple.

1. The preposition *among* always introduces a whole.
2. The preposition *to* always introduces a part.
3. The phrase *as X is* (*as is X*) indicates that X is linked to the subject so their part-whole status is the same.
4. There must be at least one whole and one part.

In making these comparisons, the subject can be either a whole or a part. The named ratio keywords (*rate, percentage, chance*) are never included. In making common part comparisons for death in the US in 1995, among those ages 25-34, we can say:

L1. Part as subject: <u>Death</u> is 2.6 times as likely among MEN as [among] WOMEN.

L2 Whole as subject: MEN are 2.6 times as likely <u>to die</u> as are WOMEN.

The two common-whole comparisons are not shown. Adverbial comparisons using *likely, risky, riskier* and *probable* have the same problem as *chance* – a time-independent presumption. Adverbial comparisons using *prevalent* avoid this difficulty.

11. CONCLUSION

Describing and comparing rates and percentages is very difficult. It is not rocket science, but it is not easy.

There are four distinct grammars (percent, percentage, rate and chance) for describing ratios. There are three types of arithmetic comparisons involving five different grammatical forms. Students must deal with the grammatical subtlety of all of this, the ambiguity of *of*, the similarities and differences between the four different sets of rules, the differences between clause-based and phrase-based descriptions and comparisons, the presence of a totally different form of comparison using *likely*, and finally the difficulty of reading tables and graphs to extract the part and whole so one can form accurate descriptions and comparisons. This is a lot!

In summary, the main cause of student difficulties is the combination of complexity, subtlety and ambiguity. *Subtle differences in syntax (grammar) can cause large difference in semantics (meaning).* Many students are not accustomed to this precision in language. Teaching this material is mathematically difficult because it entails most of the difficulty in teaching the concepts of variable¸ function¸ multivariate function, partial derivative and partial correlation. Using English to describe subtle mathematical relationships is very challenging!

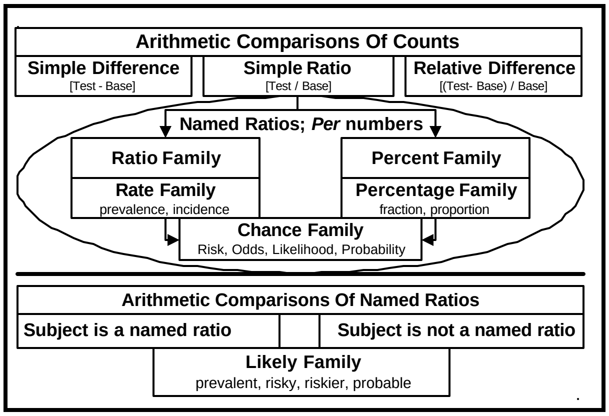
Still, this material must be taught in some detail so that students can achieve a basic goal of statistical literacy: "to construct readily understandable comparisons of ratios that follow directly from data and that can support arguments about causation."

Future steps include the development of teaching materials so this subject can be taught in courses that focus on statistical literacy and quantitative reasoning.

APPENDIX

The classifications and rules mentioned previously are descriptive – not prescriptive. Many are original. In searching widely and diligently, no such classification has been found beyond a very basic level. (See Collins Cobuild, *English Grammar*.) These classifications and rules are ones that make sense. They were tested with hundreds of college students and refined by examining their use in the Harper Collins Cobuild Corpus of English Language. See <u>http://titania.cobuild.collins.co.uk</u>.

Generating descriptions and comparisons following a single template or pattern is both straightforward and limited. Reading and decoding any and all descriptions and comparisons using parsing rules is open-ended. The rules presented here are believed adequate for the patterns shown, but their adequacy for a wider range of patterns is unknown. In short, these structures, classifications and rules are provisional. This figure illustrates a relational diagram of these many concepts:



A. *PER* NUMBERS

The simple ratio comparison of counts is so common that some forms have special grammatical signs. The most common sign is *per* so those special ratios may be designated as *per numbers*. (Tarp, 2000) Per numbers are simple ratios having special grammars such as *per*. *Percent* is the most widely used member of this family.

