

HET VERWERKEN VAN KANKERSTERFTESTATISTIEKEN

door J. V. JOOSSENS,

Katholieke Universiteit te Leuven.

De gegevens over kankersterfte die tegenwoordig, vooral dank zij de Wereldgezondheids Organisatie, ter beschikking van de onderzoekers gesteld worden, zijn zo talrijk (gegevens per land, jaartal, geslacht en leeftijd) dat ze totaal onoverzichtelijk worden. Het is dan ook nodig speciale technieken te gebruiken om de informatie uit deze gegevens te halen, in een vorm die biologische waarde heeft.

Kankersterfte gedraagt zich, tussen bepaalde leeftijden die voor de clinicus van overwegend belang zijn (namelijk 40 tot 64 jaar), als een lineaire functie van de leeftijd (op dubbel logaritmische schaal). Hiervan kan gebruik gemaakt worden om een belangrijke reductie van het aantal gegevens te bereiken. Als R_x de sterfte is op leeftijd x dan volgt uit het voorgaande dat

$$R_x = R_1 x^b \quad (1)$$

waar R_1 de berekende sterfte is op leeftijd één en b de helling is van de regressielijn.

Een belangrijk biologisch fenomeen is nu dat $\log R_1$ en b zeer sterk gecorreleerd zijn en dit voor gegevens uit verschillende landen en beide geslachten (Joossens 1963). Nu zal wel met achterdocht opgekeken worden naar een techniek die $\log R_1$ en b correleert, aangezien voor regressielijnen berekend op verschillende « random samples » van éénzelfde land, jaartal en geslacht er een evident louter mathematisch verband bestaat tussen $\log R_1$ en b . Al deze lijnen zullen immers kruisen op $\log R_x$ en $\log x$. Deze critiek geldt echter niet noodzakelijk voor de correlatie tussen $\log R_1$ en b afkomstig van verschillende landen, jaartallen en geslachten. Indien er hier een correlatie bestaat dan kan deze van biologische oorsprong zijn, namelijk door het feit dat de sterfteleeftijdslijnen gemiddeld kruisen op een bepaalde leeftijd. Deze leeftijd wordt dan de isomortaliteitsleeftijd genoemd (Joossens 1963). De regressie tussen $\log R_1$ en b zal twee constanten opleveren namelijk $\log \alpha_1$ en β_1 . Het is

Met de hulp van het Fonds voor Wetenschappelijk Geneeskundig Onderzoek, Brussel.

gemakkelijk in te zien, dat als de correlatie uitgerekend wordt tussen $\log R_x$ en b (met constanten $\log \alpha_x$ en β_x) of in het algemeen tussen $\log R_x$ en b (met constanten $\log \alpha_x$ en β_x), $\alpha_1 = \alpha_x = \alpha_x$, terwijl $\beta_1 = -\log X_m$, waar X_m de gemiddelde isomortaliteitsleeftijd is en

$$\beta_1 = \beta_x - \log x \text{ of } \log X_m = -\beta_x + \log x \quad (2)$$

Meulepas en Joossens* hebben recent aangetoond dat dit immers voortvloeit uit de duale eigenschappen van lijn en punt. Het probleem van een ge-

