# Web-based Enrollment and other types of Selection: Consequences for Generalizability

Niels Keiding

Section of Biostatistics
University of Copenhagen

nike@sund.ku.dk

Statistical Days
Turku, Finland, 18-19 May 2017

# I. Standardization vs. modelling for confounder control in observational studies: a historical perspective

Joint work with David Clayton

Keiding, N. & Clayton, D. (2014). Standardization and control for confounding in observational studies: a historical perspective. *Statist. Sci.* **29,** 529-558.

# Selected topics from the dialogue between

| | |
|---|---|
| *Weighting and standardization* | *Models, parameters, stratification* |
| *Marginal effect measures* | *Conditional effect measures* |

# Simpson (1951): Conditional or marginal effect measures

E.H. Simpson (1951). The interpretation of interaction in contingency tables. *J. Roy. Statist. Soc. B* **13**, 238-241.
M.A. Hernán, D. Clayton, N. Keiding (2011). The Simpson's paradox unraveled. *Int. J. Epid.* **40**, 780-785.

| | B=1 | B=0 |
|---|---|---|
| A=1 | 20 | 20 |
| A=0 | 6 | 6 |

OR = 1

C = 1

| | B=1 | B=0 |
|---|---|---|
| A=1 | 5 | 8 |
| A=0 | 3 | 4 |

OR = 5/6

C = 0

| | B=1 | B=0 |
|---|---|---|
| A=1 | 15 | 12 |
| A=0 | 3 | 2 |

OR = 5/6

# Simpson: baby playing cards

| | B=1 | B=0 |
|---|---|---|
| A=1 | 20 | 20 |
| A=0 | 6 | 6 |

OR = 1

C = 1

| | B=1 | B=0 |
|---|---|---|
| A=1 | 5 | 8 |
| A=0 | 3 | 4 |

OR = 5/6

C = 0

| | B=1 | B=0 |
|---|---|---|
| A=1 | 15 | 12 |
| A=0 | 3 | 2 |

OR = 5/6

| | | | |
|---|---|---|---|
| A = 0 | court cards (B, D, K) | B = 0 | red (heart, diamond) |
| A = 1 | not court cards (A, 2, 3,…, 10) | B = 1 | black (spade, club) |

C = 1 card dirty because baby played with it
C = 0 card clean

Is colour independent of court status?

Yes, marginally OR = 1
No, conditionally on dirtiness OR = 5/6 for C = 1 and for C = 0.

Relevant effect measure: Marginal

# Simpson: medical treatment

|  | B=1 | B=0 |
|---|---|---|
| A=1 | 20 | 20 |
| A=0 | 6 | 6 |

OR = 1

C = 1

|  | B=1 | B=0 |
|---|---|---|
| A=1 | 5 | 8 |
| A=0 | 3 | 4 |

OR = 5/6

C = 0

|  | B=1 | B=0 |
|---|---|---|
| A=1 | 15 | 12 |
| A=0 | 3 | 2 |

OR = 5/6

A = 0 not treated
A = 1 treated

B = 0 not dead
B = 1 dead

C = 1 male
C = 0 female

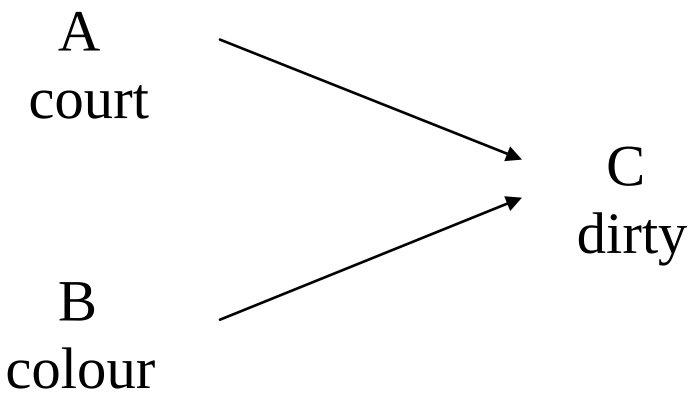
Does treatment affect death?

| | |
|---|---|
| No, marginally | OR = 1 |
| Yes for males, | OR of dying = 5/6 |
| Yes for females, | OR of dying = 5/6 |

Relevant effect measure: Conditional on sex.

*Females die more, females are treated more*

# Baby playing cards



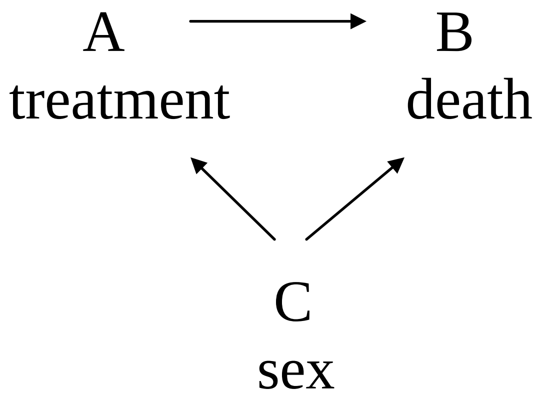
A
court

C
dirty

B
colour

Even if A and B are independent,
they become artificially associated
by conditioning on the collider C

## Medical treatment



A
treatment

B
death

C
sex

Treatment as well as death depends
on sex which is a possible confounder
that should be controlled for.

# Conditional and marginal effect measures

***Marginal**: apply age specific rates to a *target* age structure and compare the predicted *marginal* summaries in this target population*

corresponds to handling confounders by making sure their distribution is the same in study and control population

Direct standardization, randomized trials

***Conditional**: compare covariate-specific rates*

corresponds to handling confounders by stratification or restriction

Indirect standardization, regression analysis

# Direct and Indirect Standardization

| | *Study population* | *Standard population* | |
|---|---|---|---|
| no. individuals | $A_1,\cdots,A_k \quad a_i = A_i \big/ \sum A_i$ | $S_1,\cdots,S_k \quad s_i = S_i \big/ \sum S_i$ | |
| age distribution | $a_1,\cdots,a_k \quad \sum a_i = 1$ | $s_1,\cdots,s_k \quad \sum s_i = 1$ | |
| death rates | $\alpha_1,\cdots,\alpha_k$ | $\lambda_1,\cdots,\lambda_k$ | |
| Actual no. deaths | $\sum A_i \alpha_i$ | $\sum S_i \lambda_i$ | |
| Exp. no. deaths under standard death rates (*Indirect standardization*) | $\sum A_i \lambda_i$ | | *OBS/EXP*<br>$SMR = \sum A_i \alpha_i \big/ \sum A_i \lambda_i$ |
| Crude death rate | $\sum A_i \alpha_i \big/ \sum A_i$ | $\sum S_i \lambda_i \big/ \sum S_i$ | |
| Standardized death rate under standard age distribution (*Direct standardization*) | $\sum s_i \alpha_i = \sum S_i \alpha_i \big/ \sum S_i$ | | *EXP/OBS*<br>$CMF = SRR = \sum S_i \alpha_i \big/ \sum S_i \lambda_i$ |

# Indirectly Standardized death rate

$$\sum\left(\frac{A_i\alpha_i}{\sum A_i\lambda_i}\frac{\sum S_i\lambda_i}{\sum S_i}\right)=SMR\cdot\text{ Crude death rate for Standard population}$$

# Direct and Indirect Standardization in terms of target population

T. Sato & Y. Matsuyama (2003). Marginal structural models as a tool for standardization.
*Epidemiology* **14,** 680-686.

*Study (exposed) Population Target*:

$$SMR = \sum A_i \alpha_i \Big/ \sum A_i \lambda_i$$

compare observed no. deaths to counterfactual no. of deaths if unexposed

*Standard (unexposed) Population Target*:

$$CMF = SRR = \sum S_i \alpha_i \Big/ \sum S_i \lambda_i$$

compare counterfactual no. deaths if exposed to observed no. deaths.

Third possibility: *Total Population Target*:

$$SRR_T = \frac{\sum (A_i + S_i)\alpha_i}{\sum (A_i + S_i)\lambda_i} = \frac{\sum A_i \alpha_i \big/ \left(A_i / (A_i + S_i)\right)}{\sum S_i \lambda_i \big/ \left(S_i / (A_i + S_i)\right)}$$

# Total Population Target: Interpretation

$$SRR_T = \frac{\sum (A_i + S_i)\alpha_i}{\sum (A_i + S_i)\lambda_i} = \frac{\sum A_i\alpha_i \Big/ \big(A_i / (A_i + S_i)\big)}{\sum S_i\lambda_i \Big/ \big(S_i / (A_i + S_i)\big)}$$

$$= \frac{\sum (\text{deaths in Study Population in age group } i) \Big/ \big(A_i / (A_i + S_i)\big)}{\sum (\text{deaths in Standard Population in age group } i) \Big/ \big(S_i / (A_i + S_i)\big)}$$

numerator: for each age group *i*, deaths in study population weighted with inverse probability of being in study population

denominator: for each age group *i*, deaths in standard population weighted with inverse probability of being in standard population

# Comparison of several study populations

G.U. Yule (1934). On some points relating to vital statistics, more especially statistics of occupational mortality (with discussion). *J.Roy. Statist. Soc.* **97**, 1-84.

| | | |
|---|---|---|
| Study populations | A, a, $\alpha$ | B, b, $\beta$ |
| Standard population | S, s, $\lambda$ | |

$$\frac{\mathrm{SMR}(A)}{\mathrm{SMR}(B)} = \frac{\sum A_i \alpha_i \big/ \sum A_i \lambda_i}{\sum B_i \beta_i \big/ \sum B_i \lambda_i} = ?$$

no interpretation for
indirect standardization

$$\frac{\mathrm{CMF}(A)}{\mathrm{CMF}(B)} = \frac{\sum S_i \alpha_i \big/ \sum S_i \lambda_i}{\sum S_i \beta_i \big/ \sum S_i \lambda_i} = \frac{\sum S_i \alpha_i}{\sum S_i \beta_i}$$

Direct standardization yields interpretation as ratio of averages, using standard population as weights.

*Indirect and direct standardization are not just the same*

# Historical origin …

Anonymous [W. Dale] (1777). A Supplement to Calculations of the Value of Annuities, Published for the Use of Societies Instituted for Benefit of Age Containing Various Illustration of the Doctrine of Annuities, and Compleat Tables of the Value of 1£. Immediate Annuity. (Being the Only Ones Extant by Half-Yearly Interest and Payments). Together with Investigations of the State of the Laudable Society of Annuitants; Showing What Annuity Each Member Hath Purchased, and Real Mortality Therein, from its Institution Compared with Dr. Halley's Table. Also Several publications, Letters, and Anecdotes Relative to that Society. And Explanatory of Proceedings to the Present year. London: Ridley.

See N. Keiding (1987). The method of expected number of deaths, 1786-1886-1986.
*Int. Statist. Rev.* **55**, 1-20.

