
Vorsicht Falle

Dr. Dirk Hasenclever
Institut für Medizinische Informatik, Statistik &
Epidemiologie (IMISE)
Zentrum für Klinische Studien Leipzig (ZKS-KKS)

dirk.hasenclever@imise.uni-leipzig.de

Hans-Peter Beck-Bornholdt
Hans-Hermann Dubben

Der Hund, der Eier legt



Rowohlt

Erkennen von Fehlinformation
durch Querdenken

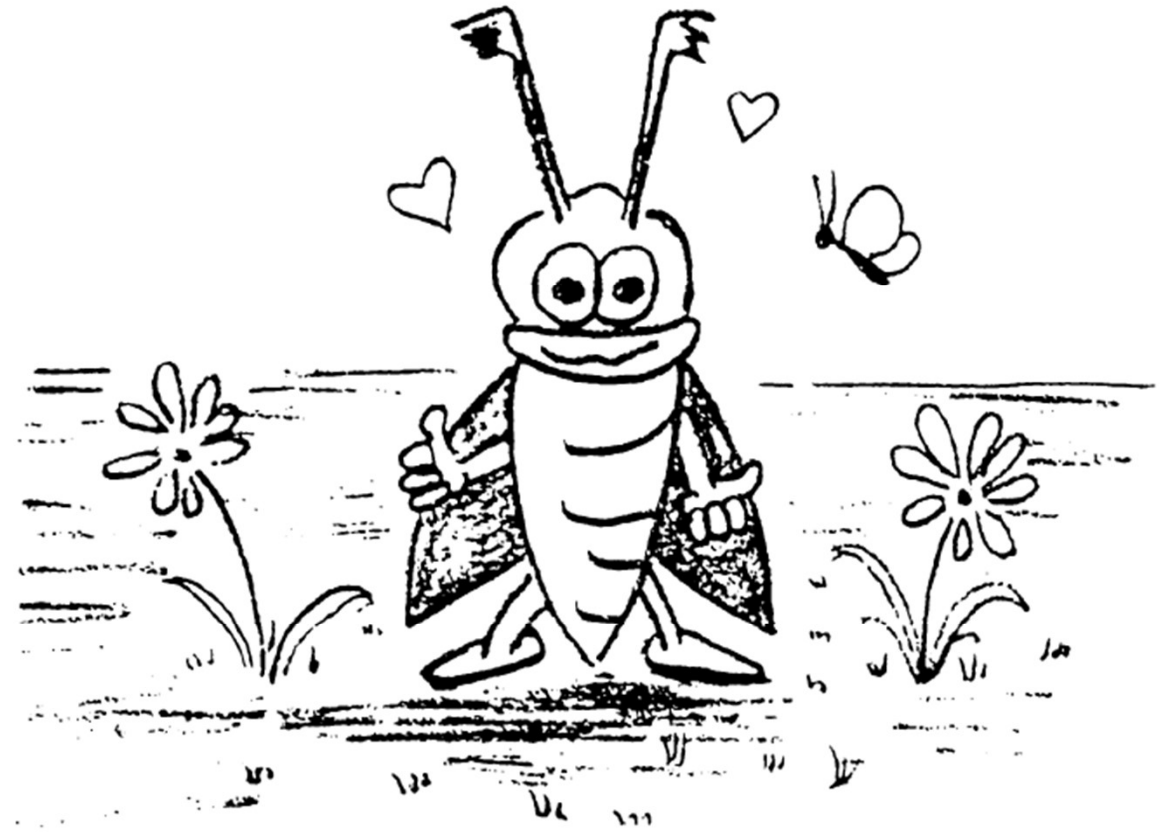
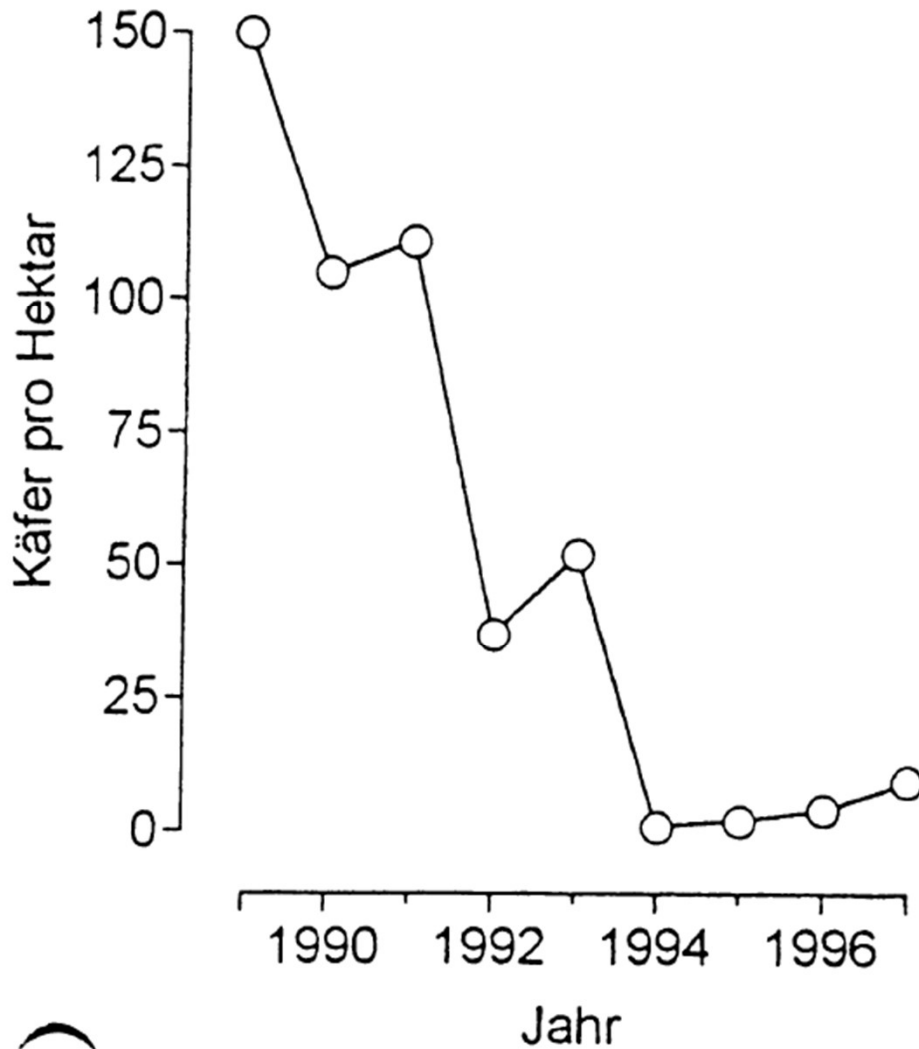


