Recall some earlier examples...

Montreal Metropolitan Population by knowledge of official language
Data collected by Statistics Canada at the 1996 census:
{numbers rounded, so subtotals do not sum exactly to total]

		English? Yes	No	Total
Français?	Oui	1,634,785	1,309,150	2,943,935
	Non	280,205	63,500	343,705
	Total	1,914,990	1,372,650	3,287,645

Stroke Unit vs. Medical Unit for Acute Stroke in elderly?
Patient status at hospital discharge (BMJ 27 Sept 1980**)**

	indept.	dependent	total no. pts
Stroke Unit	67	34	101
Medical Unit	46	45	91
total	113 (58.9%)	79	192

Bone mineral density and body composition in boys with distal forearm fractures (J Pediatr 2001 Oct;139(4):509-15**)**

		Fracture? Yes	No
Overweight?	Yes	36	14
	No	64	86
		100	100

Pour battre Roy, mieux vaux lancer bas ...
(LA PRESSE, MONTREAL, JEUDI, 21 AVRIL 1994 ... cf. Course 626)

Au cours des vingt matches des séries éliminatoires disputés l'an passé, le Canadien a accordé 51 buts... Des 51 buts alloués par le meilleur gardien au monde..

... ont vu la rondelle pénétrer dans la partie .. du filet

Haut	10	(20%)
Milieu	5	(10%)
Bas	36	(70%)
	51	(100%)

More generally...

Cross-classification of single sample of size n with respect to two characteristics, say A and B ("second" model in M&M p 641)
Test of independence of two characteristics

		B B1	B2	Total
A	A1	n_{11}	n_{12}	n_{A1}
	A2	n_{21}	n_{22}	n_{A2}
		n_{B1}	n_{B2}	n

Cohort Study: Fixed /Variable follow-up. Person (P) or P-Time denominators
(Cross-sectional Study, document states rather than events**)**

	event (or state)	non-event (or state)	Total Persons	or	Total Person-Time
	c="cases" numerator		D= Denominator		D= Denominator
"exposed" (1)	c1		D1		D1
not exposed (0)	c0		D0		D0
	c		D		D

Case-Control Study: person- or person-time "quasi-denominators"

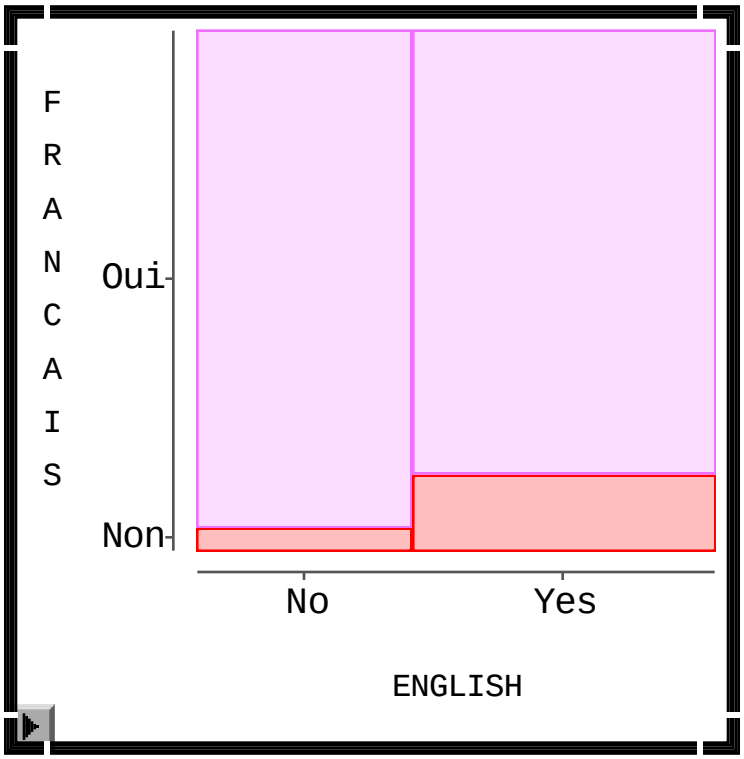
	event (or state)	quasi-denominators	or	quasi-denominators
	"c=cases" numerator	persons		Person-Time
"exposed" (1)	c1	d1		d1
not exposed (0)	c0	d0		d0
		d=sample of D		d=sample of D

e.g. languages in Montreal

create SAS file via Program Editor (Could also type directly into INSIGHT)

```
data sasuser.lang_mtl;
input Francais $ English $ number;
    /* number instead of individual per line */
    /* $ sign after name indicates character variable */
lines;
       Oui     Yes    1634785
       Oui     No     1309150
       Non     Yes     280205
       Non     No       63500
;
run;
```

then.. via SAS INSIGHT **Mosaic plot**



Statistics from 2 x 2 table via SAS Proc FREQ

```
options ls = 75 ps = 50; run;
proc freq data=sasuser.lang_mtl;
     tables Francais * English /
                all cellchi2 expected;
       /* turn on all output */
weight number; /* use weight to indicate "multiples" */
run;
```

TABLE OF FRANCAIS BY ENGLISH

FRANCAIS	ENGLISH		
(Observed) Frequency			... "obs" for short
Expected (under H_0: independence)			... "exp" for short
(Cell Chi-Square)			
Percent			
Row Pct			
Col Pct	No	Yes	Total
Non	63500	280205	343705
	143503	200202	
	44602	31970	
	1.93	8.52	10.45
	18.48	81.52	
	4.63	14.63	
Oui	1309150	1634785	2943935
	1229147	1714788	
	5207.3	3732.5	
	39.82	49.73	89.55
	44.47	55.53	
	95.37	85.37	
Total	1372650	1914990	3287640
	41.75	58.25	100.00

$$143503 = 10.45\% \text{ of } 1372650 = \frac{343705 \times 1372650}{3287640} = \frac{\text{RowTotal} \times \text{ColTotal}}{\text{OverallTotal}}$$

$$44602 = \frac{(\text{observed freq.} - \text{expected freq.})^2}{\text{expected freq.}} = \frac{(63500 - 143503)^2}{143503}$$

Statistic	DF		Value	Prob*
Chi-Square	(2-1) × (2-1) = 1	$\Sigma \frac{(obs - exp)^2}{exp}$	= 85511.881	0.001

DF = "degrees of freedom"; Σ is over all 4 cells

* Clearly, p-values are not relevant here.

e.g. Stroke Unit vs. Medical Unit for Acute Stroke in elderly?
Patient status at hospital discharge (BMJ 27 Sept 1980**)**

	independent.	dependent	total no. pts
Stroke Unit	67 (66.3%)	34	101
Medical Unit	46 (50.5%)	45	91
	113 (58.9%)	79	192

"Expected" numbers **[in bold]** under

H_0: % discharged "independent" unaffected by type of unit

	independent	dependent	total no. pts
Stroke Unit	**59.4** (58.9%)	**41.6**	101
Medical Unit	**54.6** (58.9%)	**37.4**	91
	113 (58.9%)	79	192