# The STATE of MORTALITY in the LAUDABLE SOCIETY of ANNUITANTS for Benefit of AGE, Bottom of BARTHOLOMEW-LANE, compared with Dr. HALLEY's TABLE, annually from CHRISTMAS 1766 to CHRISTMAS 1775.

| Age | 1767. | Number of Deaths and Excluſions, Anno 1768 | 1768. | Number of Deaths and Excluſions, Anno 1769 | 1769. | Number of Deaths and Excluſions, Anno 1770 | 1770. | Deaths and Excluſions in the Year 1771 | 1771. | | 1772. | Number of Deaths and Excluſions, Anno 1773 | 1773. | Number of Deaths and Excluſions, Anno 1774 | 1774. | Age | Multipliers. | 1768. | 1769. | 1770. | 1771. | 1772. | 1773. | 1774. | 1775. | Age | DEATHS at CHRISTMAS. 1774. | | 1775. | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Admitted on or ſince Chriſtmas — 1766<br>Some whoſe Ages were not known. Sect. xiii. 1. 8.<br>Admitted before Chriſtmas — 1767 | | Remaining of the paſt Year — 1767<br>Admitted in the Courſe of the Year — 1768<br>Total living at Chriſtmas — 1768 | | Remaining of the paſt Year — 1768<br>Admitted in the Courſe of the Year — 1769<br>Total living at Chriſtmas — 1769 | | | | | | | | | | | | | | | | | | | | | | Real | Expected | Real | Expected |
| 1 | | | | | | | + 0= 0 | | 0+ 2= 2 | 1 ex. | 0+ 1= 1 | | 0+ 0= 0 | | | 1 | .145 | | | | — | .2900 | .1450 | — | — | 1 | | .4350 | | .4350 |
| 2 | | | | | | | + 0= 0 | | 0+ 1= 1 | 1 d. | 1+ 0= 1 | | 1+ 0= 1 | | 0+ 0= 0 | 2 | .0666 | | | | — | .0666 | .0666 | .0666 | — | 2 | 1 | .1998 | | .1998 |
| 3 | | | | | | | + 4= 4 | 1 d. | 0+ 0= 0 | | 0+ 1= 1 | | 1+ 0= 1 | | 1+ 0= 1 | 3 | .0476 | | | | .1904 | — | .0476 | .0476 | .0476 | 3 | 1 | .2856 | | .3332 |
| 4 | | | | | | | + 1= 1 | 1 d. | 3+ 1= 4 | 1 d.<br>1 ex. | 0+ 0= 0 | | 1+ 0= 1 | | 1+ 0= 1 | 4 | .0368 | | | | .0368 | .1472 | — | .0368 | .0368 | 4 | 2 | .2208 | | .2576 |
| 5 | | | | | | | + 0= 0 | | 0+ 1= 1 | 1 ex. | 2+ 0= 2 | | 0+ 0= 0 | | 1+ 0= 1 | 5 | .03005 | | | | — | .0300 | .0600 | — | .0300 | 5 | | .0900 | | .1200 |
| 6 | | | | | | | + 0= 0 | | 0+ 0= 0 | | 0+ 0= 0 | | 2+ 1= 3 | | 0+ 0= 0 | 6 | .0253 | | | | — | — | — | .0759 | — | 6 | | .0759 | | .0759 |
| 7 | | | | | | | + 2= 2 | | 0+ 0= 0 | | 0+ 0= 0 | | 0+ 0= 0 | | 3+ 0= 3 | 7 | .0173 | | | | .0346 | — | — | — | .0519 | 7 | | .0346 | | .0865 |
| 8 | | | | | | | + 0= 0 | | 2+ 1= 3 | 1 ex. | 0+ 1= 1 | | 0+ 1= 1 | | 0+ 0= 0 | 8 | .0147 | | | | — | .0441 | .0147 | .0147 | — | 8 | | .0735 | | .0735 |
| 9 | | | | | | | + 1= 1 | | 0+ 0= 0 | | 2+ 0= 2 | | 1+ 0= 1 | | 1+ 0= 1 | 9 | .01343 | | | | .0134 | — | .0268 | .0134 | .0134 | 9 | | .0536 | | .0670 |
| 10 | | | | | | | + 1= 1 | | 1+ 0= 1 | | 0+ 0= 0 | | 2+ 0= 2 | | 1+ 0= 1 | 10 | .01210 | | | | .0121 | .0121 | — | .0242 | .0121 | 10 | | .0484 | | .0605 |
| 11 | | | | | | | + 0= 0 | | 1+ 1= 2 | 1 ex. | 1+ 0= 1 | | 0+ 0= 0 | | 2+ 0= 2 | 11 | .01072 | | | | — | .0214 | .0107 | — | .0214 | 11 | | .0321 | | .0535 |
| 12 | | | | | | | + 2= 2 | | 0+ 1= 1 | | 1+ 1= 2 | | 1+ 0= 1 | | 0+ 0= 0 | 12 | .00928 | | | | .0185 | .0092 | .0185 | .0092 | — | 12 | | .0554 | | .0554 |
| 13 | | | | | | | + 0= 0 | | 2+ 1= 3 | 1 ex. | 1+ 0= 1 | | 2+ 0= 2 | | 1+ 0= 1 | 13 | .00937 | | | | — | .0281 | .0093 | .0187 | .0093 | 13 | | .0561 | | .0654 |
| 14 | | | | | | | + 1= 1 | | 0+ 0= 0 | | 2+ 0= 2 | | 1+ 0= 1 | | 2+ 0= 2 | 14 | .00946 | | | | .0094 | — | .0189 | .0094 | .0189 | 14 | | .0377 | | .0566 |
| 15 | | | | | | | + 3= 3 | | 1+ 0= 1 | | 0+ 0= 0 | | 2+ 0= 2 | | 1+ 0= 1 | 15 | .00955 | | | | .0286 | .0095 | — | .0191 | .0095 | 15 | | .0572 | | .0667 |
| 16 | | | | | | | + 1= 1 | | 3+ 0= 3 | | 1+ 0= 1 | | 0+ 0= 0 | | 2+ 0= 2 | 16 | .00964 | | | | .0096 | .0289 | .0096 | — | .0192 | 16 | | .0481 | | .0673 |
| 17 | | | | | | | + 4= 4 | | 1+ 0= 1 | | 3+ 0= 3 | | 1+ 0= 1 | | 0+ 0= 0 | 17 | .00974 | | | | .0389 | .0097 | .0292 | .0097 | — | 17 | | .0875 | | .0875 |
| 18 | | | | | | | + 4= 4 | | 4+ 0= 4 | | 1+ 0= 1 | | 3+ 1= 4 | | 1+ 0= 1 | 18 | .00983 | | | | .0393 | .0393 | .0098 | .0393 | .0098 | 18 | | .1277 | | .1375 |
| 19 | | | | | | | + 4= 4 | | 4+ 0= 4 | | 4+ 0= 4 | | 1+ 0= 1 | | 4+ 0= 4 | 19 | .00993 | | | | .0397 | .0397 | .0397 | .0099 | .0397 | 19 | | .1290 | | .1687 |
| 20 | | | | | | | + 4= 4 | | 4+ 0= 4 | | 4+ 0= 4 | | 4+ 0= 4 | | 1+ 0= 1 | 20 | .01003 | | | | .0401 | .0401 | .0401 | .0401 | .0100 | 20 | | .1604 | | .1704 |
| | | | | | | | 32 | | 35 | | 27 | | 26 | | 22 | | | | | | .5114 | .8159 | .5465 | .4346 | .3296 | | 4 | 2.3084 | | 2 6380 |
| 21 | 2 | | 0+ 0= 0 | | 0+ 5= 5 | | 1+ 4= 5 | | 4+ 2= 6 | | 4+ 0= 4 | | 4+ 0= 4 | | 4+ 0= 4 | 21 | .01013 | .0202 | — | .0505 | .0505 | .0606 | .0404 | .0404 | .0404 | 21 | | .2626 | | .3030 |
| 22 | 2+ 1= 3 | | 2+ 2= 4 | | 0+ 2= 2 | | 5+ 6=11 | | 5+ 1= 6 | | 6+ 0= 6 | | 4+ 3= 7 | | 4+ 1= 5 | 22 | .01194 | .0358 | .0477 | .0238 | .1313 | .0716 | .0716 | .0835 | .0597 | 22 | | .4653 | | .5250 |
| 23 | 4 | | 3+ 1= 4 | | 4+ 6=10 | | 2+ 4= 6 | | 11+ 1=12 | | 6+ 0= 6 | | 6+ 1= 7 | | 7+ 0= 7 | 23 | .01036 | .0414 | .0414 | .1036 | .0621 | .1243 | .0621 | .0725 | .0725 | 23 | | .5074 | | .5799 |
| 24 | 5 | | 4+ 2= 6 | 1 ex. | 4+13=17 | | 10+ 9=19 | | 6+ 2= 8 | 1 ex. | 12+ 2=14 | 1 d. | 6+ 0= 6 | | 7+ 0= 7 | 24 | .01047 | .0523 | .0628 | .1779 | .1989 | .0837 | .1465 | .0628 | .0732 | 24 | 1 | .7849 | | .8581 |
| 25 | 5 | | 5+ 2= 7 | | 5+10=15 | | 17+ 7=24 | | 19+ 1=20 | 1 ex. | 7+ 2= 9 | | 13+ 0=13 | | 6+ 0= 6 | 25 | .01234 | .0617 | .0863 | .1851 | .2961 | .2468 | .1110 | .1604 | .0740 | 25 | | 1.1474 | | 1.2214 |
| 26 | 7 | 1 d. | 5+ 4= 9 | 1 ex.<br>1 d. | 7+ 8=15 | | 15+ 6=21 | 1 ex. | 24+ 1=25 | | 19+ 1=20 | | 9+ 0= 9 | 1 .ex | 13+ 0=13 | 26 | .01250 | .0875 | .1125 | .1875 | .2625 | .3125 | .2500 | .1125 | .1625 | 26 | 2 | 1.3250 | | 1.4875 |
| 27 | 3 | 1 ex. | 6+ 3= 9 | 1 ex. | 7+ 8=15 | | 15+12=27 | 1 d. | 20+ 0=20 | 2 ex. | 25+ 0=25 | 1 d. | 20+ 0=20 | | 8+ 0= 8 | 27 | .01265 | .0379 | .1138 | .1897 | .3415 | .2530 | .0316 | .2530 | .1012 | 27 | 2 | 1.2205 | | 1.3217 |
| 28 | 2 | 1 ex. | 2+ 2= 4 | | 8+10=18 | | 15+ 4=19 | 1 d. | 26+ 1=27 | | 18+ 0=18 | 2 ex. | 24+ 1=25 | | 20+ 0=20 | 28 | .01282 | .0256 | .0512 | .2307 | .2435 | .3461 | .2307 | .3205 | .2564 | 28 | 1 | 1.4483 | | 1.7047 |
| 29 | 3 | | 1+ 5= 6 | | 4+14=18 | 1 d. | 18+ 6=24 | 1 ex. | 18+ 4=22 | 2 d. | 27+ 0=27 | | 16+ 1=17 | | 25+ 0=25 | 29 | .01484 | .0445 | .0890 | .2671 | .3561 | .3264 | .4006 | .2522 | .3710 | 29 | 3 | 1.7359 | | 2.1069 |
| 30 | 10 | | 3+ 5= 8 | | 6+19=25 | | 17+ 5=22 | 2 d. | 23+ 1=24 | 1 ex. | 20+ 1=21 | 2 ex. | 27+ 2=29 | | 17+ 0=17 | 30 | .01506 | .1506 | .1204 | .3765 | .3313 | .3614 | .3162 | .4367 | .2560 | 30 | 2 | 2.0931 | | 2.3491 |
| | 44 | | 57 | | 141 | | 179 | | 170 | | 150 | | 137 | | 112 | | | .5575 | .7251 | 1.7924 | 2.2738 | 2.1864 | 1.6607 | 1.7945 | 1.4669 | | 11 | 10.9904 | | 12.4573 |
| 31 | 10 | | 10+ 5=15 | 2 d. | 8+11=19 | | 25+ 4=29 | 1 ex. | 20+ 2=22 | | 23+ 1=24 | | 19+ 0=19 | | 29+ 0=29 | 31 | .01529 | .1529 | .2293 | .2905 | .4434 | .3363 | .3669 | .2905 | .4434 | 31 | 2 | 2.1098 | | 2.5532 |
| 32 | 7 | 1 d.<br>1 ex. | 10+ 3=13 | | 13+18=31 | | 19+11=30 | 1 d. | 28+ 3=31 | 1 d.<br>1 ex. | 22+ 1=23 | | 24+ 0=24 | 1 ex.<br>1 d. | 19+ 0=19 | 32 | .01553 | .1087 | .2018 | .4814 | .4659 | .4814 | .3571 | .3727 | .2950 | 32 | 4 | 2.4690 | | 2.7640 |
| 33 | 10 | | 5+ 2= 7 | | 13+11=24 | 1 ex.<br>1 d. | 31+ 8=39 | | 29+ 4=33 | | 29+ 0=29 | 1 ex. | 23+ 0=23 | | 22+ 0=22 | 33 | .01577 | .1577 | .1103 | .3784 | .6150 | .5204 | .4573 | .3627 | .3469 | 33 | 1 | 2.6018 | | 2.9487 |
| 34 | 5 | | 10+ 7=17 | | 7+22=29 | 1 d. | 22+11=33 | | 39+ 0=39 | 1 d. | 33+ 3=36 | | 28+ 1=29 | | 23+ 0=23 | 34 | .01803 | .0901 | .3065 | .5228 | .5949 | .7031 | .6490 | .5228 | .4146 | 34 | 2 | 3.3892 | | 3.8038 |
| 35 | 20 | 1 ex. | 5+ 5=10 | 1 d. | 17+13=30 | | 28+11=39 | | 33+ 4=37 | 1 ex. | 38+ 0=38 | | 36+ 0=36 | | 29+ 1=30 | 35 | .01836 | .3672 | .1836 | .5508 | .7160 | .6793 | .6976 | .6609 | .5508 | 35 | 1 | 3.8554 | | 4.4062 |
| 36 | 18 | 1 ex. | 19+ 3=22 | | 9+16=25 | | 30+ 9=39 | 1 d. | 39+ 5=44 | 1 d. | 36+ 2=38 | 1 ex. | 38+ 1=39 | | 36+ 0=36 | 36 | .01871 | .3367 | .4116 | .4677 | .7296 | .8232 | .7109 | .7296 | .6735 | 36 | 2 | 4.2093 | | 4.8828 |
| 37 | 15 | | 17+ 7=24 | | 22+18=40 | | 25+10=35 | 1 d. | 38+ 2=40 | | 43+ 4=47 | 3 d. | 37+ 0=37 | 1 ex. | 39+ 2=41 | 37 | .01906 | .2859 | .4574 | .7624 | .6671 | .7624 | .8958 | .7052 | .7814 | 37 | 4 | 4.5362 | | 5.3176 |