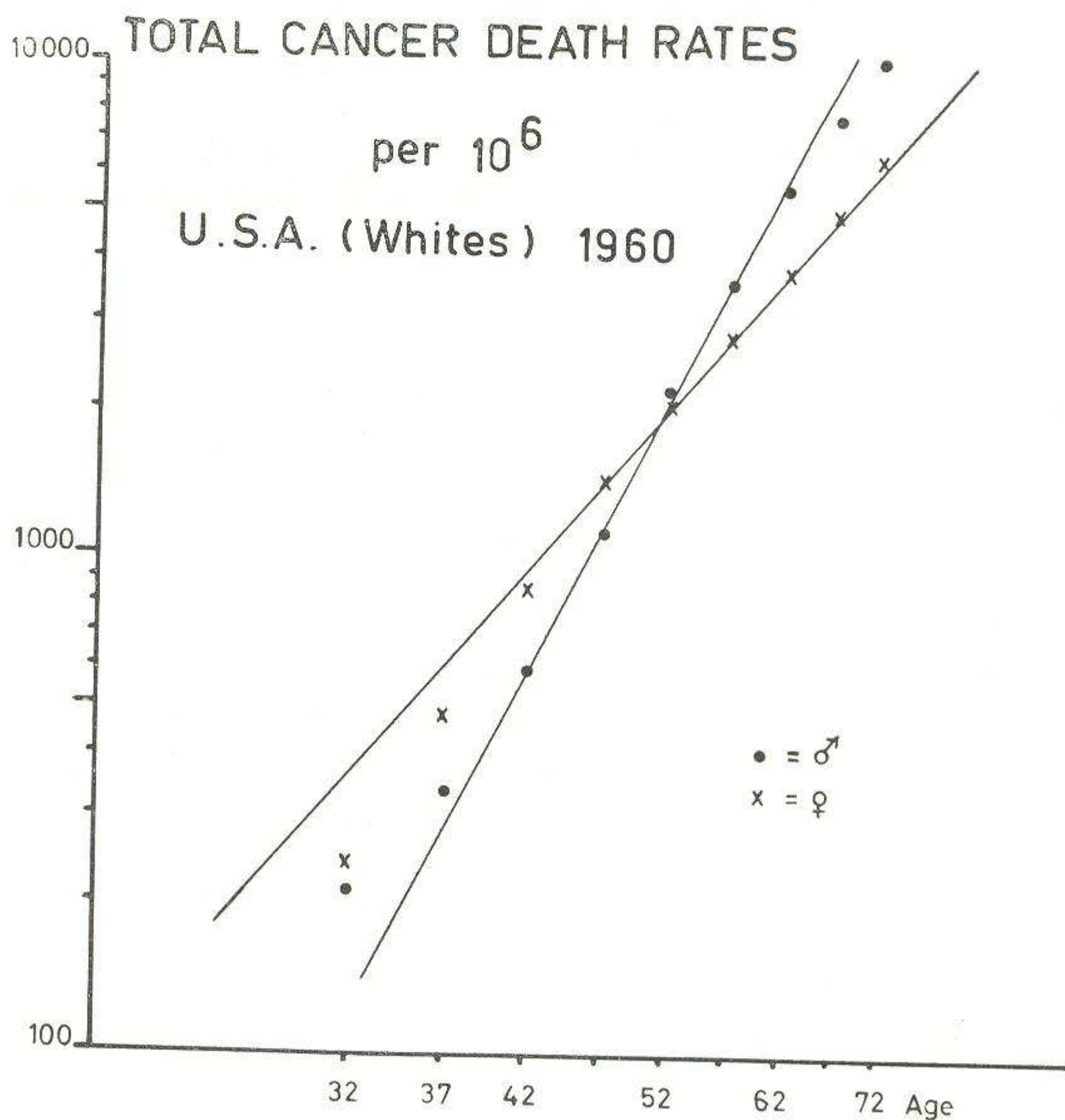


Fig. 1.

middeld snijpunt te zoeken van een bundel regressielijnen met de verzameling van de lijn parameters van deze lijnen (in casu de $\log R_1$'s en de b 's) is dan equivalent met het zoeken van één regressielijn tussen de verzameling van de punten bepaald door de $\log R_1$'s en de b 's, m.a.w. de correlatie tussen $\log R_1$ en b .

* In publicatie.