¶ $$X^2 = \frac{\{67 - \mathbf{59.4}\}^2}{\mathbf{59.4}} + \frac{\{34 - \mathbf{41.6}\}^2}{\mathbf{41.6}} + \frac{\{46 - \mathbf{54.6}\}^2}{\mathbf{54.6}} + \frac{\{45 - \mathbf{37.4}\}^2}{\mathbf{37.4}}$$

$$= 4.9268$$ [to be referred to χ^2 (1df) distr'n]

Generic formula for X^2

$$X^2 = \sum \frac{\{\text{observed} - \mathbf{expected}\}^2}{\mathbf{expected}}$$ **[Σ over all cells!]**

The **expected** number in cell in to row i and column j is:

$$\frac{\text{total row i} \bullet \text{total column j}}{\text{overall total}}$$

Continuity-corrected X^2 †

$$X^2_c = \sum \frac{\{|\text{observed} - \mathbf{expected}| - 0.5\}^2}{\mathbf{expected}}$$

¶ Use X^2 to refer to the calculated statistic in a sample, χ^2 for distribution.

† (Yates') Continuity-correction is used to reflect the fact that the binomial counts are discrete and that their probabilities are being approximated by intervals (count – 0.5, count + 0.5). The uncorrected X^2 is overly liberal i.e. it produces too large a distribution of discrepancies that is larger than the tabulated distribution... hence the reduction of each absolute deviation | observed – expected | by 0.5.

```
data sasuser.str_unit;
input Unit $ Status $ number;
lines;
       Stroke     Indep    67
       Stroke     Dep      34
       Medical    Indep    46
       Medical    Dep      45
;
run;
proc freq data=sasuser.str_unit; weight number;
tables Unit * Status / chisq cmh relrisk riskdiff nopercent nocol;
run;
```

```
                    UNIT      STATUS
                    Frequency|
                    Row Pct  |Dep      |Indep    |  Total
INDEX Category      Medical  |      45 |      46 |     91
 (see NOTE)                  |  49.45  |  50.55  |
                    Stroke   |      34 |      67 |    101
REFERENCE Category          |  33.66  |  66.34  |
                    Total         79       113       192
```

```
          Statistic                    DF      Value        Prob
          ------------------------------------------------------
          Chi-Square                    1      4.927       0.026
          Continuity Adj. Chi-Square    1      4.296       0.038
          Mantel-Haenszel Chi-Square    1      4.901       0.027
        Fisher's Exact Test Left:0.991 Right:0.019 2-Tail:  0.029
```

Column 2 **Risk Estimates**

	Risk	ASE	95% Conf Bounds (Asymptotic)		95% Conf Bounds (Exact)	
Row 1	0.505	0.052	0.403	0.608	0.399	0.612
Row 2	0.663	0.047	0.571	0.756	0.562	0.754
Total	0.589	0.036	0.519	0.658	0.515	0.659
Row 1 - Row 2	-0.158	0.070	-0.296	-0.020		

NOTE FREQ uses upper row as INDEX category; lower as REFERENCE cat.

Estimates of the **Relative Risk** (Row1/Row2)

Type of Study	Value	95% Confidence Bounds	
Cohort (Col2 Risk)	0.762 (50.55/66.34)	0.596	0.975

cf. z-test for difference of 2 proportions .. in Chapter 8

z = (0.6634 – 0.5054)/sqrt [0.5885 • 0.4115 • (1/101+ 1/91)) = 0.1580/0.0711=2.22

P = Prob[| Z | ≥ 2.22] = 0.026 (2-sided)

$z^2 = 2.22^2 =$ **4.93** (same as X^2 with 1 degree of freedom !)

OTHER EXAMPLES

Note: In the following examples, in order to keep the formulae uncluttered, I have not shown the continuity correction. In some examples, as in the one above, the sample sizes are large and so the continuity correction makes only a small change. In some others, as in the milk immunoglobulin example below, it makes a big difference. However, don't do as I do; do as I say -- use the continuity correction routinely. That way, editors and referees won't accuse you of trying to make your p-values more impressive by not using the correction.

Do infant formula samples shorten the duration of breast-feeding?
Bergevin Y, Dougherty C, Kramer MS. Lancet. 1983 May 21;1(8334):1148-51.

% still breastfeeding at 1 month in RCT which withheld free formula samples [normally given by baby-food companies to mothers leaving hospital with their infants] from a random half of those studied

	breast feeding	not breast feeding	Total mothers
given sample	175 (77%)	52	227
no sample	182 (84%)	35	217
	357 (80.4%)	87	444

"Expected" numbers under
H_0: rate not changed by giving samples

	breast feeding	not breast feeding	Total mothers
given sample	**182.5 (80.4%)**	**44.5**	227
no sample	**174.5 (80.4%)**	**42.5**	217
	357 (**80.4%**)	87	444

$$X^2 = \frac{\{175 - \mathbf{182.5}\}^2}{\mathbf{182.5}} + \frac{\{182 - \mathbf{174.5}\}^2}{\mathbf{174.5}} + \frac{\{52 - \mathbf{44.5}\}^2}{\mathbf{44.5}} + \frac{\{35 - \mathbf{42.5}\}^2}{\mathbf{42.5}}$$

$$= 3.22$$ ["NS" at 0.05 level, even with uncorrected X^2]

$X^2 = 3.22 \leftrightarrow |Z| = \sqrt{3.22} = 1.79$; Prob($|Z| > 1.79$) = 2 × 0.0367 = 0.0734

Protection by milk immunoglobulin concentrate against oral challenge with enterotoxigenic e. coli NEJM May 12 ,1988 p 1240

	developed diarrhea*	did not*	Total subjects
received milk immunoglobin concentrate	0 (**4.5**)	10(**5.5**)	10
received control immunoglobin concentrate	9 (**4.5**)	1(**5.5**)	10
All	9 (45%)	11	20

* **the numbers 4.5, 5.5, 4.5** and **5.5** in **bold** in parentheses in the table are the "**Expected**" numbers calculated on the (null) hypothesis **H_0:** rate not changed by immunoglobin concentrate

$$x^2 = \frac{\{0 - \mathbf{4.5}\}^2}{\mathbf{4.5}} + \frac{\{9 - \mathbf{4.5}\}^2}{\mathbf{4.5}} + \frac{\{10 - \mathbf{5.5}\}^2}{\mathbf{5.5}} + \frac{\{1 - \mathbf{5.5}\}^2}{\mathbf{5.5}}$$

$$= \{4.5\}^2 \left(\frac{1}{\mathbf{4.5}} + \frac{1}{\mathbf{4.5}} + \frac{1}{\mathbf{5.5}} + \frac{1}{\mathbf{5.5}} \right) = \mathbf{16.4}$$

$$\text{Continuity-corrected } x^2 = \frac{\{\,|\,0 - \mathbf{4.5}\,| - 0.5\}^2}{\mathbf{4.5}} + \ldots + =$$

$$= \{4.0\}^2 \left(\frac{1}{\mathbf{4.5}} + \frac{1}{\mathbf{4.5}} + \frac{1}{\mathbf{5.5}} + \frac{1}{\mathbf{5.5}} \right) = \mathbf{12.9}$$