| 38 | 10 | | 15+10=25 | 1 d. | 24+11=35 | | 40+ 7=47 | | 34+ 2=36 | 1 ex. | 40+ 1=41 | 2 d. | 44+ 2=46 | | 36+ 1=37 | 38 | .01943 | .1943 | .4857 | .6800 | .9132 | .6994 | .7966 | .8937 | .7189 | 38 | 3 | 4.6629 | | 5.3818 |
| 39 | 16 | | 10+15=25 | | 24+20=44 | | 35+ 8=43 | 1 d. | 47+ 4=51 | | 35+ 5=40 | | 39+ 3=42 | | 46+ 3=49 | 39 | .01982 | .3171 | .4955 | .8720 | .8522 | 1.0108 | .7928 | .8324 | .9711 | 39 | 1 | 5.1728 | | 6.1439 |
| 40 | 13 | | 16+17=33 | | 25+21=46 | 1 d. | 44+15=59 | | 42+ 8=50 | 2 d. | 51+ 6=57 | | 40+ 3=43 | 1 ex. | 42+ 3=45 | 40 | .02022 | .2628 | .6672 | .9301 | 1.1929 | 1.0110 | 1.1525 | .8694 | .9099 | 40 | 3 | 6.0859 | | 6.9958 |
| | 124 | | 191 | | 323 | | 393 | | 383 | | 373 | | 338 | | 331 | | | 2.2734 | 3.5489 | 5.9361 | 7.1902 | 7.0273 | 6.8765 | 6.2399 | 6.1055 | | 23 | 39.0923 | | 45.1978 |
| 41 | 15+ 6=21 | | 13+17=30 | | 33+12=45 | | 45+ 7=52 | 2 d. | 59+ 4=63 | 1 d. | 48+ 0=48 | | 57+ 0=57 | | 42+ 1=43 | 41 | .02064 | .4334 | .6192 | .9288 | 1.0732 | 1.3003 | .9907 | 1.1704 | .8875 | 41 | 3 | 6.5220 | | 7.4095 |
| 42 | 13+ 5=18 | | 21+11=32 | | 30+13=43 | | 45+ 8=53 | 1 d. | 50+ 3=53 | | 62+ 1=63 | | 48+ 0=48 | 1 d. | 57+ 0=57 | 42 | .02342 | .4215 | .7494 | 1.0070 | 1.2412 | 1.2412 | 1.4754 | 1.1241 | 1.3349 | 42 | 2 | 7.2598 | | 8.5947 |
| 43 | 9+ 4=13 | | 18+14=32 | | 32+10=42 | | 43+ 0=43 | 1 d. | 52+ 1=53 | 1 d.<br>1 ex. | 53+ 1=54 | 1 d. | 63+ 1=64 | | 47+ 1=48 | 43 | .02398 | .3117 | .7673 | 1.0071 | 1.0311 | 1.2709 | 1.2949 | 1.5347 | 1.1510 | 43 | 3 | 7.2177 | | 8.3687 |
| 44 | 4+ 1= 5 | 1 d. | 13+ 9=22 | | 32+10=42 | 1 d. | 42+ 3=45 | | 42+ 0=42 | 1 d. | 51+ 0=51 | 1 ex. | 53+ 1=54 | 1 d. | 64+ 0=64 | 44 | .02457 | .1228 | .5405 | 1.0319 | 1.1056 | 1.0319 | 1,2530 | 1.3267 | 1.5724 | 44 | 4 | 6.4124 | | 7.9848 |
| 45 | 11+ 4=15 | 1 d. | 4+14=18 | 1 d. | 22+ 7=29 | 1 d. | 41+ 3=44 | 1 d. | 45+ 0=45 | 1 d. | 41+ 0=41 | | 50+ 0=50 | 2 d. | 53+ 0=53 | 45 | .02518 | .3777 | .4532 | .7302 | 1.1079 | 1.1331 | 1.0323 | 1.2590 | 1.3345 | 45 | 7 | 6.0934 | | 7.4279 |
| 46 | 7+ 3=10 | | 14+15=29 | | 17+ 5=22 | | 28+ 5=33 | | 43+ 1=44 | 1 d. | 44+ 0=44 | 1 d. | 41+ 2=43 | 1 d. | 48+ 2=50 | 46 | .02584 | .2584 | .7493 | .5684 | .8527 | 1.1369 | 1.1369 | 1.1111 | 1.2920 | 46 | 3 | 5.8137 | | 7.1057 |
| 47 | 3+ 1= 4 | | 10+12=22 | | 29+ 8=37 | 1 d. | 22+ 5=27 | | 33+ 1=34 | | 43+ 1=44 | 1 d. | 43+ 2=45 | | 42+ 0=42 | 47 | .02652 | .1060 | .5834 | .9812 | .7160 | .9016 | 1.1668 | 1.1934 | 1.1138 | 47 | 2 | 5.6484 | | 6.7622 |
| 48 | 4+ 2= 6 | | 4+ 8=12 | | 22+16=38 | 1 d. | 36+ 8=44 | 1 d. | 27+ 0=27 | 1 d. | 34+ 0=34 | 1 d. | 43+ 0=43 | 1 d. | 45+ 0=45 | 48 | .02724 | .1634 | .3268 | .9261 | 1.1985 | .7354 | .9261 | 1.1713 | 1.2258 | 48 | 5 | 5.4476 | | 6.6734 |
| 49 | 4+ 1= 5 | | 6+ 7=13 | | 12+15=27 | | 37+ 2=39 | | 43+ 0=43 | 1 d. | 26+ 0=26 | 1 d. | 33+ 0=33 | | 42+ 0=42 | 49 | .03081 | .1540 | .4005 | .8318 | 1.2015 | 1.3248 | .8010 | 1.0167 | 1.2940 | 49 | 2 | 5.7303 | | 7.0243 |
| 50 | 2+ 1= 3 | | 5+ 9=14 | | 13+14=27 | | 27+13=40 | | 39+ 1=40 | 1 d. | 42+ 0=42 | 1 d. | 25+ 1=26 | | 33+ 0=33 | 50 | .03179 | .0953 | .4450 | .8583 | 1.2716 | 1.2716 | 1.3351 | .8265 | 1.0490 | 50 | 2 | 6.1034 | | 7.1524 |
| | 100 | | 224 | | 352 | | 420 | | 444 | | 447 | | 463 | | 477 | | | 2.4442 | 5.6346 | 8.8708 | 10.7993 | 11.3477 | 11.4121 | 11.7399 | 12.2549 | | 33 | 62.2487 | | 74.5036 |
| 51 | 1 | | 3+ 4= 7 | | 14+12=26 | | 27+16=43 | | 40+ 2=42 | | 39+ 0=39 | 1 d. | 41+ 0=41 | 2 d. | 26+ 0=26 | 51 | .03280 | .0328 | .2296 | .8528 | 1.4104 | 1.3776 | 1.2792 | 1.3448 | .8528 | 51 | 3 | 6.5272 | | 7.3800 |
| 52 | 1 | | 1+ 4= 5 | | 7+ 8=15 | | 26+12=38 | | 43+ 0=43 | | 42+ 2=44 | 1 d. | 38+ 0=38 | 1 d. | 39+ 1=40 | 52 | .0339 | .0339 | .1695 | .5085 | 1.2882 | 1.4577 | 1.4916 | 1.2882 | 1.3560 | 52 | 2 | 6.2376 | | 7.5936 |
| 53 | 0 | | 1+ 3= 4 | | 5+ 8=13 | 1 d. | 15+13=28 | | 38+ 2=40 | | 43+ 0×43 | | 43+ 3=46 | | 37+ 0=37 | 53 | .0351 | — | .1404 | .4563 | .9828 | 1.4040 | 1.5093 | 1.6146 | 1.2987 | 53 | 1 | 6.1074 | | 7.4061 |
| 54 | 1 | | 0+ 1= 1 | | 4+10=14 | | 12+10=22 | 1 d. | 28+ 4=32 | | 40+ 1=41 | 1 d. | 43+ 1=44 | | 46+ 1=47 | 54 | .0331 | .0331 | .0331 | .4634 | .7282 | 1,0592 | 1.3571 | 1.4564 | 1.5557 | 54 | 2 | 5.1305 | | 6.6862 |
| 55 | + 1 | | 1+ 1= 2 | | 1+ 7= 8 | | 14+ 6=20 | 1 d. | 21+ 1=22 | | 32+ 2=34 | 2 d. | 40+ 1=41 | 1 d. | 44+ 0=44 | 55 | .0342 | .0342 | .0684 | .2736 | .6840 | .7524 | 1.1628 | 1.4022 | 1.5048 | 55 | 4 | 4.3776 | | 5.8824 |
| 56 | 0 | | 1+ 2= 3 | | 2+ 6= 8 | | 8+ 4=12 | 1 d. | 19+ 3=22 | 1 ex. | 22+ 1=23 | 1 d. | 32+ 1=33 | | 40+ 1=41 | 56 | .0354 | — | .1062 | .2832 | .4248 | .7788 | .8142 | 1.1682 | 1.4514 | 56 | 2 | 3.5754 | | 5.0268 |
| 57 | 0 | | 0+ 1= 1 | | 3+ 2= 5 | | 8+ 3=11 | | 11+ 1=12 | | 21+ 2=23 | | 22+ 3=25 | 1 d. | 33+ 1=34 | 57 | .0367 | — | .0367 | .1835 | .4037 | .4404 | .8441 | .9175 | 1.2478 | 57 | 1 | 2.8259 | | 4.0737 |
| 58 | 1 | | 0+ 1= 1 | | 1+ 0= 1 | | 5+ 4= 9 | | 11+ 2=13 | | 12+ 1=13 | 1 d. | 23+ 1=24 | 1 ex.<br>1 d. | 24+ 0=24 | 58 | .0381 | .0381 | .0381 | .0381 | .3429 | .4953 | .4953 | .9144 | .9144 | 58 | 2 | 2.3622 | | 3.2766 |
| 59 | + 1 | | 1+ 2= 3 | | 1+ 0= 1 | | 1+ 0= 1 | | 9+ 0= 9 | | 13+ 1=14 | | 12+ 0=12 | | 22+ 0=22 | 59 | .0396 | .0396 | .1188 | .0396 | .0396 | .3564 | .5544 | .4752 | .8712 | 59 | | 1.6236 | | 2.4948 |
| 60 | 0 | | 1+. 1= 2 | | 3+ 8=11 | | 1+ 6= 7 | | 1+ 1= 2 | | 9+ 2=11 | | 14+ 0=14 | | 12+ 2=14 | 60 | .0413 | — | .0826 | .4543 | .2891 | .0826 | .4543 | .5782 | .5782 | 60 | | 1.9411 | | 2.5193 |
| | 6 | | 29 | | 102 | | 191 | | 237 | | 285 | | 318 | | 329 | | | .2117 | 1.0234 | 3.5533 | 6.5937 | 8.2044 | 9.9623 | 11.1597 | 11.6310 | | 17 | 40.7085 | | 52.3395 |
| 61 | | | 0+ 1= 1 | | 2+ 3= 5 | | 11+ 0=11 | 1 d. | 7+ 1= 8 | | 2+ 0= 2 | 1 d. | 11+ 1=12 | 1 d. | 14+ 0=14 | 61 | .0431 | — | .0431 | .2155 | .4741 | .3448 | .0862 | .5172 | .6034 | 61 | 3 | 1.6809 | | 2.2843 |
| 62 | | | 0+ 0= 0 | | 1+ 2= 3 | | 5+ 2= 7 | 1 ex. | 10+ 2=12 | 1 ex. | 8+ 0= 8 | | 1+ 3= 4 | | 11+ 0=11 | 62 | .04504 | — | — | .1351 | .3152 | .5404 | .3603 | .1801 | .4954 | 62 | | 1.5311 | | 2.0265 |
| 63 | | | 0+ 0= 0 | | 0+ 0= 0 | | 3+ 0= 3 | | 6+ 0= 6 | | 11+ 0=11 | | 8+ 0= 8 | 1 d. | 4+ 0= 4 | 63 | .04717 | — | — | — | .1415 | .2830 | .5188 | .3773 | .1886 | 63 | 1 | 1.3206 | | 1.5092 |
| 64 | | | 0+ 2= 2 | | 0+ 0= 0 | | 0+ 1= 1 | 1 d. | 3+ 0= 3 | | 6+ 1= 7 | | 11+ 0=11 | | 7+ 0= 7 | 64 | .04950 | — | .0990 | — | .0495 | .1485 | .3465 | .5445 | .3465 | 64 | 1 | 1.1880 | | 1.5345 |
| 65 | | | 0+ 1= 1 | | 2+ 0= 2 | | 0+ 0= 0 | | 0+ 0= 0 | | 3+ 1= 4 | 1 ex. | 7+ 0= 7 | | 11+ 0=11 | 65 | .05208 | — | .0520 | .1041 | — | — | .2083 | .3645 | .5728 | 65 | | .7289 | | 1.3017 |
| 66 | | | 4 | | 1+ 0= 1 | 1 ex. | 2+ 0= 2 | | 0+ 1= 1 | | 0+ 0= 0 | | 3+ 0= 3 | | 7+ 0= 7 | 66 | .05494 | — | — | .0549 | .1098 | .0549 | — | .1648 | .3845 | 66 | | .3844 | | .7689 |
| 67 | | | | | 11 | | 0+ 0= 0 | | 2+ 0= 2 | | 1+ 0= 1 | | 0+ 0= 0 | | 3+ 0= 3 | 67 | .05814 | — | — | — | — | .1162 | .0581 | — | .1744 | 67 | | .1743 | | .3487 |
| 68 | | | | | | | + 1= 1 | | 0+ 0= 0 | | 2+ 0= 2 | | 1+ 0= 1 | | 0+ 0= 0 | 68 | .06172 | — | — | — | .0617 | — | .1234 | .0617 | — | 68 | | .2468 | | .2468 |
| 69 | | | | | | | 25 | | 1+ 0= 1 | | 0+ 0= 0 | | 2+ 0= 2 | | 1+ 0= 1 | 69 | .06578 | — | — | — | — | .0657 | — | .1356 | .0657 | 69 | | .2013 | | .2670 |
| 70 | | | | | | | | | 33 | | 1+ 0= 1 | | 0+ 0= 0 | | 2+ 0= 2 | 70 | .07746 | — | — | — | — | — | .0774 | — | .1549 | 70 | | .0774 | | .2323 |
| 71 | | | | | | | | | | | 36 | | 1+ 0= 1 | | 0+ 0= 0 | | | | .1941 | .5096 | 1.1518 | 1.5535 | 1.7790 | 2.3457 | 2.9862 | | 5 | 7.5337 | | 10.5199 |
| 72 | | | | | | | | | | | | | 49 | | 1+ 0= 1 | | Totals at Chriſtmas 1774. | 5.4868 | 11,1261 | 20.6622 | 28.5202 | 31.1352 | 32,2372 | 33.7143 | 34.7741 | Totals | 93 | 162.8820 | | 197.6561 |
| 73 | | | | | | | | | | | | | | | 61 | | | | | | | | | | | | | | | |
| Admitted | 274 | | 240 | | 433 | | 321 | | 88 | | 50 | | 42 | | 21 | in all | 1469 | | | | | | | | | | | | | |
| Deaths | | 4 | | 6 | | 9 | | 21 | | 17 | | 21 | | 15 | | in all | 93 | | | | | | | | | | | | | |
| Excluſions | | 5 | | 3 | | 2 | | 4 | | 17 | | 8 | | 5 | | in all | 44 | | | | | | | | | | | | | |
| Total Deaths & Excl. | | 9 | | 9 | | 11 | | 25 | | 34 | | 29 | | 20 | | in all | 137 | | | | | | | | | | | | | |
| N°. Deaths *expected* | | 5.4868 | | 11.1261 | | 20.6622 | | 28.5202 | | 31.1352 | | 32.2372 | | 33.7143 | | in all | 162.8820 | | | | | | | | | | | | | |