Abbildung 13: Bestand einer inzwischen seltenen, vom Aussterben bedrohten und völlig frei erfundenen Käferart im Harz

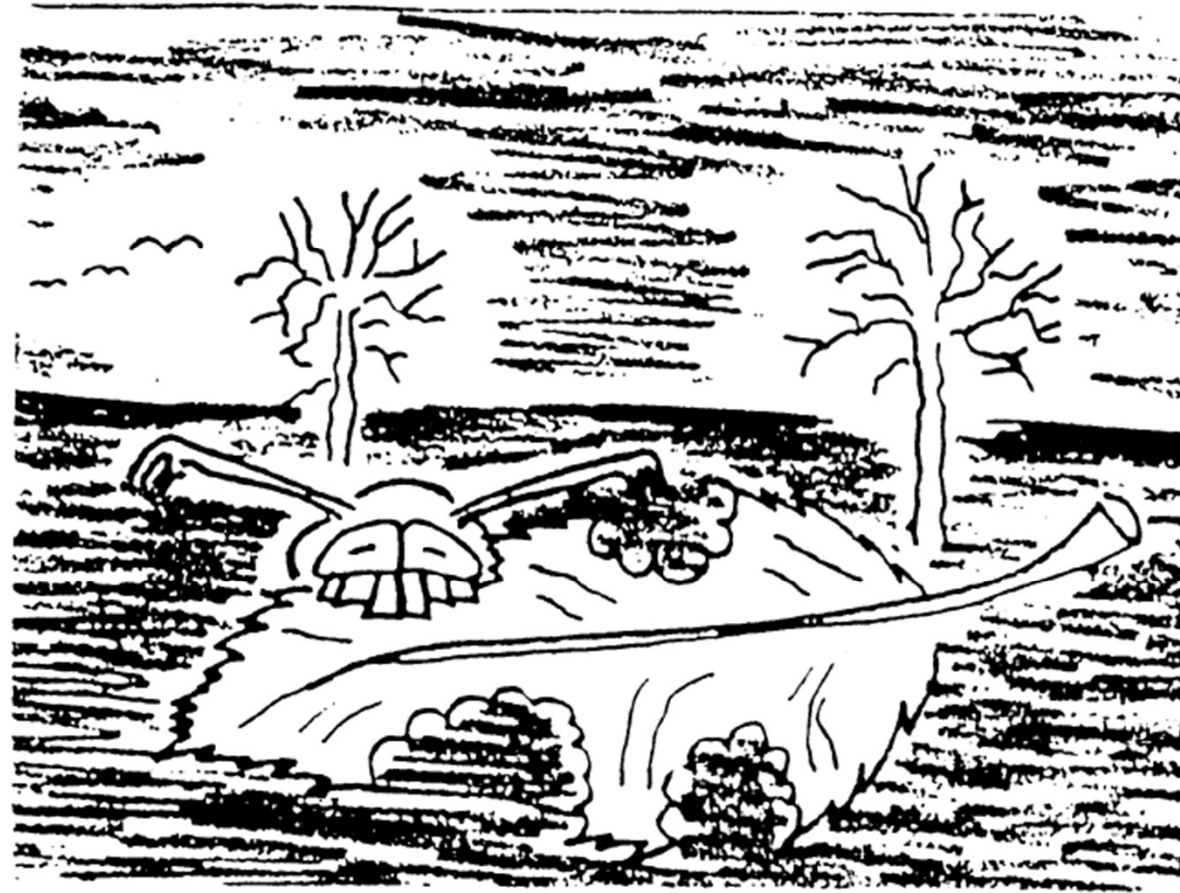
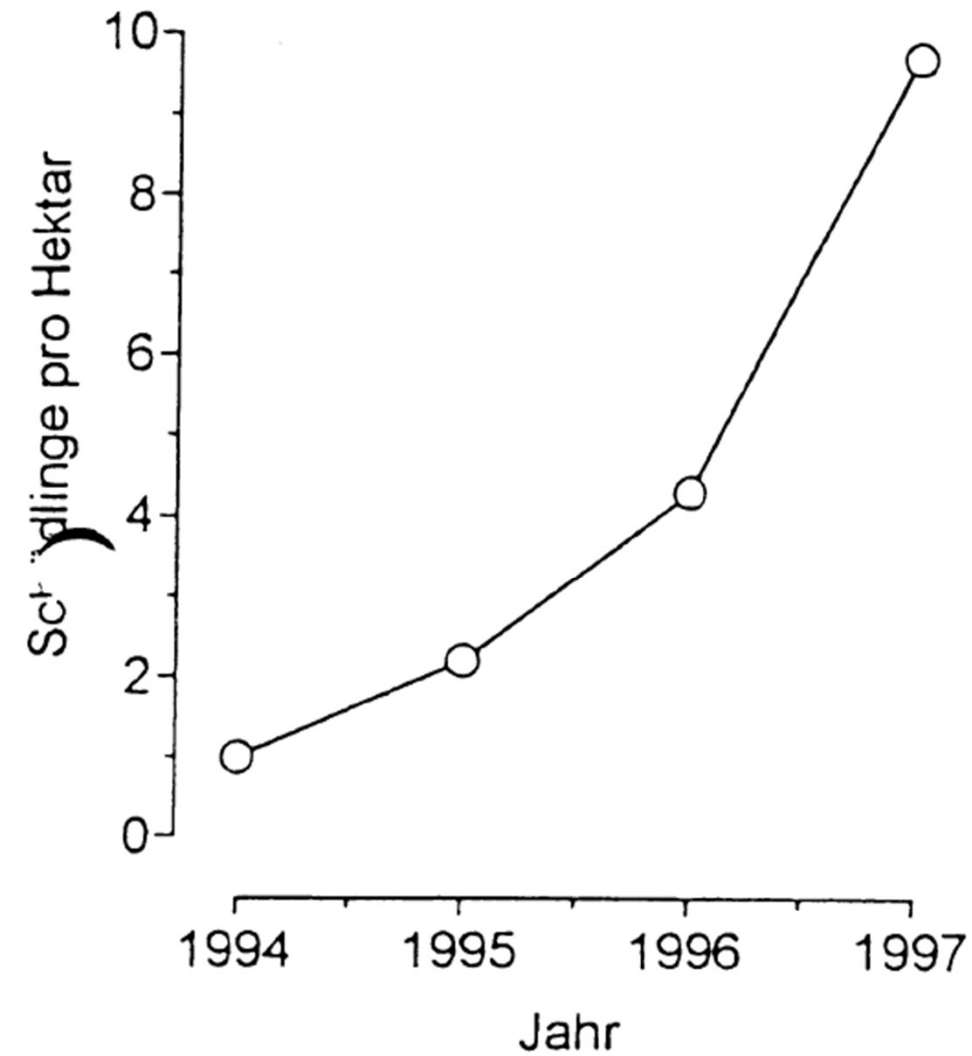