Attack rates of ophthalmia neonatorum among exposed newborns receiving silver nitrate, or tetracycline NEJM March 11 ,1998 653-7

	Silver Nitrate	Tetracycline
exposed to N. gonorrhœa attack rate	5 / 71 (7%)	2 / 66 (3%)
exposed to C. trachomatis attack rate	10 / 99 (10%)	8 / 111 (7%)

- If use x^2 , **it must be based on counts, not on %'s**
- a **short-cut method of calculation** for the 2x2 table with 'generic' entries a, b, c, d, and with row, column and overall totals r1, r2, c1, c2 and N respectively and overall totals is (with stroke data as e.g.):

$$x^2 = \frac{N\ \{\ a\bullet d - b\bullet c\ \}^{\ 2}}{r1\bullet r2\bullet c1\bullet c2} = \frac{192\ \{\ 67\bullet 45 - 34\bullet 46\ \}^{\ 2}}{101\bullet 91\bullet 113\bullet 79} = 4.93$$

For the **continuity corrected** version, the **shortcut formula** is:

$$\frac{N\{\ |\ a\bullet d - b\bullet c\ |\ -\frac{N}{2}\ \}^{\ 2}}{r1\bullet r2\bullet c1\bullet c2} = \frac{192\{\ |67\bullet 45 - 34\bullet 46\ |\ -\ 96\}^{\ 2}}{101\bullet 91\bullet 113\bullet 79} = 4.30$$

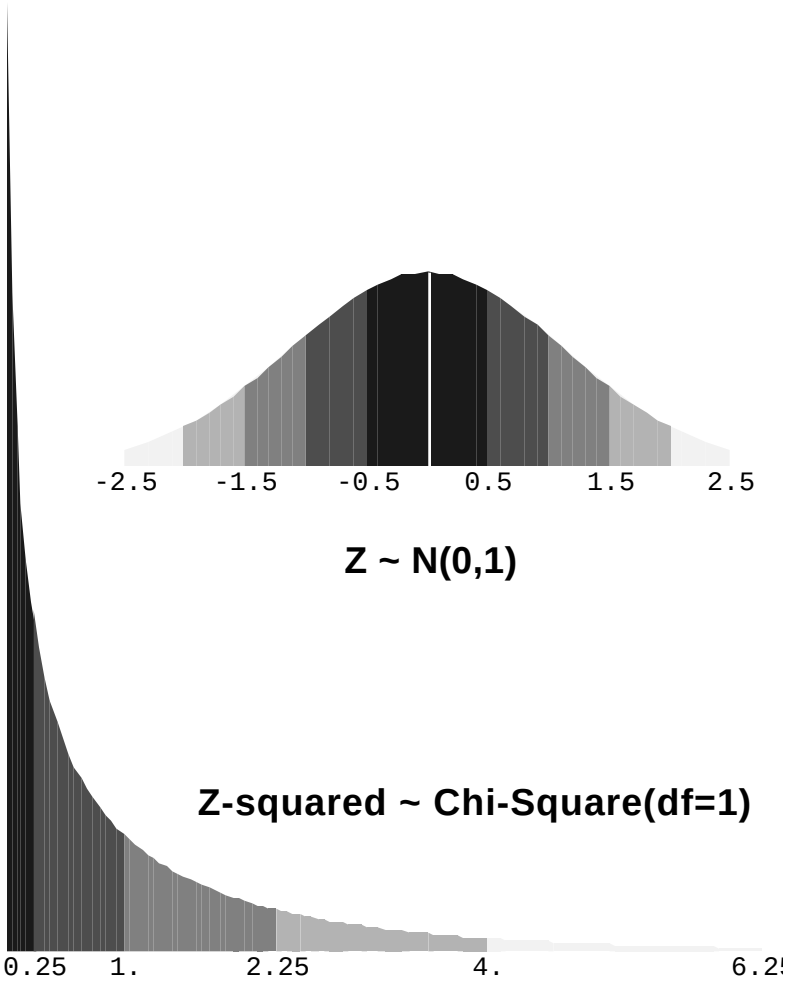
where | x | means 'the absolute value of'.

The formula involves the crossproducts a•d and b•c. If their ratio (empirical **odds ratio**) is 1, their difference is zero. The direction of the difference in proportions is given by the sign of ad – bc.

These formulae avoid fractions -- one doesn't see expectations or deviations, or the magnitude of the difference. Presumably for this reason, some books, such as Norman and Streiner's Statistics: the Bare Essentials, classify x^2 as "non-parametric" or "distribution-free", and so put it in the non-parametric chapter. After their first edition, I pointed out to them that the above use of the chi-square test is as a test of the difference in two binomial proportions.. how much more parametric or distribution-specific can that be? Look up the index in the latest edition to see if my arguments convinced them. The direction is given by the sign of ad – bc.

- The uncorrected version of the 2-sided z-test for comparing two proportions gives the same p-value as the uncorrected version of the x^2 test. One can check that $Z^2 = x^2$. Likewise, the corrected version of the 2-sided z-test for comparing 2 proportions gives the same p-value as the corrected version of the x^2 test.

The chi-square random variable (r. v.) is the square of the N(0,1) r. v. The very high "probability density", and the rapid change in this density, just to the right of $Z^2 = 0$, (cf. diagram) is a result of the Z ->Z^2 transformation. For example, the 3.98% of the "probability mass" between Z=0 and Z=0.1 is transferred to the small interval 0^2 to 0.1^2, or 0 to 0.01, an width of 0.01 (an identical amount gets transferred from the Z interval -0.1 to 0). The 3.94% of the "probability mass" between Z=0.1 and Z=0.2 is transferred to the small interval 0.1^2 to 0.2^2, or 0.01 to 0.04, an width of 0.03. The is an identical transfer from the Z interval -0.2 to -0.1, for an "average" chi-square density of approx. (2 x 0.0394)/0.03 = 2.6. You can track the 'block transfers' using the different shades.



- By construction, the x^2 is a 2-sided test, unless one uses x and refers it to the z - table.
- There are other chi-square distributions (with df > 1). See later.

- If use x^2 , it must be based on <u>all</u> cells, not just on numerators -- unless the more common type of outcome is so much more common that the contribution for these cells is negligible.
- The x^2 test is a large sample test i.e. it is always an approximation . Since x^2 (1 f) is just Z^2, one is no more exact than the other.
- In t-tests, n= 30 is often considered 'large enough' for large-sample procedures -- it depends on skewness of data are & whether the Central Limit Theorem will "Gaussianize" the distribution of the statistic . For 0/1 data, such as those above, the 'real' or 'effective' sample sizes are not the denominators 71 and 66, but rather the numerators 5 and 2 [e.g. changing the 2 to a 1 would make the ratio of attack rates appear twice as good] It doesn't matter if the 5 and 2 came from n's of 710 and 660 or 7100 and 6600 . The 'effective sample size' for binary data is the number of subjects having the less common outcome.
- The "guidelines" (such as they are) about when it is appropriate to use x^2 are based on "**Expected**" numbers **not on the observed numbers**. One quoted rule [often used by computer programs to generate warning messages] is that the expected numbers in most of the cells should exceed 5 for the χ^2 to be accurate. Thus, the 2x2 table on the left below will generate a 'warning' ; that on the right will not.