| N°. of Deaths in SEVEN Years | Real | Expected |
|---|---|---|
| Deaths under 21 Years of Age | 4 | 2.3084 |
| D°. from 21 to 30 incluſive | 11 | 10.9904 |
| D°. from 31 to 40 incluſive | 23 | 39.0923 |
| D°. from 41 to 50 incluſive | 33 | 62.2487 |
| D°. from 51 to 60 incluſive | 17 | 40.7085 |
| D°. upwards of 60 Years of Age | 5 | 7.5337 |
| Totals | 93 | 162.8820 |

| State of the SOCIETY at Chriſtmas 1774. | | |
|---|---|---|
| | Living Members | 1332 |
| | Deaths and Excluſions | 137 |
| | Five Papiſts, and Three unentered | 8 |
| | Final Number in the Society's Book | 1477 |

| | |
|---|---|
| From the Number expected to die in *Eight* Years | 197.6561 |
| Subtract the Number expected to die in *Seven* Years | 162.8820 |
| The Number expected to die Anno 1775, is | 34.7741 |

Printed for the SUPPLEMENT to DALE's CALCULATIONS of the Value of ANNUITIES, publiſhed by J. RIDLEY, in St. James's-Street.

to CHRISTMAS 1775.

| Multipliers. | 1771. |
|---|---|
| | — |
| | — |
| .145 | .190 |
| .0666 | .036 |
| .0476 | — |
| .0368 | — |
| .03005 | .034 |
| .0253 | — |
| .0173 | .013 |
| .0147 | .012 |
| .01343 | — |
| .01210 | .018 |
| .01072 | — |
| .00928 | .009 |
| .00937 | .028 |
| .00946 | .009 |
| .00955 | .038 |
| .00964 | .039 |
| .00974 | .039 |
| .00983 | .040 |
| .00993 | .511 |
| .01003 | |
| .01013 | .050 |
| .01104 | .131 |

| N°. of Deaths in Seven Years | Expected | Real |
|---|---|---|
| Deaths under 21 Years of Age — | 2.3084 | 4 |
| D°. from 21 to 30 inclusive — | 10.9904 | 11 |
| D°. from 31 to 40 inclusive — | 39.0923 | 23 |
| D°. from 41 to 50 inclusive — | 62.2487 | 33 |
| D°. from 51 to 60 inclusive — | 40.7085 | 17 |
| D°. upwards of 60 Years of Age — | 7.5337 | 5 |
| Totals | 162.882[illegible] | 93 |

| Multipliers. | 1768. | 1769. | 1770. | 1771. |
|---|---|---|---|---|
| | | | | .511 |
| .01013 | .0202 | — | .0505 | .050 |
| .01104 | .0368 | .0477 | .0238 | .131 |

# Direct standardization

F.G.P. Neison (1844). On a method recently proposed for conducting inquiries into the comparative sanatory condition of various districts, with illustrations, derived from numerous places in Great Britain at the period of the last census. *J. Statist. Soc. (Lond.)* **7**, 40-68.

# *On a Method recently proposed for conducting Inquiries into the Comparative Sanatory Condition of various Districts, with Illustrations, derived from numerous places in Great Britain at the period of the last Census.* By F. G. P. Neison, F.L.S., &c.

[*Read before the Statistical Society of London, January* 15*th*, 1844.]