Abbildung 14: Rasante Entwicklung von Schädlingen im Harz

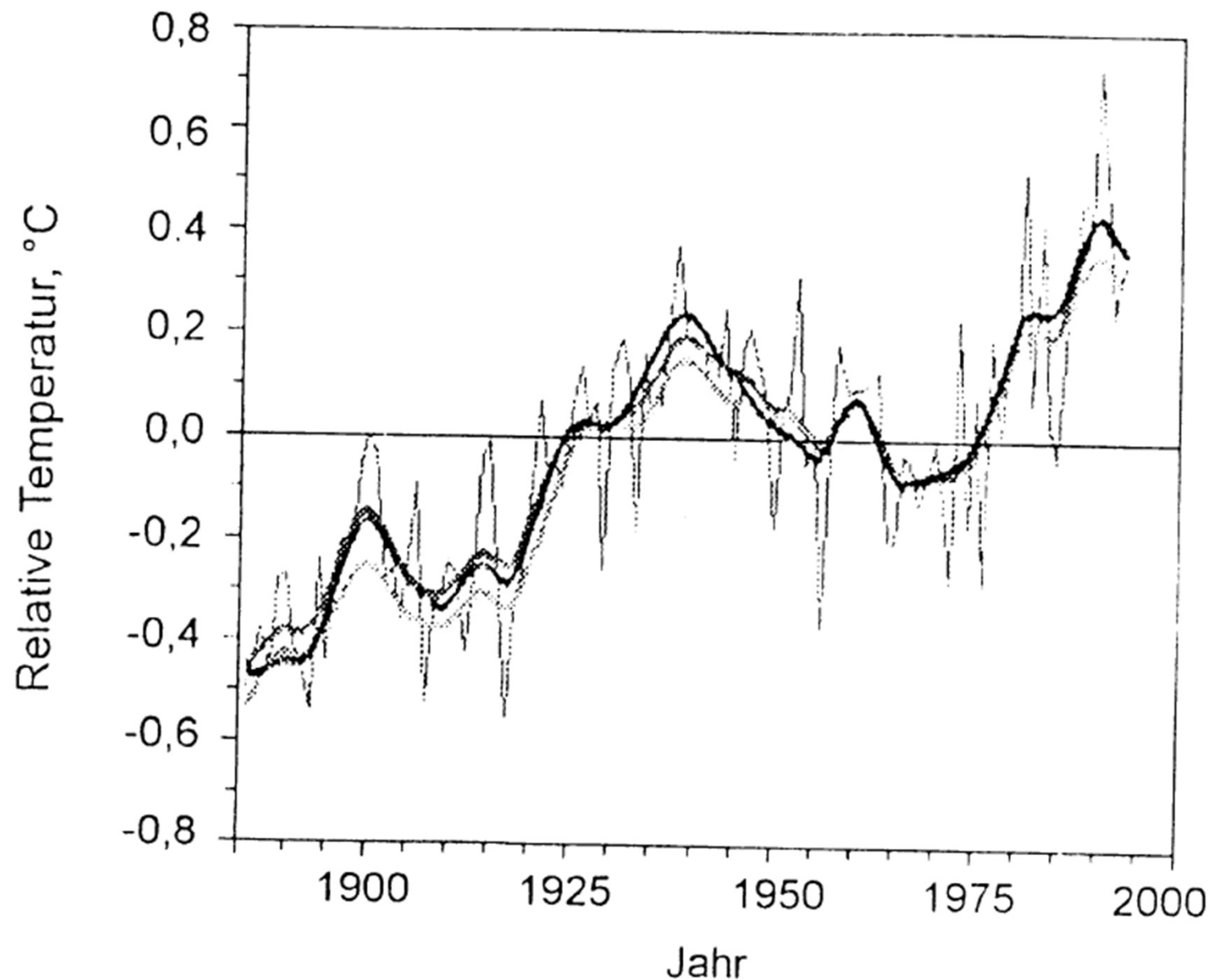


Abbildung 15: Globale Erwärmung seit 1885 (Karl & Baker 1996)