5	2	1	11
66	64	11	1

- The regular uncorrected x^2 test statistic for a single 2x2 table can be written in a seemingly very different format, as

$$x^2 = \frac{\{\, a - E[a \mid H_0] \,\}^2}{\text{Variance}[\, a \mid H_0 \,]} = \frac{\{\, a - E[a \mid H_0] \,\}^2}{r1 \bullet r2 \bullet c1 \bullet c2 \,/\, N^3}$$

The Variance in the denominator of this statistic can be viewed as arising from a statistical model in which the 2 compared proportions are separate independent random variables; i.e. the 'unconditional' or '2-independent binomials' model .

Just like the formula with 4 O's and 4 E's, this format is not as calculator-friendly as the shortcut (integer-only) one . <u>But</u>... cf Mantel Haenszel.

Mantel-Haenszel Test Statistic for a single 2 x 2 table

Preamble

Some of the paragraphs on the next page more appropriately belong in the notes for Ch 8.2, where Fisher's exact test was introduced , but the reasons become important when we come to using the above formulation of the x^2 test for a <u>single</u> table to combining evidence over <u>several</u> (possibly sparse) 2x2 tables. Cochran first proposed combining evidence from 2 x 2 tables in 1954. His aim was to combine a small number of 'large' tables, and he did not anticipate that this technique could also be used to combine a large number of quite 'small' 2 x 2 tables (each one with quite sparse information) , with the combination of data from n matched pairs as the limiting case. Thus, he was a wee bit careless about variances. It took the now famous Mantel-Haenszel paper of 1959 to make a variance correction that for 'large 'tables was trivial, but for matched pair tables, was critical.

SAS and others rightly acknowledge Cochran's role in the test statistic, calling it the 'Cochran-Mantel-Haenszel' or 'CMH' statistic. (Indeed 'CMH' is the option one uses with the PROC FREQ to obtain the summary measures and the overall test statistic). This formulation is also the one most commonly used for the log-trank test used to compare two survival curves.

Most consider that the biggest legacy of "M-H" paper is to the Mantel-Haenszel summary measure (point estimate) of Odds Ratio. We will come back a little later to this issue of combining data from 2 x 2 tables .

Mantel-Haenszel Test Statistic for a single 2 x 2 table

Preamble: Conditional vs. Unconditional? continued... In the separate binomials model, the only marginal totals that are fixed ahead of time are the two sample sizes. In most instances, this model reflects reality. The only exception I know of is the design exemplified by the psychophysics study of the lady tasting tea . If she is told that there are 4 cups where the tea is poured first, and 4 where it is poured second, then she will arrange her responses so that there are 4 of each. Thus, in this instance, both the row totals and the column totals are "fixed" ahead of time, and so it makes sense that the (frequentist) inference be limited to the (only) 5 possible data tables that have all margins fixed.

This is the statistical model behind Fisher's exact test, and indeed Fisher used the tea-tasting example to explain it. But this test is now used for data situations where one cannot -- at least ahead of time -- consider both sets of marginal totals fixed. For example, in the food sensitivity study in the ch 8.2 notes, from the answers given, it appears that the subjects were not told that there were 3 three injections of extract and nine of diluent, but the authors used the conditional test anyway.

Many of the reasons put forward for using the conditional test based on all margins fixed (i.e. the hypergeometric model, with only one random variable) involve practicality rather than adherence to a coherent set of inferential principles. They mostly have to do with one of the following 'supposed' difficulties (a) using the normal approximation when the expected numbers are low (b) the fact that there are two parameters, but one is only interested in their difference, or ratio, or odds ratio, and so the 'remaining' parameter is just a 'nuisance' (c) how to order or rank the tables by their degree of evidence against H_0. For example, in a 2x2 table with n0=23, and n1=24 (as in the bromocryptine and infertility study) , there are theoretically 24 x 25 = 600 possible tables. However, if one -- after the fact -- restricts the analysis to only those tables where the total number of "successes" is 12 (12 pregnancies) , then there are only 13 possible tables (see notes and Excel spreadsheet for Fisher's exact test). And, by reducing the problem from a 2-dimensional one to a 1-dimensional one, is also becomes possible to more easily rank the tables by their degree of evidence against H0, something that is supposedly more difficult when the tables are simultaneously arrayed along both dimensions. (d) a fourth reason, which I will illustrate with the Marvin Zelen "Marbles in the Folger's Coffee Can" model, is that, after the fact, it is much easier to empirically -- and heuristically -- demonstrate - a low p-value using the single random variable, conditional (hypergeometric), model than it is with the '2-separate binomials' model.

In fact there are many ways to circumvent these objections without having to 'condition' on all margins, and there is still a considerable debate, much of it philosophical, on this 100 years after analyses of 2x2 data were first introduced. However, since we often combine information from data arranged as matched pairs or 'finely stratified ' strata, we do need to consider this one setting where conditioning is the 'right thing to do'. In the example here, there will only be 1 large table, so the difference will not be important. But when we come to matched pairs, the implications are large.

Mantel-Haenszel Test Statistic for a single 2 x 2 table

Details

In the **conditional model**, with **both margins fixed**, there is only one cell entry that can vary independently. Without loss of generality, we focus on the frequency in the 'a' cell. Then, under the null hypothesis,

a ~ Hypergeometric[parameters given by marginal totals]

i.e. by r1=Row1Total, r2, c1=Col1Total, c2, and N=OverallTotal.

Thus

$$\text{Expected value}[\, a \mid H_0 \,] = E[\, a \mid H_0 \,] = \frac{r1 \times c1}{N}$$

$$\text{Variance}_{\text{condn'l}}[\, a \mid H_0 \,] = \{\, r1 \bullet r2 \bullet c1 \bullet c2 \,\} \,/\, \{\, N^2(N-1) \,\}$$

Under the null, the expectation is the same with the conditional as the unconditional models. **Note** however the difference in the variance: under the conditional model it is different, since it uses **$N^2(N-1)$ rather than N^3**, reflecting the different pattern of variation in the frequency in the 'a' (and consequently in the other 3) cell(s) if all margins are fixed (vs. what would happen if the lady were not told "4 1st; 4 2nd").