The present contribution has been made in consequence of the discussion which followed the reading of Mr. Chadwick's paper, at the last meeting of this Society. A new method of measuring the comparative value of life in various districts, was proposed in that communication, in the following terms: "If the ages of any class, or of the general population, living in any district, and the ages of those who die were reduced to the simplest proportions; that is, if the total years of age whether of the living or dying, were divided by the total number of individuals from which the returns were made, the public would be enabled to make comparisons between district and district, and to judge of the relative degrees of pressure in each, of the causes of mortality." It is also stated "That the average ages of death are found to maintain a comparatively steady course, always nearest to the actual condition of the population, and give the most sure indications." And that "the chief test of the pressure of the causes of mortality is the duration of life in years; and whatever age may be taken as the standard of the natural age or the average age of the individual in any community may be taken to correct the returns of the proportions of death in that same community."*

It appears to me that there will be little difficulty in showing that the method proposed is fallacious in principle, and that its practical application to vital statistics will produce contradictory results; and that consequently it cannot be used as a test of the sanatory condition of a community.

J. Statist. Soc. 7, 40-68, 1844.

On this series of tables is founded another table, which I would beg also to submit to consideration. It exhibits the average age at death in the preceding 12 principal towns and counties, also what would have been the average age at death if placed under the same population as the metropolis.

*Effects of the irregularities of distribution of Population according to Ages in the under-mentioned Districts, arranged to show the fallacy of proposing the average Age at Death of the whole Population, as a Test of the Sanatory Condition of a Locality.*

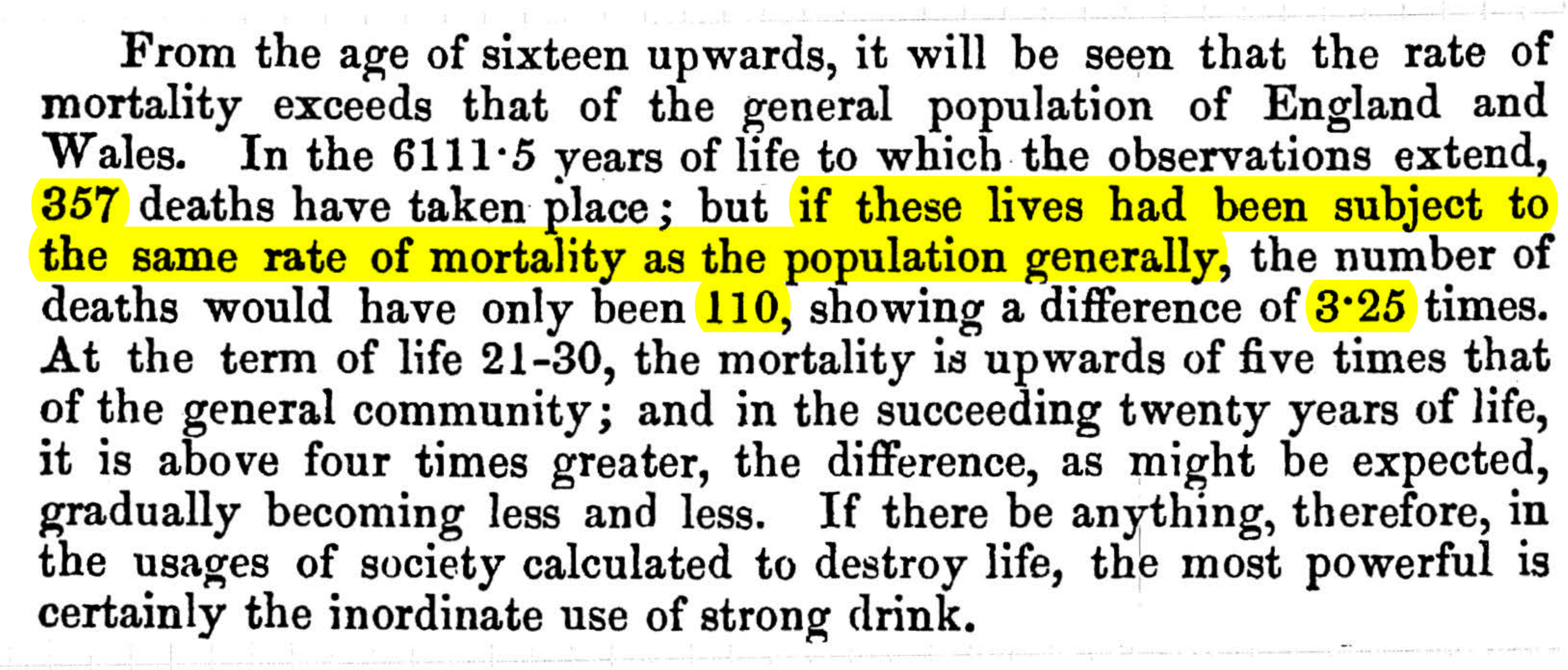
| District. | Average Age at Death.* | | Number out of which one will die. | | Mortality per Cent. | |
|---|---|---|---|---|---|---|
| | Actual. | Transferred. | Actual. | Transferred. | Actual. | Transferred. |
| Metropolis | 29·06 | .. | 39·10 | .. | 2·55 | .. |
| Liverpool | 20·67 | 25·07 | 35·36 | 34·92 | 2·82 | 2·58 |
| Sheffield | 23·19 | 28·14 | 28·51 | 29·28 | 3·50 | 3·75 |
| Manchester | 22·86 | 27·37 | 40·76 | 39·93 | 2·45 | 2·50 |
| Birmingham | 23·70 | 26·82 | 48·65 | 50·63 | 2·06 | 1·97 |
| Leeds | 22·51 | 26·01 | 33·59 | 35·44 | 2·95 | 2·82 |
| City of Exeter | 30·56 | 26·24 | 39·88 | 41·79 | 2·50 | 2·39 |
| Devon | 37·97 | 31·48 | 57·14 | 66·57 | 1·75 | 1·50 |
| Essex | 30·82 | 28·55 | 51·12 | 56·34 | 1·95 | 1·77 |
| Suffolk | 33·24 | 29·51 | 48·96 | 54·57 | 2·04 | 1·83 |
| Norfolk | 31·80 | 26·71 | 48·82 | 56·38 | 2·04 | 1·77 |
| Hereford | 38·42 | 30·54 | 57·86 | 68·49 | 1·72 | 1·46 |

This table contains some interesting results, and I beg to cite two cases. The average age at death in the metropolis is 29·06 years, but in the town of Sheffield it is only 23·19 years; however, if Sheffield were placed under the population of the metropolis the average age at death would be raised to 28·14 years, approaching close to the metropolis.

Another method of viewing this question would be to apply the same rate of mortality to different populations. The populations of the same place at different periods will be found to change in their distribution over the various periods of life, in the same way that distinct districts, as before noticed, do at the same period of time.

# Neison on alcoholism

F.G.P. Neison (1851). On the rate of mortality among persons of intemperate habits. *J. Stat. Soc. Lond.* **14**, 200-219.

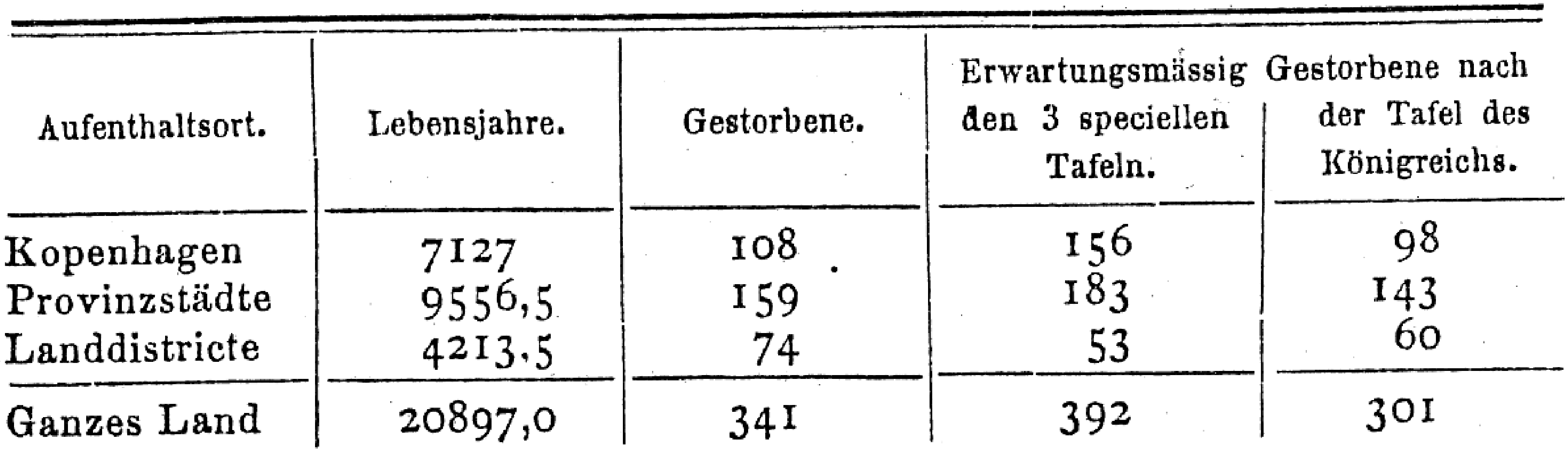


From the age of sixteen upwards, it will be seen that the rate of mortality exceeds that of the general population of England and Wales. In the 6111·5 years of life to which the observations extend, 357 deaths have taken place; but if these lives had been subject to the same rate of mortality as the population generally, the number of deaths would have only been 110, showing a difference of 3·25 times. At the term of life 21-30, the mortality is upwards of five times that of the general community; and in the succeeding twenty years of life, it is above four times greater, the difference, as might be expected, gradually becoming less and less. If there be anything, therefore, in the usages of society calculated to destroy life, the most powerful is certainly the inordinate use of strong drink.

SMR = 3.25

# Westergaard: method of expected number of deaths

H. Westergaard (1882). Die Lehre von der Mortalität und Morbilität. Jena: Fischer



| Aufenthaltsort. | Lebensjahre. | Gestorbene. | Erwartungsmässig Gestorbene nach | |
|---|---|---|---|---|
| | | | den 3 speciellen Tafeln. | der Tafel des Königreichs. |
| Kopenhagen | 7127 | 108 | 156 | 98 |
| Provinzstädte | 9556,5 | 159 | 183 | 143 |
| Landdistricte | 4213,5 | 74 | 53 | 60 |
| Ganzes Land | 20897,0 | 341 | 392 | 301 |

Man sieht hieraus, wie schwierig es ist, eine wissenschaftliche statistische Berechnung zu unternehmen. Die zwei Methoden dürften beide richtig scheinen, und doch geben sie ausserordentlich verschiedene Resultate. Nach der einen Methode würde man schliessen, dass der ärztliche Stand unter sehr ungünstigen sanitären Verhältnissen lebe, nach der anderen, dass der Gesundheitszustand verhältnissmässig gut sei.

Die Schwierigkeit lag darin, dass es zwei Ursachen gab: die ärztliche Profession und der Aufenthaltsort; beiden Ursachen muss man gerecht werden, und wenn man von der einen, dem Aufenthaltsorte, absieht und nur mit Hülfe der allgemeinen Sterblichkeitstafel den Einfluss der anderen betrachtet, begeht man einen fehlerhaften Schluss.