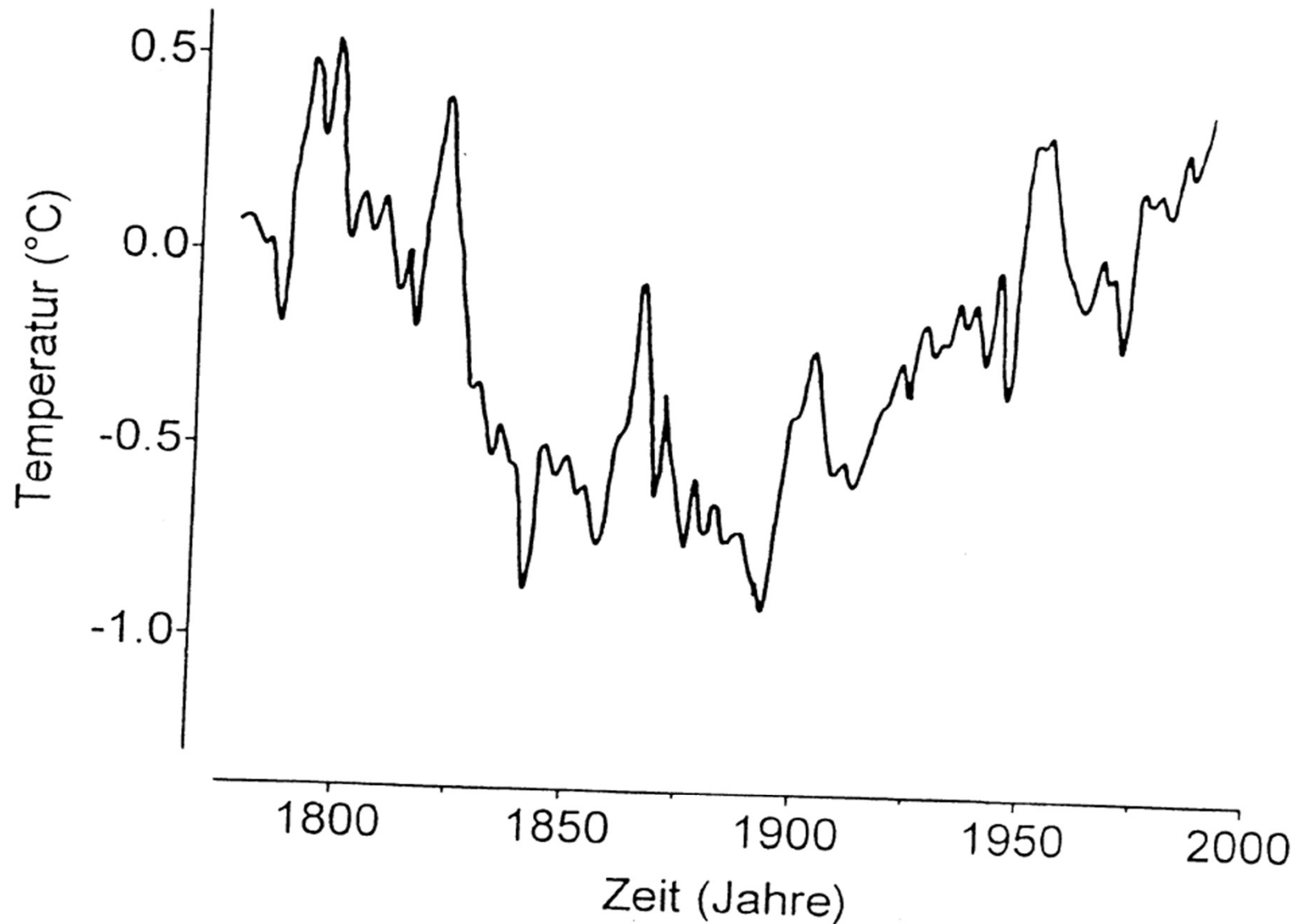


Abbildung 16: Mittlere Lufttemperatur in Österreich von 1778 bis 1991 (nach Sommaruga-Wögrath et al. 1997)