The test statistic using this conditional variance can be computed as a Z statistic

$$Z = X = \text{"chi"} = \frac{a - E[\, a \mid H_0]}{SD_{\text{condn'l}}\ [\, a \mid H_0]},$$

which has the same form as the critical ratios used in the z-test for proportions or means, or as the more traditional square

$$X^2_{\,MH} \qquad = \qquad \frac{\{\, a - E[\, a \mid H_0]\,\}^2}{\text{Variance}_{\text{condn'l}}\,[\, a \mid H_0]}\ .$$

Example

In our stroke vs. medical unit example above, the marginal totals were r1=101, r2=91, c1=113, and c2=79, so N=192. These yield the "excess in the a cell" of

67 - (101•113)/192 = 67 - 59.44 = 7.56

and conditional variance

$\{\, 101 \bullet 91 \bullet 113 \bullet 79\,\} \;/\; \{\, 192^2(191)\,\} = 11.6528$

giving

$X^2_{\,MH} = 7.56^2 / 11.6528 = 4.901,$

in agreement with the printout from Proc FREQ in SAS.

Note:

The MH test does not use the continuity corrected with the {a - E[a]}. Part of the justification for this is that when the point estimate of the odds ratio falls at the null, i.e. a•d = b•c, so that $E[\, a \mid H_0\,] = a$, it would be good if the test statistic also had a value of zero. A continuity correction would force the test-statistic to have a positive value even when the "observed a" = "expected under the null" !

2x2 *samples reasonably equal in size, two types of outcome common*
e.g. outcomes in trial of stroke vs. medical unit .

	BAD OUTCOME	GOOD OUTCOME	Total persons	or	Total Person-time
sample 1	bad1	good1	n_1		n_1
sample 2	bad2	good2	n_2		n_2
	bad	good	n		n

"**Expected**" numbers of outcomes under **H_0: rates not different (split the events across 2 samples in ratio of $n_1 : n_2$)**

	BAD OUTCOME	GOOD OUTCOME	Total persons or person-time
sample 1	**bad1**	**good1**	n_1
sample 2	**bad2**	**good2**	n_2
	bad	good	n

$$x^2 = \frac{\{bad1 - \mathbf{bad1}\}^2}{\mathbf{bad1}} + \frac{\{good1 - \mathbf{good1}\}^2}{\mathbf{good1}} + \frac{\{bad2 - \mathbf{bad2}\}^2}{\mathbf{bad2}} + \frac{\{good2 - \mathbf{good2}\}^2}{\mathbf{good2}}$$

2x1 *samples large and reasonably equal in size,*
BAD outcome uncommon : e.g. leukemias and breast cancers

	BAD OUTCOME	GOOD OUTCOME	Total persons	or	Total Person-time
sample 1	bad1	MOST	n_1		n_1
sample 2	bad2	MOST	n_2		n_2
	bad	MOST	n		n

$$x^2 \approx \frac{\{bad1 - \mathbf{bad1}\}^2}{\mathbf{bad1}} + \text{minimal contribution} + \frac{\{bad2 - \mathbf{bad2}\}^2}{\mathbf{bad2}} + \text{minimal contribution}$$

"**Expected**" numbers of outcomes under **H_0: rates not different (only need ratio $n_1 : n_2$ to get expected split of BAD events)**

[see A&B §4.10; WE WILL REVISIT THIS 2x1 TABLE , AND THE 1 x 1 TABLE, WHEN COMPUTING EFFECT MEASURES for INCIDENCE RATES]

1x2 1 *sample ; two types of outcome common*
e.g. male and female births with specific timing of conception

	"LESS GOOD OUTCOME	GOOD OUTCOME	Total persons	or	Total Person-time
sample	less_good	good	n		n

$$x^2 = \frac{\{less\_good - \mathbf{less\_good}\}^2}{\mathbf{less\_good}} + \text{minimal contribution} + \frac{\{good - \mathbf{good}\}^2}{\mathbf{good}} + \text{minimal contribution}$$

"**Expected**" numbers of outcomes under **H_0: rate not different from EXTERNAL rate (use EXTERNAL rate, based on LARGE amount of data (e.g. national rates), to calculate the expected split of events). If use internal comparison, then we have full 2 x 2 table.**

1x1 *1 large sample , BAD outcome uncommon*
*e.g. 78 cancers observed in Alberta study, **83.5 expected***

	BAD OUTCOME	GOOD OUTCOME	Total persons	or	Total Person-time
sample	bad	MOST	n		n

$$x^2 \approx \frac{\{bad - \mathbf{bad}\}^2}{\mathbf{bad}}$$

"**Expected**" number of outcomes under **H_0: rate not different from EXTERNAL rate (use EXTERNAL rate to calculate expected number of BAD events)**

This $x^2 = \frac{\{observed - \mathbf{expected}\}^2}{\mathbf{expected}}$ **is equivalent to the large sample approximation to the Poisson distribution** [A&B §4.10]

i.e. $z = \frac{observed - \mathbf{expected}}{\sqrt{\mathbf{expected}}}$ so that

$$z^2 = \frac{\{observed - \mathbf{expected}\}^2}{\mathbf{expected}} = x^2$$

† This terminology is my own: don't try it out on an editor!

e.g. *Development of leukemia during a 6-year period following drug-rx for cancer*

	leukemia	not	Total persons
drug rx	14	2053	2067
no drug rx	1	1565	1566
	15	3618	3633

"Expected" numbers of leukemia under
H_0: rate not increased by drug

	leukemia	not	Total persons
drug rx	**8.53**	**2058.47**	2067
no drug rx	**6.47**	**1559.53**	1566
	15	3618	3633

$$x^2 = \frac{\{14 - \mathbf{8.53}\}^2}{\mathbf{8.53}} + \frac{\{2053 - \mathbf{2058.47}\}^2}{\mathbf{2058.47}}$$

$$+ \frac{\{1 - \mathbf{6.47}\}^2}{\mathbf{6.47}} + \frac{\{1565 - \mathbf{1559.53}\}^2}{\mathbf{1559.53}}$$

$$= 8.17$$

e.g. MI in the first 56 months of US MDs' study of aspirin

	MI	not	Total MDs
aspirin	104	remainder	11K
placebo	189	remainder	11K
	293	remainder	22K

"Expected" numbers of MI under **H_0: rate not affected by aspirin**

	MI	not	Total persons
drug rx	**146.5**	**rest**	11K
no drug rx	**146.5**	**rest**	11K
	293	**rest**	22K

e.g. US MDs' study of aspirin ... continued

$$x^2 = \frac{\{104 - \mathbf{146.5}\}^2}{\mathbf{146.5}} + \frac{\{189 - \mathbf{146.5}\}^2}{\mathbf{146.5}} + \frac{\{42.5\}^2}{\mathbf{11K}} + \frac{\{-42.5\}^2}{\mathbf{11K}} = \mathbf{24.7}$$

In effect, testing whether 293 MI's could distribute this unevenly if used a coin.

e.g. *Breast Cancer in women repeatedly exposed to multiple X-ray fluoroscopies Boice and Monson 1977*

	cancers	Women-years (WY)
exposed	41	28,010
not exposed	15	19,017
	56	47,027

"Expected" numbers of MI under **H_0: rate not affected by X-rays**

	cancers	Women-years (WY)
exposed	**33.4**	28,010
not exposed	**22.6**	19,017
	56	47,027

$$\chi^2 = \frac{\{41 - \mathbf{33.4}\}^2}{\mathbf{33.4}} + \frac{\{15 - \mathbf{22.6}\}^2}{\mathbf{22.6}} + \frac{\{\text{deviation}\}^2}{\approx \mathbf{28K}} + \frac{\{\text{deviation}\}^2}{\approx \mathbf{19K}}$$

$$\approx \frac{\{41 - \mathbf{33.4}\}^2}{\mathbf{33.4}} + \frac{\{15 - \mathbf{22.6}\}^2}{\mathbf{22.6}} = 4.29$$

[3.74 with continuity correction]

Equivalent to testing whether the a+b events could split in this extreme or more extreme a way.. would expect under H_0 that the split would be (apart from random variation) in the ratio of $WY_{exposed} : WY_{non\text{-}exposed}$.