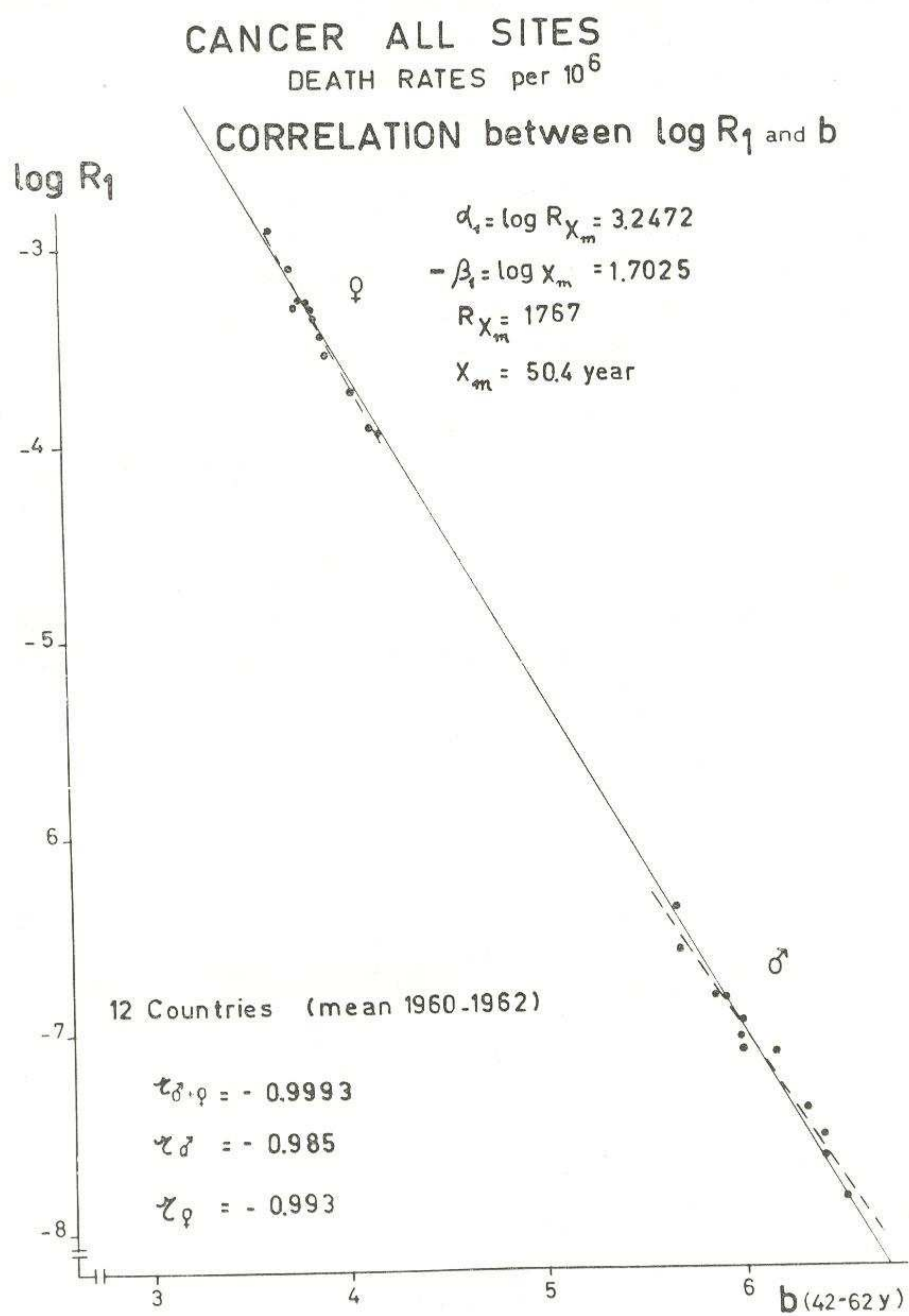


Fig. 2.

De berekening bevestigt dat de correlatie tussen $\log R_1$ en b en tussen $\log R_x$ en b (welke laatste mathematisch niet gecorreleerd zijn) identische waarden oplevert voor X_m , rekening houdend met de vergelijking (2).

Een tweede methode om aan te tonen dat het bestaan van X_m een biologisch fenomeen is en geen technisch artefact kan gezocht worden in de berekening van het snijpunt * van slechts twee sterfte-leeftijd lijnen van éénzelfde land en jaartal, maar van beide geslachten. Fig. 1 geeft de kankersterfte weer voor beide geslachten in de Verenigde Staten in 1960. De lijnen snijden elkaar in de omgeving van 50 jaar. Welnu de berekening bewijst dat voor n landen en k

$$\text{jaartallen } \log X_m = \frac{\sum_{i=1}^n \sum_{j=1}^k \log X_{ij}}{n \cdot k}.$$

Fig. 2 geeft de correlatie weer tussen $\log R_1$ en b voor mannen en vrouwen. Eerst voor beide geslachten afzonderlijk dan, aangezien $b \delta$ niet significant verschillend is van $b \varphi$, ook voor beide samen. De berekende isomortaliteitsleeftijd, X_m , is hier 50,4 jaar. Ook voor andere aandoeningen kan een isomortaliteitsleeftijd berekend worden; zo is deze voor infectieziekten rond de 30 jaar, voor diabetes rond de 50 jaar en voor atherosclerosis (degeneratieve hartaandoeningen) rond de 85 jaar.