Am sichersten ist es, die Theilung des Materials fortzusetzen, bis es keine störende Ursachen mehr giebt; hat man keinen anderen Beweis, so wird man ein sicheres Kennzeichen, dass dies erreicht ist, darin finden, dass eine erneuerte Theilung des Materials nicht die Resultate ändert.

|  |  |  | Expected Number of Deaths according to |  |
|---|---|---|---|---|
|  | Years at risk | Dead | the 3 special districts | the whole country |
| Copenhagen | 7127 | 108 | 156 | 98 |
| Provincial towns | 9556.5 | 159 | 183 | 143 |
| Rural districts | 4213.5 | 74 | 53 | 60 |
| Whole country | 20897.0 | 341 | 392 | 301 |

It is seen from this how difficult it is to conduct a scientific statistical calculation. The two methods both look correct, and still yield very different results. According to one method one would conclude that the medical profession live under very unhealthy conditions, according to the other, that their health is relatively good.

The difficulty derives from the fact that there *exist two causes:* the medical profession and the place of residence; both causes have to be taken into account, and if one neglects one of them, the place of residence, and only with the help of the general life table considers the influence of the other, one will make an erroneous conclusion.

The safest is to continue the stratification of the material until no further disruptive causes exist; if one has no other proof, then a safe sign that this has been achieved, is that further stratification of the material does not change the results.

In Kopenhagen ist die Sterblichkeit etwas grösser gewesen, als im übrigen Dänemark, aber nicht so viel grösser, dass nicht zufällige Ursachen den Unterschied erklären könnten. Auf dem Lande ist die Sterblichkeit grösser, als erwartungsmässig, aber auch hier lässt sich nicht entscheiden, ob diese höhere Sterblichkeit einer verschiedenen Norm zuzuschreiben ist. Um dieses nun noch besser zu erkennen, können wir die Anzahl der erwartungsmässigen Todesfälle nach der Tabelle des Königreichs um 13,2 Procent erhöhen. Wir erhalten dann folgendes Resultat:

| Aufenthaltsort | Erfahrung | Berechnung |
|---|---|---|
| Kopenhagen | 108 | 110,94 |
| Provinzstädte | 159 | 161.60 |
| Landdistricte | 74 | 68,50 |
| Das ganze Land | 341 | 341 |

Die Abweichungen sind so unwesentlich, dass man sie mit gutem Grunde als eine Wirkung zufälliger Ursachen betrachten kann, und man darf ohne grosses Risico die Sterblichkeit in allen drei Landestheilen, wenigstens in Kopenhagen und den Städten, für gleich gross annehmen.

Wenn wir jetzt, wie wir es gelernt haben, irgend eine Ursache, z. B. das Alter, durch Anwendung der Methode der erwartungsmässig Gestorbenen eliminiren, so muss der mittlere Fehler die Form:

$$M = \sqrt{S \cdot p \cdot q \ldots} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (43)$$

erhalten, wenn das Alter dadurch zu einer zufälligen Ursache reducirt worden ist; in der Formel (43) wird $S$ die Anzahl der Lebensjahre, $p$ die Wahrscheinlichkeit zu sterben und $q$ die Wahrscheinlichkeit zu leben bedeuten. Aber in der That hat der mittlere Fehler die Form:

$$M = \sqrt{a_1 p_1 q_1 + a_2 p_2 q_2 + \ldots .} \quad . \quad . \quad . \quad . \quad . \quad (44)$$

Diese beiden Formeln werden indessen unter gewissen Voraussetzungen identisch. Wenn man z. B. berechtigt ist, $q$ hinwegzunehmen, werden beide Ausdrücke für den mittleren Fehler zur Quadratwurzel der Gestorbenen reducirt. Mit Annäherung, aber auch nur mit Annäherung, wird also das Alter als eine zufällige Ursache betrachtet werden können. Wie mit dem Alter, so geht es mit allen anderen Factoren, die in die Sterblichkeit eingreifen, so mit den Vermögensverhältnissen, Erwerb, Wohnung, Kleidung u. s. w. u. s. w. — alle üben ihre Wirkungen, und es ist unmöglich sie gänzlich hinwegzuschaffen; selbst dann, wenn zwei Bevölkerungen auf dieselbe Weise, verhältnissmässig aus derselben Anzahl von Armen und Reichen, Trunkenbolden und Nüchternen, zusammengesetzt sind, werden doch alle diese Ursachen vorhanden sein, und es ist nur eine Annäherung an die Wahrheit, wenn man davon spricht, dass sie eliminirt sind, eine Annäherung, deren Berechtigung mehr oder minder auf subjektivem Ermessen beruht.

# Standard error of the expected number of deaths

If we now, as we have learnt about it, have eliminated some cause, e.g. age, through application of the method of expected number of deaths, then the standard error will take the form

$$M = \sqrt{S \cdot p \cdot q}$$

since age has been reduced to a random cause.

[random cause: zufällige Ursache]

Standard error of SMR reinvented by Yule (1934).

# Yule: Association measures

(Galton-)Pearson correlation coefficient was forcefully marketed by Karl Pearson

G.U. Yule (1900). On the association of attributes in statistics: with illustrations from the material of the Childhood Society, &c. *Phil.Trans.Roy.Soc.Lond. A*, **194**, 257-319.

$2 \times 2$ table

|   |   |
|---|---|
| a | b |
| c | d |

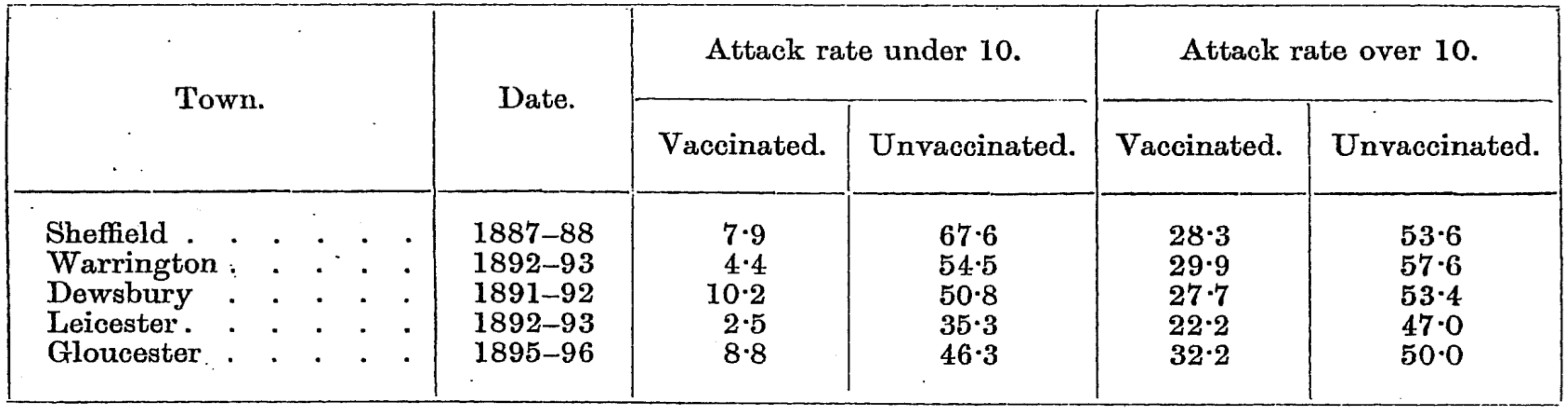
define $Q = \frac{ad - bc}{ad + bc}$

As the correlation coefficient, $Q = 0$ under independence, $= \pm 1$ under total positive/negative dependence.

Note for later use: $Q = \frac{OR - 1}{OR + 1}$ where $OR$ is the odds ratio

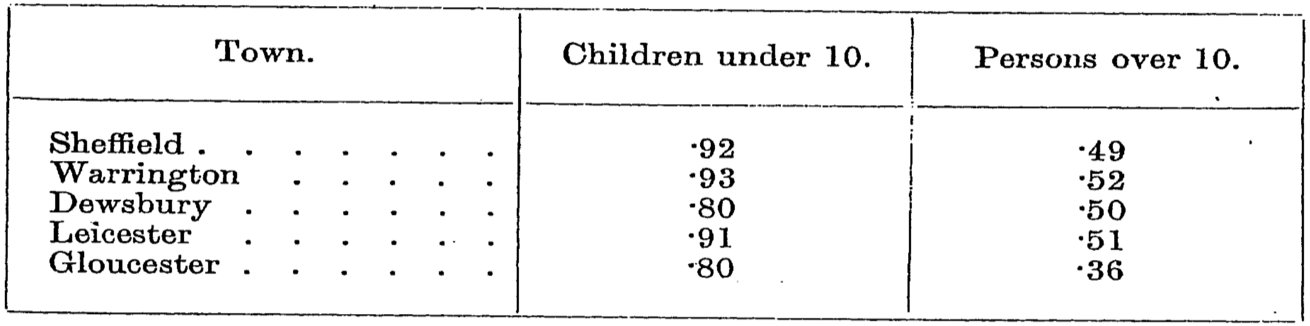
(Pearson preferred the tetrachoric correlation generated by assuming double dichotomy of a bivariate normal. Yule and Pearson fought about this for more than a decade, cf. MacKenzie (1978, 1981).)

# Example: association between smallpox vaccination and attack rate



| Town. | Date. | Attack rate under 10. | | Attack rate over 10. | |
|---|---|---|---|---|---|
| | | Vaccinated. | Unvaccinated. | Vaccinated. | Unvaccinated. |
| Sheffield . . . . . . . | 1887–88 | 7·9 | 67·6 | 28·3 | 53·6 |
| Warrington . . . . . | 1892–93 | 4·4 | 54·5 | 29·9 | 57·6 |
| Dewsbury . . . . . . | 1891–92 | 10·2 | 50·8 | 27·7 | 53·4 |
| Leicester . . . . . . . | 1892–93 | 2·5 | 35·3 | 22·2 | 47·0 |
| Gloucester . . . . . . | 1895–96 | 8·8 | 46·3 | 32·2 | 50·0 |

Association between Non-vaccination and Attack in Infected Households.



| Town. | Children under 10. | Persons over 10. |
|---|---|---|
| Sheffield . . . . . . . . | ·92 | ·49 |
| Warrington . . . . . . | ·93 | ·52 |
| Dewsbury . . . . . . . | ·80 | ·50 |
| Leicester . . . . . . . | ·91 | ·51 |
| Gloucester . . . . . . | ·80 | ·36 |

Notice the widely differing attack rates for Sheffield, Warrington, Leicester, but quite similar values of *Q*.

# Kermack, McKendrick, McKinlay: multiplicative age-cohort model for death rates

W.O. Kermack, A.G. McKendrick, P.L. McKinlay (1934). Death-rates in Great Britain and Sweden: expression of specific mortality rates as products of two factors, and some consequences thereof. *J. Hygiene (Camb.)* **34**, 433-457.

Death-rate at age $\theta$ and time $t$ may be described by

$$f\left(t,\theta\right)=\alpha\left(t,\theta\right)\beta_{\theta}$$

The paper contains careful discussion of estimation procedures (with standard errors) and goodness of fit evaluation.

There is no reference to the lively current discussion of standardization.

Until 1979 this paper was not quoted outside a narrow circle of cohort analysts.