Per numbers are more than fractions or ratios.[11] *Per numbers* are ratios having a particular context. While 0.6 + 0.6 = 1.2, we do not say that 60% *plus* 60% equals 120%. Having a market share of 60% in each of two territories does not give a market share of 120% in the combined territory.

B. NAMED-RATIOS

Named-ratios are ratios that have their own proper names. Twelve ratios were analyzed.[12] These were grouped, based on their grammars, into four families.[13]

1. ***Ratio***
2. ***Rate****, incidence and prevalence*.
3. ***Percentage,*** *proportion and fraction*
4. ***Chance****, risk, odds, likelihood and probability*.

Although *percent* is not the name of a ratio, it is used to describe a ratio so it is included. *Share* designates a portion – not a proportion – so it is not a named ratio.[14] Since it often describes a ratio, it is also included.

Descriptions are classified as clause-based or phrase-based depending on whether the part and whole require the main clause or not. There are many similarities between comparisons using phrase-based descriptions.

REFERENCES

Collins Cobuild. *English Grammar*. Harper Collins Publ., 1990. [2.221 (part of), 2.257-266 (fractions)]

Durkin, Kevin, and Beatrice Shire (1991). *Language in Mathematical Education*. Open University Press.

Orr, Eleanor Wilson (1987). *Twice as Less*. Norton.

Piaget, Jean (1971). *Understanding Causality*. Norton. Section 12: Linearity, Proportionality & Distributivity.

Schield, Milo (1999a). *Statistical Literacy: Thinking Critically about Statistics*. *Of Significance* journal. The Association of Public Data Users, Volume 1.

Schield, Milo (1999b). *Common Errors in Forming Arithmetic Comparisons*. *Of Significance* journal. The Association of Public Data Users, Volume 1.

Schield, Milo (2000). *Statistical Literacy and Mathematical Reasoning*. University Working Group, International Conference on Mathematics Education (ICME-9), Tokyo.

Tarp, Alan (2000). *Organic Bottom-Up Mathematics - A Three Level Bundling and Totaling Curriculum*. Short Presentation, International Conference on Mathematics Education (ICME-9), Tokyo.

Acknowledgments: To the hundreds of Augsburg students who helped me understand this topic. To Leonard Peikoff for his deft analysis of English grammar. To Harper Collins Cobuild for generating the Cobuild machine-readable corpus of the English Language. To Thomas V.V. Burnham who independently analyzed this material in depth and made major conceptual distinctions and integrations. And to my wife and editor, Linda, who has lived all too many years with my unrelenting desire to understand how one should describe and compare rates and percentages in tables and graphs.

The author can be reached at schield@augsburg.edu. This paper is at www.augsburg.edu/ppages/~schield.

---

11 These can be proper fractions (values between 0 and 1) or improper fractions (30 miles per gallon).

12 These 12 ratios (and their relative usage in the Cobuild Corpus) are chance (27%), rate (22%), *percent of* (16%), risk (12%), *share* (11%), odds (3%), proportion (3%), percentage (2%), ratio (2%), likelihood (1%), fraction (0.9%), probability (0.7%), incidence (0.6%), and prevalence (0.2%). Many uses are ordinal (e.g., a small chance). Uses include both descriptions and comparisons. *Percent of* and *share* are also shown.

13 The distribution of these named ratios by family along with *percent of* and share is chance (44%), rate (23%), *percent of* (16%), share (11%), percentage (6%) and ratio (2%).

14 *Share*, *proportion* and *fraction* are used with percentage grammar at least 10% of the time they appear. But only *proportion* and *fraction* indicate a ratio. *Share* may use the grammar of percentage, but refers to only the part as do *portion, piece*, or *slice.* "A 10% share of the million dollar estate was his" means "A $100,000 share/portion was his." "His 10% share of the million dollar estate was ample" means "His share/portion, which was $100,000, was ample."