Door X_1 of X_m in vergelijking (1) te brengen bekomt men :

$$R_{X_1} = R_1 X_1^b \quad (3) a$$

$$\text{of } R_{X_m} = R_{X_m} X_m^b \quad (3) b$$

Uit (1) en (3) volgt

$$R_x = R_{X_1} \left(\frac{x}{X_1} \right)^b \quad (4) a$$

$$\text{of } R_x = R_{X_m} \left(\frac{x}{X_m} \right)^b \quad (4) b$$

Vergelijking (4) is biologisch veel interessanter dan (1), aangezien R_1 een fictieve waarde is, terwijl R_{X_1} of R_{X_m} meestal rechtstreeks observeerbaar is.

Welke van de twee vergelijkingen (4) zal verkozen worden hangt af van het feit of X_1 in alle landen en jaartallen een constante is. In dat geval is (4) b te verkiezen. Door verder onderzoek is het echter gebleken dat X_1 niet constant is, noch in de tijd, noch onder verschillende landen. Daarom zal (4) a de beste keuze zijn.

* De leeftijd overeenkomend met dit snijpunt wordt dan X_1 genoemd.

(4) a is ook van nut voor de studie van de sex ratio van de sterfte (verhouding mannelijke op vrouwelijke sterfte). Vermits R_{X1} , per definitie, gelijk is voor beide geslachten, herleidt de sex ratio zich tot

$$\left(\frac{x}{\bar{X}_1} \right) b_{\delta} - b_{\varphi} \quad (5)$$

ten minste voor het gebied van x waar aan (1) voldaan wordt.

Door Joossens en Van den Berghe en door Joossens en Meulepas werd recent een hypothese voorgesteld waarbij b , de exponent van de leeftijd, als genetisch van oorsprong beschouwd wordt. b , equivalent met de elasticiteit van de economisten, zou volgens deze hypothese gedragen worden door de X chromosomen en namelijk dominant zijn voor degeneratieve cardiovasculaire aandoeningen en diabetes, maar recessief voor kanker en infectieziekten. Nu is het gemakkelijk te bewijzen dat b_{δ} in dit geval de waarde van b_{φ} zal zal bepalen en namelijk zo dat $b_{\delta} - b_{\varphi}$ des te groter zal zijn in absolute waarde naarmate de variance van b_{δ} groter is. Voor een rechthoekige verdeling en een dominant verschijnsel zoals in atherosclerose (tussen 50 en 84 jaar) is

$$b_{\delta} - b_{\varphi} = -\frac{\sigma_{b_{\delta}}}{\sqrt{3}} \quad (6)$$

voor een recessief, zoals kanker tussen 40 en 64 jaar,

$$b_{\delta} - b_{\varphi} = +\frac{\sigma_{b_{\delta}}}{\sqrt{3}}$$

Voor een normale verdeling is $b_{\delta} - b_{\varphi}$ iets kleiner voor éénzelfde variance, dan voor een rechthoekige verdeling.

Deze formule werd afgeleid uit een meer algemeen werk van Moriguti (1951).

Hieruit volgt, als de variance voor kanker gelijk is aan die van atherosclerose, dat $(b_{\delta} - b_{\varphi})_{ca} = -(b_{\delta} - b_{\varphi})_{at}$.

Onder dezelfde voorwaarde is dan

$$b_{at\delta} + b_{ca\delta} = b_{at\varphi} + b_{ca\varphi}$$

Figuur 3 toont dat aan de eis van gelijke variance voor $b_{ca\delta}$ en $b_{at\delta}$ ongeveer voldaan wordt als het gemiddelde van 12 landen genomen wordt en dit voor alle onderzochte jaren tussen 1956 en 1962.

b is een vrij grote waarde, gemiddeld ongeveer 6 voor kanker bij mannen (4 bij vrouwen) en ongeveer 7 voor atherosclerose bij mannen (9 bij vrouwen). Daaruit volgt dat b_{at} vooral de maximaal bereikbare levensduur duidelijk beperkt. b kan dan ook beschouwd worden, in de hoger gemaakte hypothese een eigenschap van de X chromosoom, als de genetisch limiterende factor van de levensduur en dit karakter zou dan voor elke diersoort verschillend zijn. Bomen, die geen X chromosomen bezitten zoals dieren, hebben theoretisch dan ook geen beperkte levensduur.