J. Hygiene (Cambridge)

VOLUME XXXIV DECEMBER, 1934 No. 4

433-457

# DEATH-RATES IN GREAT BRITAIN AND SWEDEN: EXPRESSION OF SPECIFIC MORTALITY RATES AS PRODUCTS OF TWO FACTORS, AND SOME CONSEQUENCES THEREOF

BY W. O. KERMACK, A. G. McKENDRICK
AND P. L. McKINLAY

*From the Laboratory of the Royal College of Physicians of Edinburgh,
and the Department of Health for Scotland*

(With 6 Figures in the Text)

CONTENTS

## INTRODUCTION

IN a preliminary paper[1] the specific death-rates of England and Wales, of Scotland, and of Sweden of the various age groups for different years have been analysed, and the chief result which emerged may be stated in the following terms. If $v_{t,\theta}d\theta$ denote the number of persons at a time $t$, between the ages $\theta$ and $\theta+d\theta$, then[2]

$$-\frac{1}{v_{t,\theta}}\left(\frac{\partial v_{t,\theta}}{\partial t}+\frac{\partial v_{t,\theta}}{\partial \theta}\right)=f(t,\theta) \qquad \text{......(1)}$$

is the specific death-rate for the age $\theta$ at the time $t$. It appears from our previous paper that $f(t,\theta)$ may, to a close approximation, be represented as the product of two factors, of which one, $\beta_\theta$, is a function of the age alone, whilst the other, $\alpha(t-\theta)$, is a function of the date of birth $(t-\theta)$. Thus

$$f(t,\theta)=\alpha(t-\theta)\,\beta_\theta \qquad \text{......(1 a).}$$

Clearly both $\alpha$ and $\beta$ are arbitrary to the extent of a multiplying constant. In the case of the statistics of England and Wales, and of Scotland, the deviations from this form appear to be of the nature of random irregularities,

## SUMMARY

1. The specific mortality rates for males, females and the total population for England and Wales, for Scotland and for Sweden, have been fitted to a formula $f(t, \theta) = \alpha(t-\theta)\beta_\theta$, where $f(t, \theta)$ is the specific mortality rate at a time $t$ for age $\theta$, $\beta_\theta$ is a function depending solely on the age $\theta$, and $\alpha(t-\theta)$ depends only on the time of birth $(t-\theta)$. The results are in substantial agreement with those obtained by less refined methods in the previous paper. The probable errors of the values found for $\alpha$ and for $\beta$ have been calculated.

2. It is shown that the $\beta_\theta$ curves for the Scottish and the English males are approximately represented by the Makeham-Gompertz formula $A + Be^{c\theta}$, where $A$, $B$ and $c$ have suitable values. The other $\beta_\theta$ curves do not appear to conform exactly to a formula of this type.

3. With the help of the representation of $\beta_\theta$ by the Makeham-Gompertz expression the effect of variation of $\alpha$ on the survival curves, the death curves, and the expectation of life has been determined. It is shown that with the range of values of $\alpha$ experienced in Britain during the last 50 years, the most marked effect is most likely to be experienced in the future between the ages of 65 and 85, a very considerable increase of people of these ages being likely provided that the relationship exhibited by the statistics up to the present date is maintained in the future.

Though the Makeham-Gompertz formula does not hold in the case of the English and Scottish females, nor for the Swedish statistics, these approximate sufficiently closely to the values for the English and Scottish males, to allow of the conclusion deduced in the latter case being extended to the former.

4. It is strongly emphasised that the validity of all the predictions depends upon a hypothesis of extrapolation which, however attractive in the light of the figures so far available, might not be fulfilled under certain contingencies.

# Kitagawa-decomposition of rate-differences

E.M. Kitagawa (1955). Components of the difference between two rates. *JASA* **50**, 1168-1174.

$$\text{Crude rate}^{A} - \text{Crude rate}^{S} = \sum a_i \alpha_i - \sum s_i \lambda_i$$

$$= \sum \left( a_i - s_i \right) \frac{\alpha_i + \lambda_i}{2} + \sum \left( \alpha_i - \lambda_i \right) \frac{a_i + s_i}{2}$$

| | | | | |
|---|---|---|---|---|
| = | difference in age distribution | weighted with average age specific death rate | difference in age specific death rate | weighted with average age distribution |

= contribution from difference in age distribution + contribution from difference in pattern of death rates

# Miettinen: Components of the crude risk ratio

*Amer. J. Epid.* **96** (1972), 168-172

| | | exposed | non exposed | |
|---|---|---|---|---|
| Cohort study | Events | $e_j$ | $g_j$ | $j=1,\cdots,k$ |
| | Denominator<br>(No. individuals or person-years) | $F_j$ | $H_j$ | |

$$e=\sum e_j,\ F=\sum F_j,\ g=\sum g_j,\ H=\sum H_j,$$

Crude risk ratio $\hat{\rho}_c=\dfrac{e}{F}\Big/\dfrac{g}{H}=eH/gF$

Under no effect in stratum $j$, $e_j$ would be $e_j^*=g_jF_j/H_j$ (expected no. of events in the exposed if non-exposed rates applied) so with $e^*=\sum e_j^*$, the component of $\hat{\rho}_c$ attributable to confounding would be

$$\hat{\rho}^*=e^*H/gF$$

# Components of the crude risk ratio

Repeat: component of $\hat{\rho}_c = e H / gF$ due to confounding is

$$\hat{\rho}^* = e^* H / gF$$

Rate ratio cleaned of confounding by direct standardization based on stratification in the exposed group

$$\hat{\rho}_s = e / e^*$$

and we have

$$\hat{\rho}_c \quad = \quad \underbrace{\hat{\rho}^*}_{\text{due to confounding}} \quad \cdot \quad \underbrace{\hat{\rho}_s}_{\text{due to exposure}}$$

# Mantel and Haenszel

N. Mantel & W. Haenszel (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *J. Nat. Cancer Inst.* **32**, 719-748.

Stratify $2 \times 2$ table into $2 \times 2 \times k$ table

$$\begin{matrix} A_i & B_i \\ C_i & D_i \end{matrix}, \; i = 1, \cdots, k$$

Use as a "summary estimate of the relative risk"

$$\frac{\sum A_i D_i \big/ \left(A_i + B_i + C_i + D_i\right)}{\sum B_i C_i \big/ \left(A_i + B_i + C_i + D_i\right)}$$

which makes sense in a model with odds ratios (or Yule's $Q$) constant across strata.

Likewise, the famous one-degree of freedom Mantel-Haenszel test for unity relative risks can be explained by first assuming common odds ratio, the testing whether this = 1.

# Mantel & Haenszel on the Mantel-Haenszel procedure

*If one could assume that the increased relative risk associated with a factor was constant over all subclassifications, the estimation problem would reduce to weighting the several subclassification estimates according to their respective precisions. The complex maximum likelihood iterative procedure necessary for obtaining such a weighted estimate would seem to be unjustified, since the assumption of a constant relative risk can be discarded as usually untenable.*

# The Standardized Mortality Ratio as maximum likelihood estimator

S.J. Kilpatrick (1962). Occupational mortality indices. *Population Studies* **16**, 175-187.

Population Studies 16, 175-187, 1962

# Occupational Mortality Indices*

### By S. J. KILPATRICK

## INTRODUCTION

The detection of associations between environment and specified causes of death often provides important clues to the ætiology of fatal disease. The mortality experienced by different groups of individuals is best compared, using specific death rates of sub-groups alike in every respect, apart from the single factor by which the total population is divided. This situation is rarely, if ever, realised and we have to be satisfied with mortality comparisons between groups of individuals alike with regard to the two, three or four major factors known to affect the risk of death.

In this paper groups are defined as aggregates of occupations (social classes). It is assumed that age is the only factor related to an individual's mortality within a group. This example may readily be extended to other factors such as sex, marital status, residence, etc. Although the association of social class and age-specific mortality may be evaluated by comparisons between social classes, specific death rates of a social class are more frequently compared with the corresponding rates of the total population. It is this type of comparison which is considered here.

### *Derivation of Weights*

Weights which minimise the variance of $I_{us}$ may be derived as follows. Assume that the $\theta_x$ are homogeneous and that $M_{ux}$ may be used as the risk of death from the cause specified.[8] Then $\theta\, M_{sx}$ is the probability of death in group $u$ between the ages $x$ and $x+n_x$ and the logarithm of the likelihood of the observed series of deaths $D_{ux}$ is

$$L = \text{constant} + \Sigma\, D_{ux} \log\, (\theta\, M_{sx}) + \Sigma\, (P_{ux} - \tfrac{1}{2}\, D_{ux}) \log\, (1 - \theta M_{sx})$$

Hence $\dfrac{\partial L}{\partial \theta} = \Sigma \dfrac{D_{ux}}{\theta} - \Sigma \dfrac{(P_{ux} - \frac{1}{2}\, D_{ux}) M_{sx}}{(1 - \theta\, M_{sx})}$

To the first order of $\theta$, $\frac{1}{2}\, D_{ux}$ may be neglected relative to $P_{ux}$, and $\theta\, M_{sx}$ relative to 1, so that an approximate solution of the maximum likelihood equation $\dfrac{\partial L}{\partial \theta} = 0$ is

$$\hat{\theta} \doteq \frac{\Sigma\, P_{ux}\, M_{ux}}{\Sigma\, P_{ux}\, M_{sx}} = \frac{D_u}{E_u}$$

This first-order estimate of $\theta$ is easily recognised as the usual expression for the S.M.R., being the ratio of the total observed deaths in group $u$ to the total deaths expected in group $u$ by indirect standardisation.[9] Therefore, the " best " system of weights (in the sense of *iv*) is that equal or proportional to $P_{ux}$, the age structure of the group $u$. It follows that the S.M.R. has the smallest standard error of all mortality indices defined by (1).

The variance of $\hat{\theta}$ may be obtained from the second derivative's expectation by

$$(\text{Var } \hat{\theta})^{-1} = -\mathrm{E}\left\{\frac{\partial^2 L}{\partial \theta^2}\right\}$$

$$= \frac{\Sigma D_{ux}}{\hat{\theta}^2} + \Sigma \frac{(P_{ux} - \frac{1}{2} D_{ux})\, M_{sx}^2}{(1 - \hat{\theta}\, M_{sx})^2}$$

$$= \frac{E_u^2}{D_u} + \Sigma\, E_{ux}\, M_{sx} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

neglecting terms of higher order.

Taking the reciprocal gives $D_u / E_u^2$ as a slightly overestimated minimum variance. This, however, is the conventional expression used to estimate the variance of the S.M.R. and is easily derived from (2) by substitution of $w_x = P_{ux}$.

This paper proposes a simple test for heterogeneity in $\theta_x$ and shows that the S.M.R. is equivalent to the maximum likelihood estimate of a common $\theta$ when the $\theta_x$ do not differ significantly. It follows therefore that the S.M.R. has a minimum standard error.

# The National Halothane Study

J. Bunker, W. Forrest Jr., F. Mosteller & L. Vardam (eds.) (1969). *The National Halothane Study*.
National Institute of General Medical Sciences.

Halothane is an anaesthetic which around 1960 was suspected in the U.S. for causing increased rates of hepatic necrosis, sometimes fatal. A large cooperative study was started in July 1963. We shall here focus on the study of ‘surgical deaths’, i.e. deaths during the first 6 weeks after surgery. The study was based on retrospective information from 34 participating medical centres, who reported all surgical deaths during the four years 1959-62 as well as provided information on a random sample of about 38,000 from the total of about 856,000 operations at these centres during the four years. The study was designed and analysed in a collaborative effort between leading biostatisticians at Stanford University, Harvard University and Princeton University/Bell Labs and the report (Bunker et al., 1969) is unusually rich in explicit discussions about how to handle the adjustment problem with the many variables registered for the patients and the corresponding ‘thin’ cross-classifications.