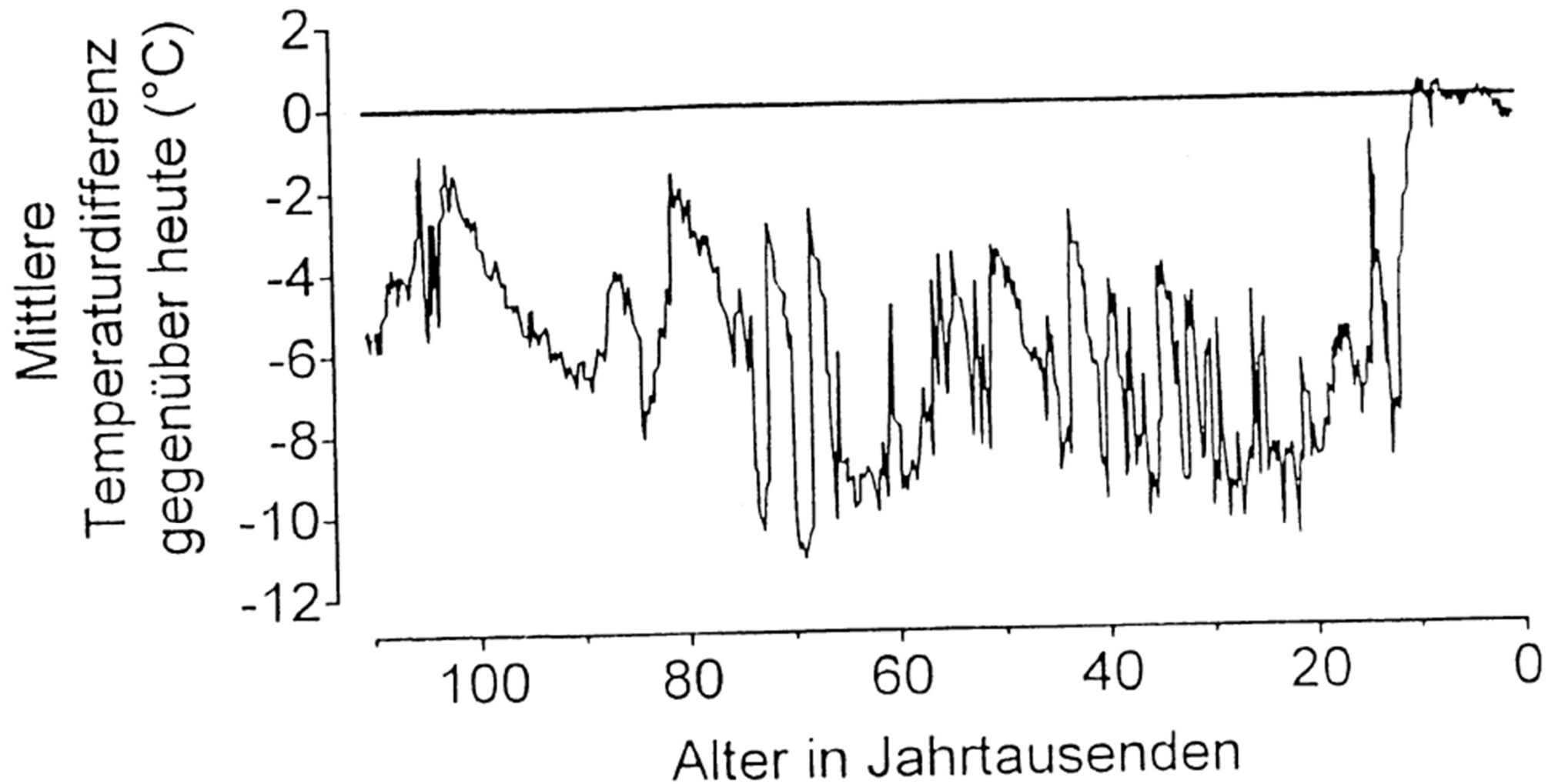