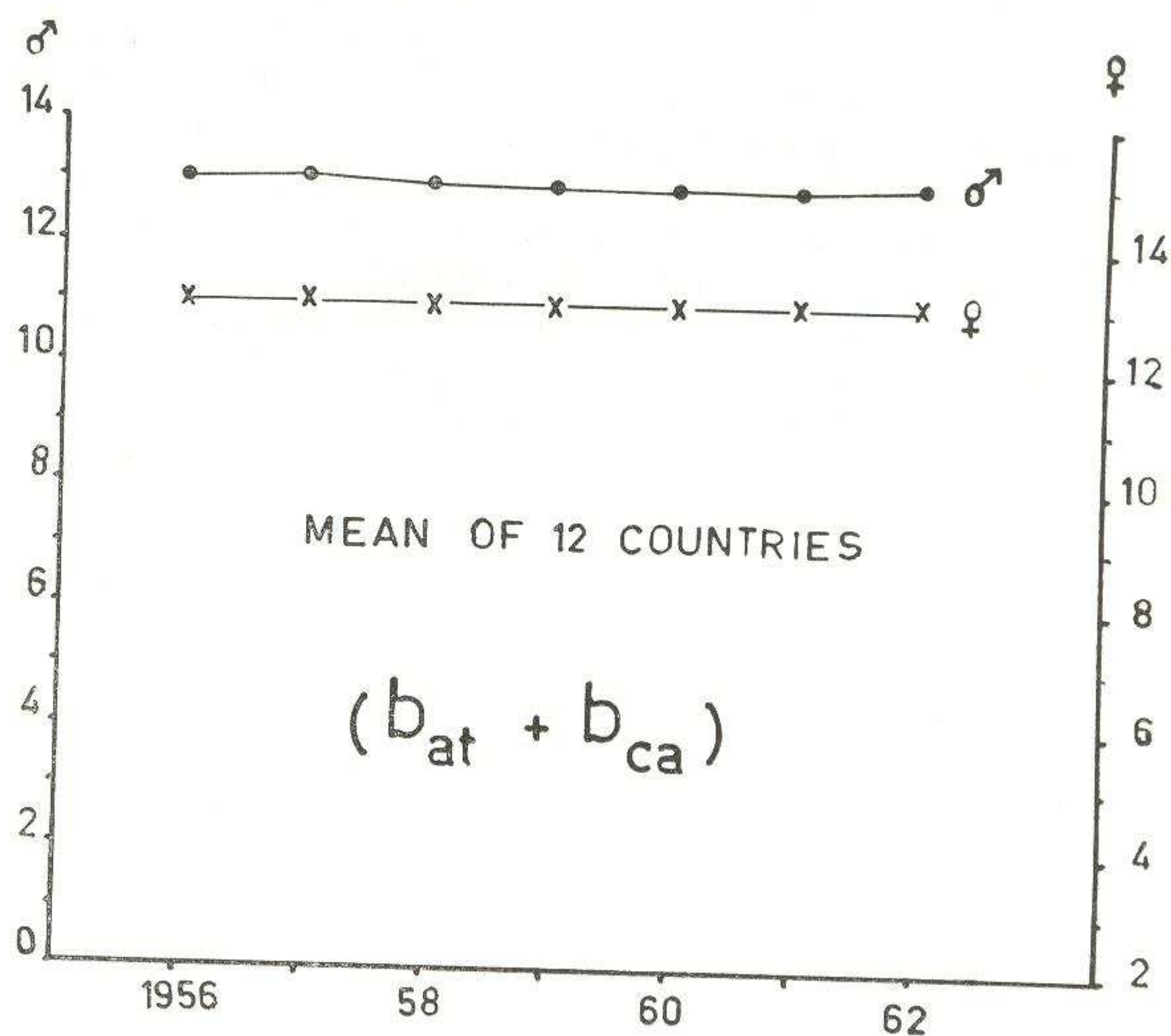


Fig. 3.

Een gans andere methodiek om kankersterftestatistieken, meer bepaald sommige kankerlocalisaties, te verwerken is de volgende (Joossens 1965).

Bepaalde localisaties kunnen per leeftijd, tussen 40 en 74 jaar, uitgedrukt worden in % van de totale kankersterfte op die leeftijd. Voor elk van de 21 onderzochte landen, blijkt nu dat de som van de % van maag-, long- en geslachtskankers ongeveer gelijk is op elke leeftijd van 40 tot 74 jaar, op voorwaarde dat de gegevens van mannen en vrouwen samengeteld worden. Fig. 4

geeft hiervan enkele voorbeelden. Voor mannen stijgt het % met de leeftijd, voor vrouwen daalt het juist evenveel. Dit wijst erop dat kankersterfte biologische informatie uitwisselt tussen beide geslachten. Dit is dan een ander argument voor de hogergemaakte hypothese.