WE WILL REVISIT THE 2x1 TABLE , AND THE 1 x 1 TABLE, WHEN COMPUTING EFFECT MEASURES for INCIDENCE RATES]

e.g. *Response of <u>same</u> subject in each of 2 conditions (self-paired)*

Responses of matched pair, one in 1 condition, 1 in other

Δ's in paired responses on interval scale, reduced to <u>sign</u> of Δ

		Result in Other PAIR Member: Positive	Result in Other PAIR Member: Negative	Total PAIRS
Result in One PAIR Member	Positive	a	b	
	Negative	c	d	n PAIRS

		Exposure in "Control": Positive	Exposure in "Control": Negative	Total PAIRS
Exposure in "Case"	Positive	a	b	
	Negative	c	d	n PAIRS

extreme situations (1 or other / forced choice e.g. exercise 8.18, or who dies first among twin pairs discordant for handedness)

		Shorter: Won	Shorter: Lost	Total PAIRS
Taller	Won	-	b	
	Lost	c	-	n PAIRS

Can also turn this table 'inside-out' and analyze using case-control approach

		Loser: Taller	Loser: Shorter	Total PAIRS
Winner	Taller	-	b	
	Shorter	c	-	n PAIRS

(McNemar) Test of equality of proportions:

-1- discard the concordant pairs (+,+) and (–,–) as being "un-informative" (this point is somewhat controversial)

-2- analyze split of (b+c) discordant pairs (under H_0, expect 50:50)

Example: HIV in twins in relation to order of delivery [LancetDec14'91]

Mother -> infant transmission of HIV-infection: 66 sets of twins

		Result in 2nd-born Twin: HIV +	Result in 2nd-born Twin: HIV –	Total Sets
Result in 1st born Twin	HIV +	10	18	
	HIV –	4	34	66 Sets

To analyze the 'split' of discordant pairs:

<u>(if n small)</u>

Binomial probabilities with "n" = b+c and $\pi = 0.5$

(Table C or Table for Sign Test is helpful here)

<u>(if n larger)</u>

- Z test of observed proportion p = b / (b+c) vs $\pi = 0.5$
- χ^2 test on observed 1x2 table [b | c]

versus 1x2 table expected if H_0 holds

$$\left[\frac{b+c}{2} \mid \frac{b+c}{2} \right]$$

Note that the Z^2 and x^2 are equivalent

McNemar) Test of equality of proportions: worked example

Result in 1st born Twin	Result in 2nd-born Twin: HIV +	HIV –	Total Sets
HIV +	10	18	
HIV –	4	34	66 Sets

Analysis using exact binomial

Binomial probabilities with "n" = b+c = 28 with discordant outcomes
Under H_0 that order makes no difference to likelihood of HIV transmission, the split among these 22 should be like that obtained by tossing 22 coins, each with

π (first born is the one to have the HIV transmitted) = 0.5

The Binomial(n=22, π = 0.5) distribution is not available in Table C, but can be obtained from Excel. Of interest is the sum of the probabilities for 18/22, 19/22, 20/22, 21/22 and 22/22 (1-sided) then doubled if dealing with a 2-sided alternative, i.e. 2 x (0.00174 + 0.00036 + negligible terms) = 0.004.

Gaussian (or equivalently, Chi-square) Approximation to Binomial

$$Z = \frac{18 - E[b]}{SD[b]} = \frac{18 - \frac{22}{2}}{\sqrt{22 \times 0.5 \times 0.5}} = \frac{18 - 11}{\sqrt{5.5}} = 2.98 \; ; Z^2 = \frac{7^2}{5.5} = 8.91$$

Prob (| Z | > 2.98) = 0.003; From Table F, Prob[X^2 (1 df)> 8.91] ≈ 0.003

χ^2 test on observed 1x2 table [18 4]
versus 1x2 table **expected** [**11** **11**]
if H_0 holds

$$X^2 \text{ test} = \frac{\{18 - \mathbf{11}\}^2}{\mathbf{11}} + \frac{\{4 - \mathbf{11}\}^2}{\mathbf{11}} = \frac{7^2}{5.5} = 8.91$$

Notice that because of the symmetry involved in testing $\pi = 0.5$ versus a 2-sided alternative, the test statistics have a particularly simple form:

$$Z^2 = x^2 = \frac{(b - c)^2}{b + c} \; ; \quad \frac{(18 - 4)^2}{22} = 8.91 \text{ in our example}$$

With continuity correction $Z^2_c = x^2_c = \frac{(|b - c| - 1)^2}{b + c} = \frac{169}{22} = 7.68$
(2-sided P ≈ 0.005)

Q: Why a continuity correction of 1 rather than usual 0.5?

A: The difference b–c jumps in 2's rather than 1's

e.g. if b+c =18, then b – c =18, 16, 14, ... ,-14, -16, -18)

Situation	Question	Gaussian Approximation: no	Gaussian Approximation: yes
1 Popln.	**CI** for π	• Nomograms/tables/spreadsheet	• $p \pm z\sqrt{\frac{p[1-p]}{n}}$ (cf. asymmetric)
π (see notes on 8.1)	**Test** π_0	• Binomial distribution	• $z = \frac{p - \pi_0}{\sqrt{\frac{\pi_0[1-\pi_0]}{n}}}$ $\{z^2 = x^2\}$
(sample of n; p = y/n are "positive")			
2 Populations (matched samples) or 1 population under 2 conditions			
OR = $[\pi_1/(1-\pi_1)]/[\pi_2/(1-\pi_2)]$			
	CI for **OR**	• b ~ Bin(n',[**OR**/(1 + **OR**)])	• CI for **OR**/(1+**OR**) => CI for **OR**
Test $\pi_1 = \pi_2$	**Test** **OR=OR**$_0$	• b ~ Bin(n',[**OR**$_0$/(1 + **OR**$_0$)])	• $z = \frac{b/n' - \mathbf{OR}_0/[1+\mathbf{OR}_0]}{SE[\, b/n' \mid H_0\,]}$ or x^2
(sample of n pairs n(++)=a; n(+-)=b; n(-+)=c; n(--)=d; b+c=n' ; or (i.e. est. of **OR**) = b/c)			
2 Poplns.	**CI** for $\Delta = \pi_1 - \pi_2$	• Miettinen and Nurminen	• $p_1 - p_2 \pm z\sqrt{\frac{p_1[1-p_1]}{n_1} + \frac{p_2[1-p_2]}{n_2}}$ (also via Binomial regression** ... RD)
π_1 and π_2	**CI** for **RR**	•	• cf Rothman p134; regression (RR)
	CI for **OR**	• Conditional	• Condnl[Approx.]/Woolf/Miettinen (or via Binomial(logistic)regression**)
	Test RR or **OR**$_0$ or Δ_0	• Fisher's Exact Test (cond'nl)	• $z = \frac{[\, p_1 - p_2\,] - \Delta_0}{\sqrt{\frac{p[1-p]}{n_1} + \frac{p[1-p]}{n_2}}}$ (*) {or x^2}
		• Unconditional methods (Suissa and Schuster)	
		• Permutational (StatExact software)	
(independent samples of n_1 and n_2)			