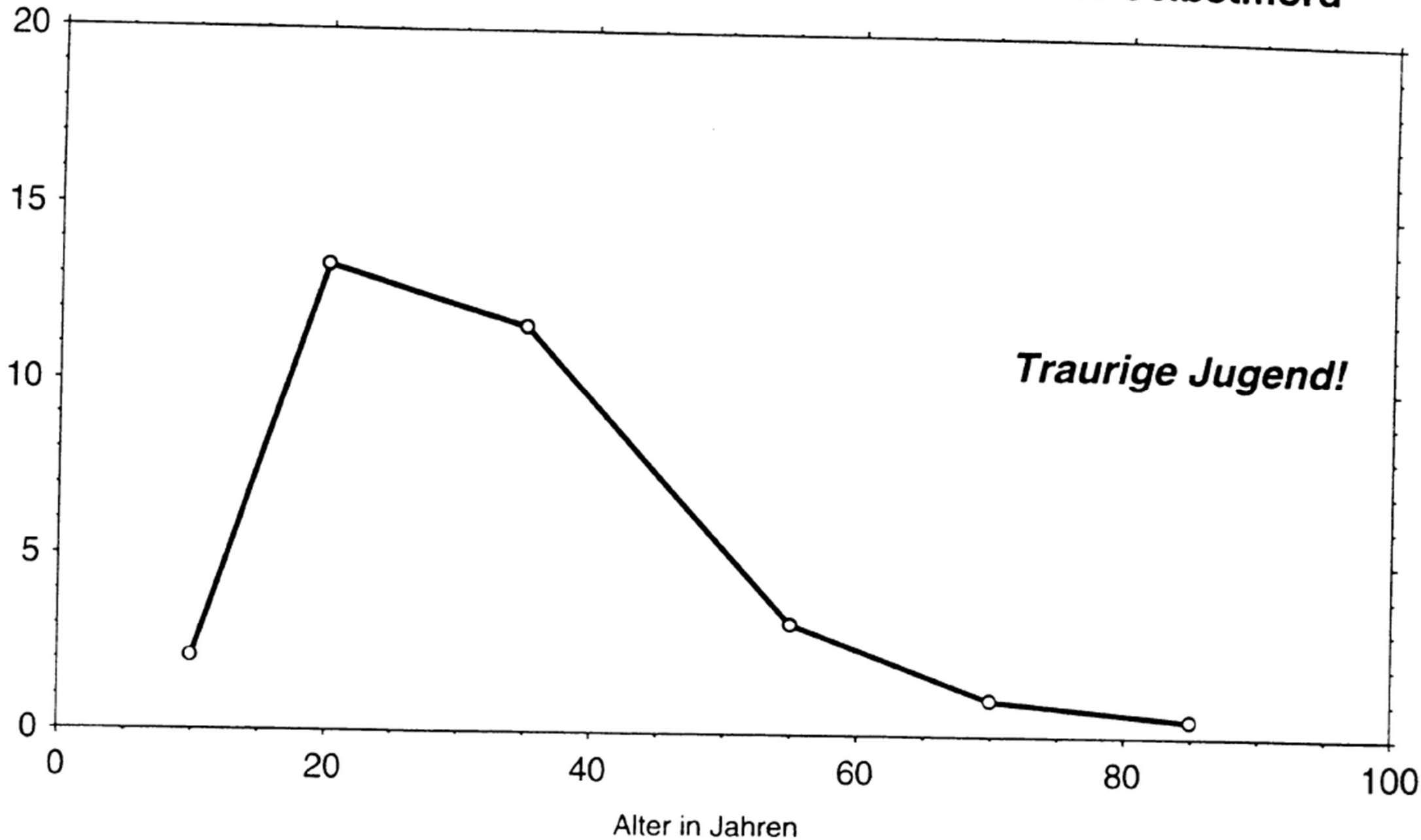