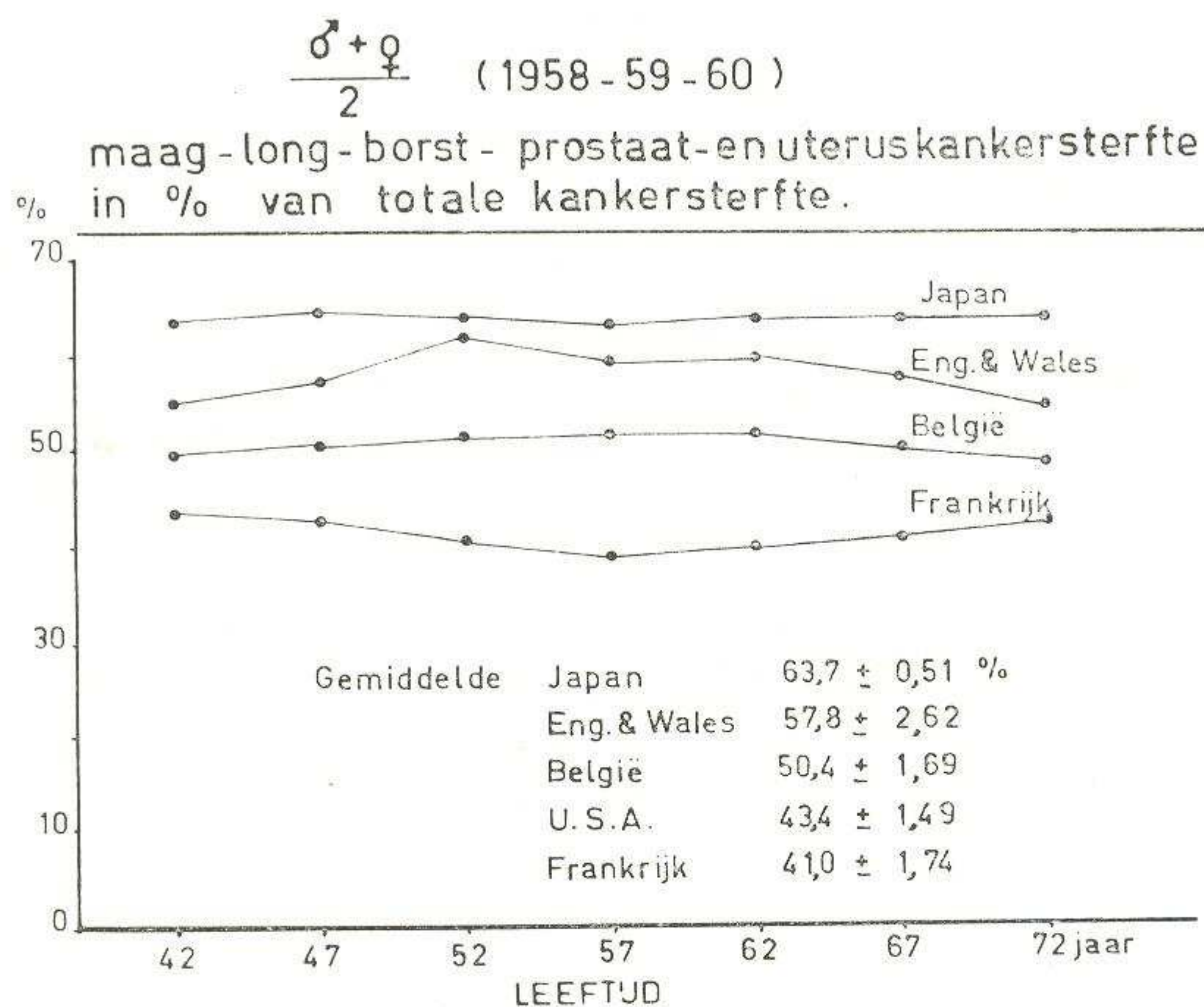


Fig. 4.