Notes:

(*) p in combined data = $\frac{n_1p_1 + n_2p_2}{n_1 + n_2} = \frac{\Sigma \text{ numerators}}{\Sigma \text{ denominators}}$ (weighted average of two p 's)

** Binomial Regression: extension [to come] of 1-parameter binomial regression models described in notes for 8.1

e.g. *Independence of classification on 2 variables*
Similarity of multinomial profiles

(generic) Relationship between one factor (rows) and another (columns) in n observations; crossclassified into an

r(=# of rows) x c(=# of columns) table.

	Col_1	Col_2	...	Col_c	*Total*
Row_1	n_{11}	n_{12}	...	n_{1c}	N_{row1}
Row_2	n_{21}	n_{22}	...	n_{2c}	N_{row2}
.....	...	...	...	...	...
Row_r	n_{r1}	n_{r2}	...	n_{rc}	N_{rowr}
Total	N_{col1}	N_{col2}		N_{colc}	N

(e.g. 1) Relationship between laterality of hand and laterality of eye (measured by astigmatism, acuity of vision, etc.) in 413 subjects crossclassified into a 3x3 table. [data from Woo, Biometrika 2A 79-148]

	Left-eyed	Ambiocular	Right-eyed	*Total*
Left handed	34	62	28	124
Ambidextrous	27	28	20	75
Right handed	57	105	52	214
Total	118	195	100	413

$$\chi^2_{df} = \sum \frac{\{ \text{observed} - \mathbf{expected} \}^2}{\mathbf{expected}}$$

- **expected** number in cell = $\frac{N_{row} \bullet N_{column}}{N}$
- summation is over all r x c cells
- degrees of freedom (df) = (r–1)(c–1). In above eg., r=3; c=3 => df:4

The χ^2 statistic measures the deviation from independence of row and column classifications (e.g. 1) and dissimilarity of the distributions (profiles) of responses (e.g. 2 and 3). However, omnibus chi-square tests (H_0: identical response profiles) with large df are seldom of interest, since the alternative hypothesis (profiles are not identical) is so broad, and the chi-square tests are invariant to the ordering of the rows and columns. More often, a specific alternative hypothesis is of interest; omnibus tests penalize one for looking in all directions, when in fact one's focus is narrower, and aiming to pick up a specific 'signal'. The next 2 examples (>2 ORDERED response categories in each of 2 groups; binary responses in > 2 ORDERED exposure categories are a more fruitful step in this direction.

Analyzing data from ORDERED categories

Using a chi-square test for the following 2x3 table ignores the ordered nature of the responses

e.g. 2 Quality of sleep before elective operation. [BMJ]

	Bad	Reasonably good	Good	*Total*
Patients given Triazolam	2	17	12	31
Patients given Placebo	8	15	8	31
Total	10	32	20	62

See article by Moses L et al NEJM 311 442-448 1984 (also published as Chapter in Medical Uses of Statistics by J Bailar and F Mosteller}.

e.g. 3 Outcome after 2 to 7 days of Rx in 20 patients with chronic oral candidiasis.

	Outcome category				
	1(good)	2	3	4 (poor)	*Total*
Clotrimazole	6	3	1	0	10
Placebo	1	0	0	9	10
Total	7	3	1	9	20

Any dichotomization of outcomes loses information and statistical power. Moses et al. suggest using the Mann-Whitney U test (also known as the Wilcoxon Rank sum test) to take account of ordered nature of response categories.

Suppose that, in a k x 2 contingency table the k groups fall into a natural order. They may correspond to different values, or groups of values, of a quantitative variable like age; or they may correspond to qualitative categories, such as severity of a disease, which can be ordered but not readily assigned a numerical value. The usual $\chi^2_{(k-1)}$ test is designed to detect differences between the k proportions --- without taking the 'ordering' of the rows into account. It is an 'omnibus' test and is unchanged even if we interchange the order of the columns. More specifically one might ask whether there is a significant trend in these proportions from group 1 to group k. Let us assign a quantitative variable, x, to the k groups. If the definition of groups uses such a variable, this can be chosen to be x. If the definition is qualitative, x can take integer values from 1 to k. The notation is as follows:

		Frequency			proportion
Group	X	Pos	Neg	Total	positive
1	x_1	r_1	$n_1 - r_1$	n_1	p_1
1	x_2	r_2	$n_2 - r_2$	n_2	p_2
.	.	.	.	.	.
.	.	.	.	.	.
i	x_i	r_i	$n_i - r_i$	n_i	p_i
k	x_k	r_k	$n_k - r_k$	n_k	p_k
All		R	N – R	N	P(=R/N)

The $\chi^2_{(1)}$ **statistic for trend**, $X^2_{(1)}$, which forms part of the overall X^2, can be computed as follows:

$$X^2_{(1\ df)} = \frac{N\{N\sum r_i x_i - R\sum n_i x_i\}^2}{R\{N-R\}[N\sum n_i x_i^2 - (\sum n_i x_i)^2]}$$

From SAS

if 1 line of data for each of 119 individuals

```
PROC FREQ DATA= ... ;
TABLES falls*sick /TREND;
```

if enter a variable (say "number" to indicate how many persons has each exposure/response pattern, then syntax is

```
PROC FREQ DATA=   ;
TABLES falls*sick / TREND;
 WEIGHT number;
```

From Stata

```
input falls ill number
        0    0   33
        0    1    8
        5    0   19
        5    1   15
       15    0   11
       15    1    9
       25    0    4
       25    1   10
       40    0    0
       40    1   10
end

tabodds ill falls [freq=number]
```

Example [jh]

Distribution of subjects with polluted-water exposure-related symptoms among Competitors and Employees and Relative Risk (RR) According to Number of Falls in the Water Data from article "Health Hazards Associated with Windsurfing on Polluted Water " AJPH 76 690-691, 1986 -- research conducted at the Windsurfer Western Hemisphere Championship held over 9 days in August 1984. During the championships, the same single-menu meals were served to both competitors and employees]

Groups of subjects	No. of subjects with symptoms	No without	Total	RD	RR	OR
Employees (ref gp)	8 (20%)	33	41	--	1.0	1.0
Competitors :						
0-10 falls	15 (44%)	19	34	24%	2.3	3.3
11-20 falls	9 (45%)	11	20	25%	3.5	3.4
21-30 falls	10 (71%)	4	14	51%	3.7	10.3
> 30 falls	10 (100%)	0	10	80%	5.1	inf.