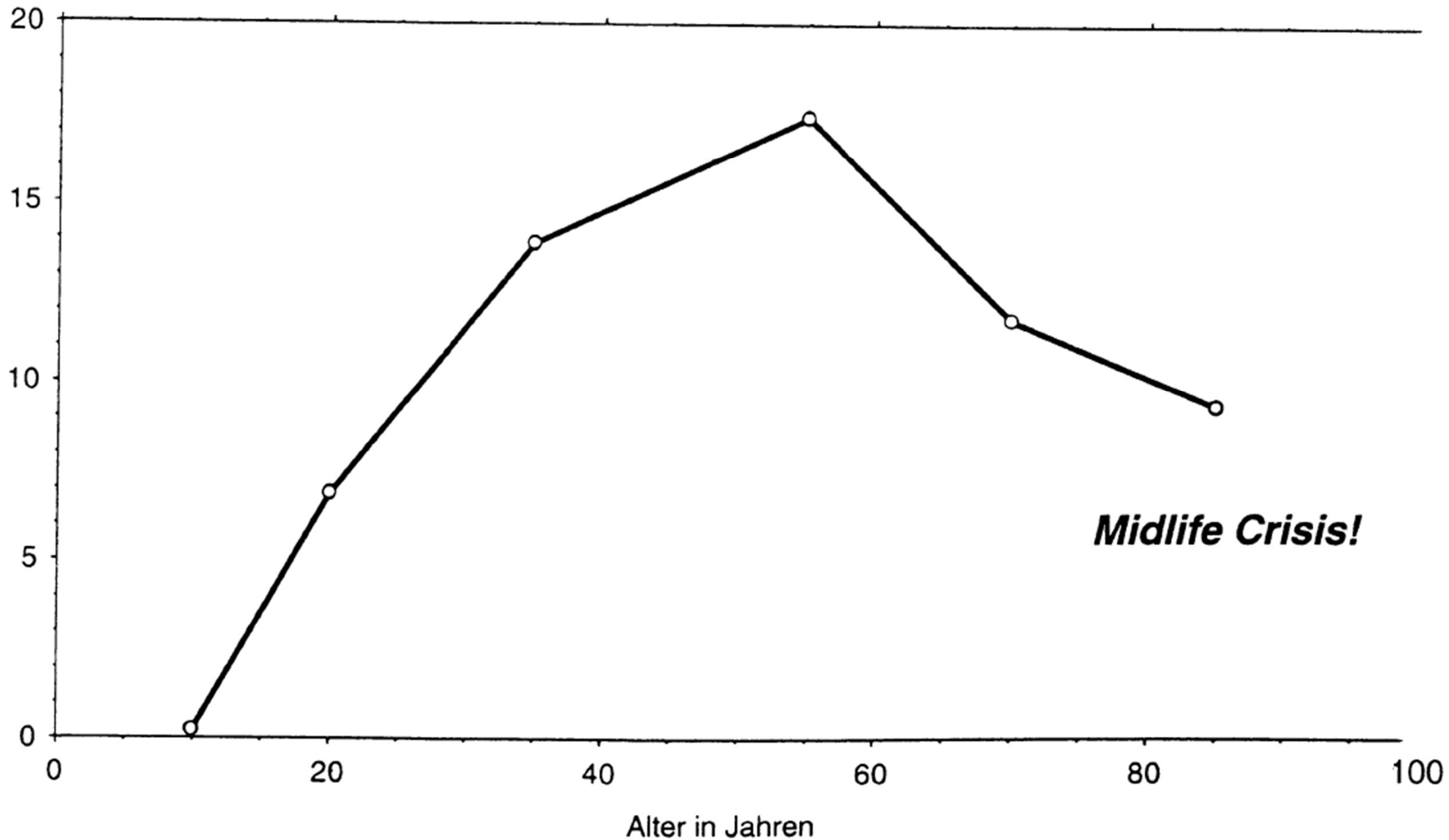