Het belang van maag-, long- en geslachtskankersterfte ($\pm 50\%$ van de totale kankersterfte) wordt door de volgende regressie-analyse aangetoond. Gegevens van 20 of 21 landen van de jaren '58, '59 en '60 en waargenomen op leeftijden gaande van 42 tot 72 jaar ($n = 427$) werden in een veelvoudige lineaire regressie verwerkt. De regressie werd eerst per leeftijd uitgevoerd ($n = 61$) om de vergelijkbaarheid van de bekomen regressiecoëfficiënten na te gaan van leeftijd tot leeftijd. Dit bleek het geval te zijn voor de coëfficiënten van maag-, long- en geslachtskankersterfte, maar niet voor de constante factor. Deze werd namelijk groter met de leeftijd volgens een machtsfunctie van de leeftijd. Daarom werd in de regressieanalyse een machtsfunctie van de leeftijd ingevoerd. De waarde van deze exponent werd door de elektronische ordinator door « error and trial » gezocht totdat de R^2 maximaal was en zo mogelijk tevens de nieuwe constante factor niet significant verschillend was van nul.

De waarde van de regressiecoëfficiënten wordt in tabel I aangegeven.

TABEL I

Beste regressie-vergelijkingen

 $n = 427$

$$\delta \quad R_x = 0,811 \left(\frac{x}{48,9} \right)^{4,90} + 0,796 M + 1,155 L + 1,380 G (*) \quad R^2 = 0,991$$

$$\varphi \quad R_x = 0,819 \left(\frac{x}{48,9} \right)^{4,90} + 0,873 M + 1,607 L + 1,696 G - 0,258 \quad R^2 = 0,986$$

$$\frac{\delta + \varphi}{2} \quad R_x = 0,414 \left(\frac{x}{48,9} \right)^{5,40} + 0,872 M + 1,260 L + 1,993 G \quad R^2 = 0,993$$

In tabel II wordt één regressievergelijking gebruikt om de kankersterfte per leeftijd te berekenen in drie verschillende rassen, die niet in de oorspronkelijke regressieberekening bevat waren. De overeenkomst met de waargenomen waarden is bevredigend.

TABEL II

Totale kankersterfte ‰ berekend per leeftijd uit de maag-, long- en genitaalkankersterfte (met één vergelijking, de laatste van tabel I)

$$\frac{\delta + \varphi}{2} \quad 1959$$

Leeftijd	Verenigde Staten (niet-blanke bevolking)		Israël		Nieuw Zeeland	
	Waarge- nomen	Berekend	Waarge- nomen	Berekend	Waarge- nomen	Berekend
42	1,17	1,15	0,71	0,73	0,64	0,73
47	1,78	1,75	1,13	1,05	1,42	1,22
52	2,86	2,64	1,99	1,63	1,79	1,87
57	4,47	4,32	2,90	2,72	2,71	2,83
62	5,79	5,67	3,82	3,96	4,62	4,68
67	8,60	8,78	6,57	5,89	6,03	6,63
72	9,44	10,91	9,19	8,27	8,52	8,98

(*) R_x = leeftijd-specifieke sterfte ‰ voor totale kanker op leeftijd x , M = maagkankersterfte ‰, L = longkankersterfte ‰, G = borst- en prostaatkankersterfte bij de man, borst- en uterus kankersterfte bij de vrouw ‰, alle waargenomen op leeftijd x .
 R^2 is het kwadraat van de totale veelvoudige correlatie-coëfficiënt tussen de verschillende veranderlijken.

LITERATUUR

JOOSSENS, J.V. : Mortaliteit en leeftijd. Verh. Kon. Vlaamse Akad. voor Gen. van Belg. 1963 : 25, 175.

JOOSSENS J.V. : Het probleem van de kankersterfte. Verh. Kon. Vlaamse Akad. voor Gen. van Belg. 1965 : 27, 489.

JOOSSENS, J.V. en H. VAN DEN BERGHE : Exogenous and endogenous factors in atherosclerosis. Acta Cardiologica (in druk).

JOOSSENS, J.V. en E. MEULEPAS : The sex ratio of mortality. A genetic hypothesis. Kon. Vlaamse Akad. voor Gen. van Belg. (in druk).

MORIGUTI, S. : Extremal properties of extreme value distributions. Ann. Math. Stat. 1951 : 22, 523.

De hier aangehaalde berekeningen werden door het « Rekencentrum » van de Leuvense Universiteit uitgevoerd (Prof. H. Florin en J. Mainguet). Dr G. Peeters en Dr A. Vuylsteke voerden de programmaties uit.