Any dichotomization of exposure loses information and statistical power. Authors correctly used Chi-square test for trend, yielding $\chi^2_{1df} = 25.3$, $P = 10^{-6}$. I get 24.58 with the "spacing" 0, 5, 15, 25 and 40. SAS*, using the "Cochran-Armitage Trend Test", with same spacing, gives a Z statistic of -4.969, ($Z^2 = 24.69$). The entire variation among the 5 proportions in the table (ignoring ordering) is approximately $X^2(4\ df) = 27$, but it is almost all explained by the exposure gradient. In smaller datasets, even if the overall X^2 is not significant, the trend portion can be. In this e.g. there was such a strong relationship that even the overall test was significant. The same is true in the example overleaf (dealing with birth date and sporting success), where again the sample sizes are large and the signal strong.

* Syntax `proc freq DATA= ... ; tables falls*sick /trend;`

This syntax assumes you enter data for each of the 119 individuals; if instead you enter a variable (say you call it "number" to indicate how many persons has each exposure/response pattern, then the required syntax is

```
proc freq DATA= ..; tables falls*sick /trend;
 weight number;
```

PS: If you look up A&B, you will find another x^2 [Eqn. 12.2]. This value, calculated as the difference between the trend and the overall x^2 statistics, can be is used to test if there is serious non-linear variation over and above the linear trend.

Example **Birth date and sporting success**

SCIENTIFIC CORRESPONDENCE in NATURE • VOL 368 • 14 APRIL 1994 p592

Sir — I have found a significant relationship between birth date and success in tennis and soccer. In the Netherlands and England, players born early in the competition year are more likely to participate in national soccer leagues. The high incidence of elite athletes born in the first quarter of the competition year can be explained by the effects of age-group position.

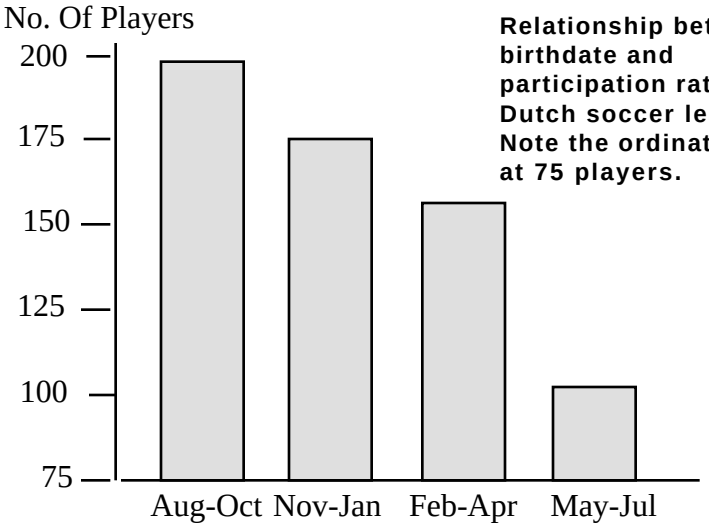
In organized sport. talent is considered predominantly in terms of physical skills. and the influence of social and psychological factors is often ignored or underestimated[1]. Various studies have investigated the psychological characteristics of elite athletes[2], but none has looked for an effect of age. I discovered a strikingly skewed distribution of the dates of birth of 12- to l6-year-old tennis players in the top rankings of the Dutch youth league. Half of a sample of 60 tennis players were born in the first 3 months of the year.

This discovery led me to consider the distribution of the dates of birth of professional soccer players. In the Netherlands, there are two leagues comprising a total of 36 clubs. I found a striking difference between participation rates of those born in August and July. The Dutch soccer competition year starts on the first of August. A chi-square test indicates that the distribution is not uniform ($P<0.001$); and a regression analysis demonstrates a clear linear relationship between month of birth and number of participants. The dates of birth of 621 players, compiled into quarters, are shown in the figure. This relationship cannot be attributed to the distribution of births in the Netherlands, as this is highly uniform.

We also inspected the distribution of the dates of birth of English football players in league clubs in the period 1991-92 (ref.3). Birth dates for all players were tabulated by month and compiled into quarters. The results (table) show the significant effect of date of birth on participation rate of soccer players within each of the national leagues, indicating that. as in the Netherlands, significantly more football players are born in the first quarter of the competition year (which starts in September in England).

There is a known relationship between date of birth and educational achievement[5]. implying that the younger children in any school year group are at a disadvantage compared to the older children. Children who participate in sports are also placed in age groups, and my results imply many athletes in organized sports may never get a fair chance because of this method of classification. Very little attention has been drawn to this problem. One of the few studies done in this area analysed the dates of birth of young Canadian hockey players in the 1983-84 season[6]. Players possessing a relative age advantage (born in the months lanuary-June) were more likely to participate in minor hockey and more likely to play for top teams than players in July-December.

More than 20 years ago, this journal published an article concerning the relationship between season of birth and cognitive development[7]. The authors attributed this relationship to a fault in the British educational system. A similar relationship was found[5] in the Netherlands. Despite this, no action was undertaken to change the educational system. One can only hope that this will not he the case for sports.



Relationship between birthdate and participation rates in Dutch soccer league. Note the ordinate begins at 75 players.

PARTICIPATION RATES IN ENGLISH SOCCER LEAGUES

	Players in birthdate quarters				Statistics		
League	Sep-Nov	Dec-Feb	Mar-May	Jun-Aug	Total	Chi-Square	Sig. Level
FA premier	288	190	147	136	761	75.5	P<0.0001
Division 1	264	169	154	147	734	48.47	P<0.0001
Division 2	251	168	123	131	673	61.11	P<0.0001
Division 3	217	169	121	102	609	52.38	P<0.0001
Total	1,020	696	545	516	2,777	230.77	P<0.0001

References: **1** Dudink A Fur J High Ability 1, 144-150 (1990). **2** Dudink A & Bakker. F. Ned. Tschr. Psychol 48. 55 -69 (1993). **3** Rollin,J *Rothmans Football Yearbook 1992-93* (Headline. London. 1992). **4** Shearer.E Educ Res 10. 51-56 (1967) **5** Doornbos, K. [Date of birth and scholastic performance (Wolters-Noordhoff, Groningen. 1971). **6** Barnsley. R. H. & Thompson A. H. Can. J. Behav. Sci 20. 167-176 (1988). **7** Williams. Ph.. Davies P., Evans, R & Ferguson, N. Nature 228. 1033-1036 (1970).

Ad Dudink Faculty of Psychology, University of Amsterdam, 1018 WB 3 Amsterdam, The Netherlands

——

For an example of an analysis of seasonal variation, see the article by H T Sørensen et al. Does month of birth affect risk of Crohn's disease in childhood and adolescence? p 907 BMJ VOLUME 323 20 OCTOBER 2001 bmj.com (copy of article, and associated dataset, on course 626 website).