# The National Halothane Study:
# introduction by B.W. Brown et al.

‘the analysis of rates and counts associated with many background variables is a recurring and very awkward problem. …It is appropriate to create new methods for handling this nearly universal problem at just this time. High-speed computers and experience with them have now developed to such a stage that we can afford to execute extensive manipulations repeatedly on large bodies of data with many control variables, whereas previously such heavy arithmetic work was impossible. The presence of the large sample from the National Halothane Study has encouraged the investigation and development of flexible methods of adjusting for several background variables. Although this adjustment problem is not totally solved by the work in this Study, substantial advances have been made and directions for further profitable research are clearly marked.’

# The National Halothane Study: introduction by B.W. Brown et al. (cont.)

‘Pure or complete randomization does not produce either equal or conveniently proportional numbers of patients in each class; attempts at deep post-stratification are doomed to failure because for several variables the number of possible strata quickly climbs beyond the thousands.

…

Insofar as we want rates for special groups, we need some method of estimation that borrows strength from the general pattern of the variables. Such a method is likely to be similar, at least in spirit, to some of those that were developed and applied in this Study.

At some stage in nearly every large-scale, randomized field study (a large, randomized prospective study of postoperative deaths would be no exception), the question arises whether the randomization has been executed according to plan. Inevitably, adjustments are required to see what the effects of the possible failure of the randomization might be. Again, the desired adjustments would ordinarily be among the sorts that we discuss.’

# The National Halothane Study: methods

*Many versions of standardization* (L. Moses, J.W. Tukey et al.)

*'Smoothed' contingency table analysis* (F. Mosteller, Y.M.M. Bishop) leading to the important monograph

Y.M.M. Bishop, S.E. Fienberg & P. Holland (1975). *Discrete Multivariate Analysis: Theory and Practice*. MIT Press, Cambridge, MA.

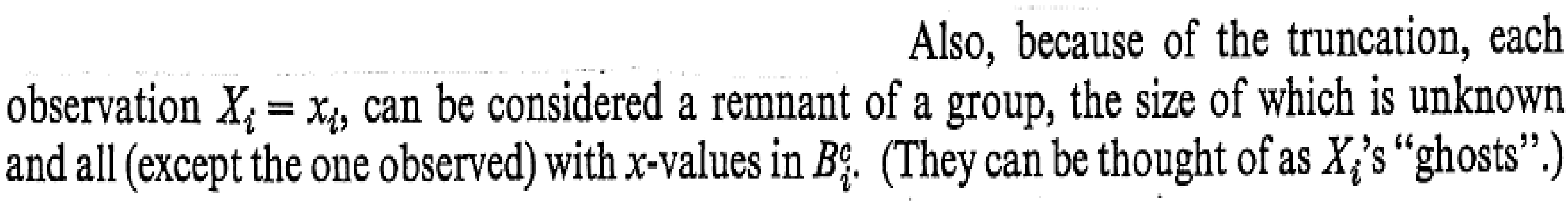
# Loglinear models

S.E. Fienberg (1975). Comment [on paper by Sonja McKinley], *JASA* **70**, 821-823.

*The reader should be aware that standardization is basically a descriptive technique that has been made obsolete, for most of the purposes to which it has traditionally been put, by the ready availability of computer programs for loglinear model analysis of multidimensional contingency tables.*

# Inverse probability weighting: ghosts

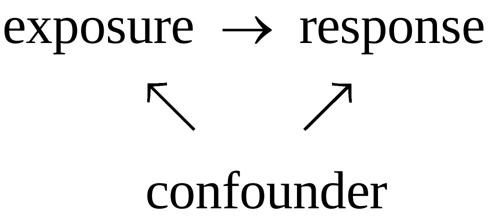
B.W. Turnbull (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *JRSS B* **38**, 290-295.



Also, because of the truncation, each observation $X_i = x_i$, can be considered a remnant of a group, the size of which is unknown and all (except the one observed) with $x$-values in $B_i^c$. (They can be thought of as $X_i$'s "ghosts".)

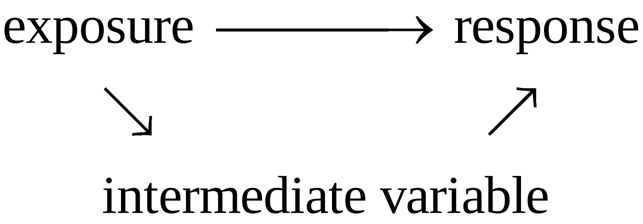
# Weighting methods today: time-dependent confounding

*Conventional wisdom on confounders and intermediate variables*



exposure → response
↖ ↗
confounder

Control for confounder



exposure ⟶ response
↘ ↗
intermediate variable

Do not control for intermediate variable

# Time-dependent confounders, *or* feed-back

Feed-back from outcome $Y_i$ to covariate $X_i$

$$\begin{tikzcd}
 & Y_i \arrow[rr] \arrow[dr] & & Y_{i+1} \\
X_{i-1} \arrow[ur] \arrow[rr] & & X_i \arrow[ur] &
\end{tikzcd}$$

$Y_i$ intermediate between $X_{i-1}$ and $Y_{i+1}$
so do not condition on $Y_i$

$Y_i$ confounder for effect of $X_i$ on $Y_{i+1}$
So “control” for $Y_i$ e.g. by conditioning

# Robins: *g*-computation, marginal structural models

J.M. Robins (1986). A new approach to causal inference in mortality studies with a sustained exposure period – application to control of the healthy worker survivor effect. *Math.Modell.* **7**, 1393-1512.
J.M. Robins, M.A. Hernán, B. Brumback (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology* **11,** 550-560.
R.M. Daniel et al. (2013). Methods for dealing with time-dependent confounding. *Statist. Med.* **32**, 1584-1618.

Robins (1986) generalized direct standardization to the time-dependent confounding situation and formulated other approaches using inverse probability weighting such as marginal structural models (2000). A very readable survey is by Daniel et al. (2013).

The target is here *marginal*, at least within strata, and to me this seems a major reason why marginal effect measures and direct standardization methods are back in business.
(Note: Robins et al. also studied conditional targets via *structural nested models).*

# Simplest possible marginal structural model

Effect of dichotomous exposure $E$ on dichotomous outcome $D$
covariates $z_1, \cdots, z_k$. Assume no unmeasured confounders
given observation of $z_i : P\left(D=1 \middle| \text{set } E=e, Z=z_i\right) = P\left(D=1 \middle| E=e, Z=z_i\right)$.

Marginal structural model

$$P\left(D=1 \middle| \text{set } E=e\right) = e^{\beta_0+\beta_1 e}$$

Estimation via Inverse Probability of Treatment Weights (IPTW).

Each subject $j$ with covariate $z_j$ and exposure $e_j$ is assigned a weight

$$w_j = 1 \Big/ P\left(E=e_j \middle| Z=z_j\right)$$

Let exposed $\sim$ Study group, unexposed $\sim$ Standard group. Then if $z_j = 1$

$$P\left(E=1 \middle| Z=i\right) = A_i \big/ \left(A_i + S_i\right), \;\; P\left(E=0 \middle| Z=i\right) = S_i \big/ \left(A_i + S_i\right)$$

# Derivation of IPTW estimator

$$P\left(D=1 \middle| \text{set } E=e\right)=\sum_z P\left(D=1 \middle| E=e, Z=z\right)P\left(Z=z\right)$$ (no unmeas. conf.)

$$= \quad \sum_z \frac{P\left(D=1, E=e, Z=z\right)P\left(Z=z\right)}{P\left(E=e \middle| Z=z\right)P\left(Z=z\right)}$$

$$= \quad \sum_z \frac{P\left(D=1, E=e, Z=z\right)}{P\left(E=e \middle| Z=z\right)}$$

each individual is weighted with the Inverse Probability of being exposed (Treated).

# Direct and Indirect Standardization in terms of target population

T. Sato & Y. Matsuyama (2003). Marginal structural models as a tool for standardization.
*Epidemiology* **14,** 680-686.

*Study (exposed) Population Target*:

$$SMR = \sum A_i \alpha_i \Big/ \sum A_i \lambda_i$$

compare observed no. deaths to counterfactual no. of deaths if unexposed

*Standard (unexposed) Population Target*:

$$CMF = SRR = \sum S_i \alpha_i \Big/ \sum S_i \lambda_i$$

compare counterfactual no. deaths if exposed to observed no. deaths.

Third possibility: *Total Population Target*:

$$SRR_T = \frac{\sum (A_i + S_i)\alpha_i}{\sum (A_i + S_i)\lambda_i} = \frac{\sum A_i \alpha_i \Big/ \big(A_i \big/ (A_i + S_i)\big)}{\sum S_i \lambda_i \Big/ \big(S_i \big/ (A_i + S_i)\big)}$$

# Total Population Target: Interpretation

$$SRR_T = \frac{\sum (A_i + S_i)\alpha_i}{\sum (A_i + S_i)\lambda_i} = \frac{\sum A_i \alpha_i \Big/ \big(A_i / (A_i + S_i)\big)}{\sum S_i \lambda_i \Big/ \big(S_i / (A_i + S_i)\big)}$$

$$= \frac{\sum \left(\text{deaths in Study Population in age group } i\right) \Big/ \big(A_i / (A_i + S_i)\big)}{\sum \left(\text{deaths in Standard Population in age group } i\right) \Big/ \big(S_i / (A_i + S_i)\big)}$$

numerator: for each age group *I*, deaths in study population weighted with inverse probability of being in study population

denominator: for each age group *I*, deaths in standard population weighted with inverse probability of being in standard population

# Likelihood interpretation of continuous-time marginal structural models

K. Røysland (2011). A martingale approach to continuous-time marginal structural models.
*Bernoulli* **17**, 895-915

Defines two martingale measures:

the observational measure
the randomised trial measure

and assumes the latter absolutely continuous wrt the former.

The resulting likelihood ratio process corresponds to the weights in discrete time marginal structural models.

# Summary

| *Weighting and standardization* | | *Models and parameters* | |
|---|---|---|---|
| 1777-1900 | Prehistory | | (Pearson correlation) |
| | | 1900 | Yule: association measure Q |
| much discussion that I have skipped | | | |
| | | 1934 | Kermack, McKendrick, McKinlay<br>Multiplicative model for rates |
| 1955 | Kitagawa | | |
| 1959 | Mantel & Haenszel | | |
| | | 1962 | Kilpatrick: SMR as MLE |
| *1969 The National Halothane Study* | | | |
| | | 1960s | Loglinear models, logistic regression |
| 1972 | Miettinen | 1972 | Cox |
| 1976 | Turnbull: ghosts | | |
| | | 1970s, 1980s | Models can solve everything |
| 1986 | Robins: time-dependent confounding | | |
| 1999 | Robins: marginal structural models | | |

# Epilogue

To most biostatisticians the conditional approach (stratification, regression analysis) is the preferred one for the daily work analysing observational studies. BUT more complex situations have shown that the parameters of logistic and multiplicative intensity models no longer have interpretation as causal effects (*e.g.* Aalen et al., 2015). This has brought the marginal approach back into the limelight, most spectacularly with the breakthrough by Robins (1986) on time-dependent confounding.

Aalen, O.O., Cook, R.J. & Røysland, K. (2015). Does Cox analysis of a randomized survival study yield a causal treatment effect? *Lifetime Data Analysis* **21,** 579-593.