Abbildung 17: Globale Temperatur während der letzten 110 000 Jahre, ermittelt durch Messungen an Bohrkernen im grönländischen Eisschild (Broecker 1996)

Von 100 Todesfällen einer Altersklasse entfielen auf Selbstmord

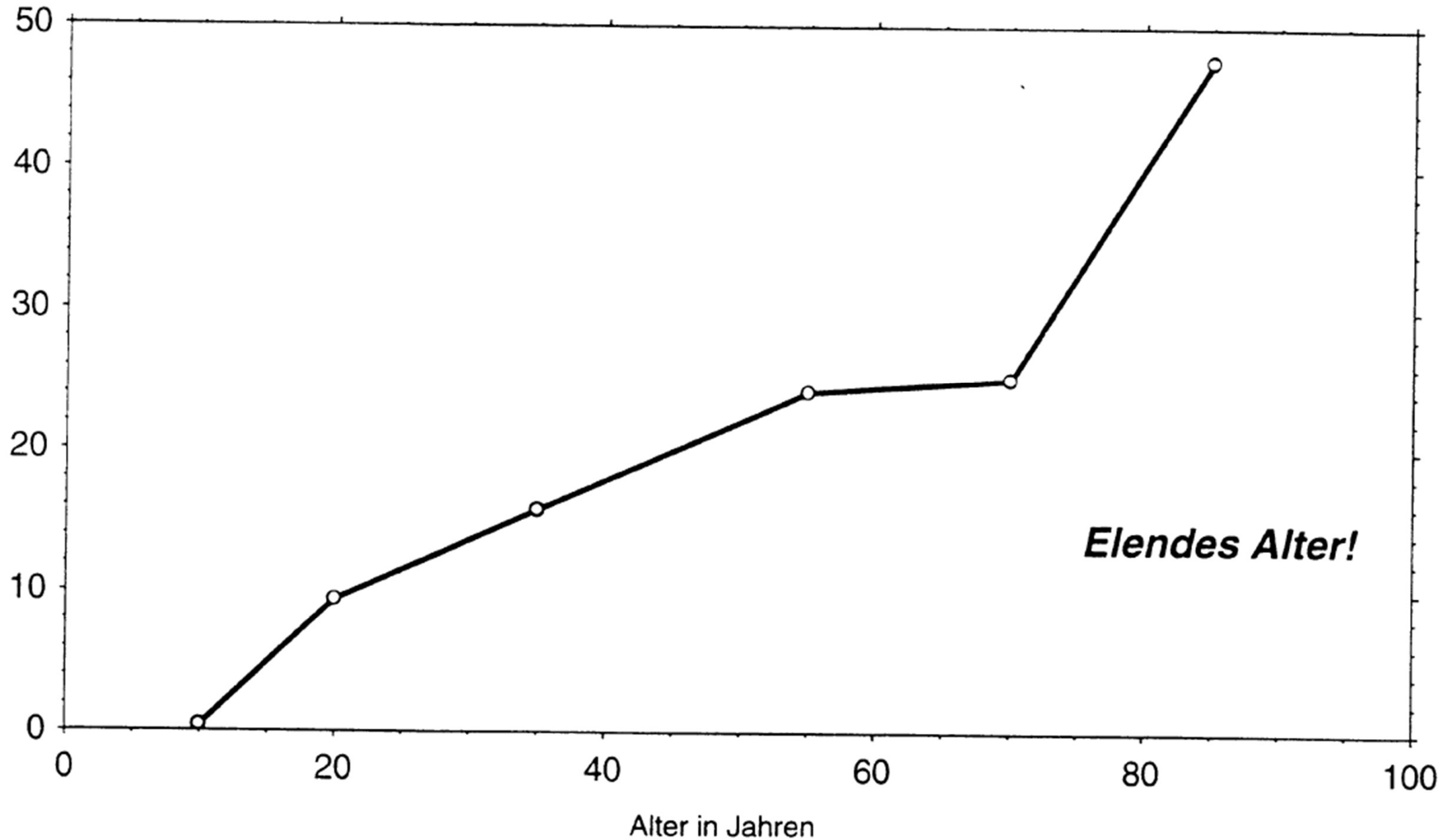


Von 100 Selbstmördern entfielen auf die betroffene Altersklasse

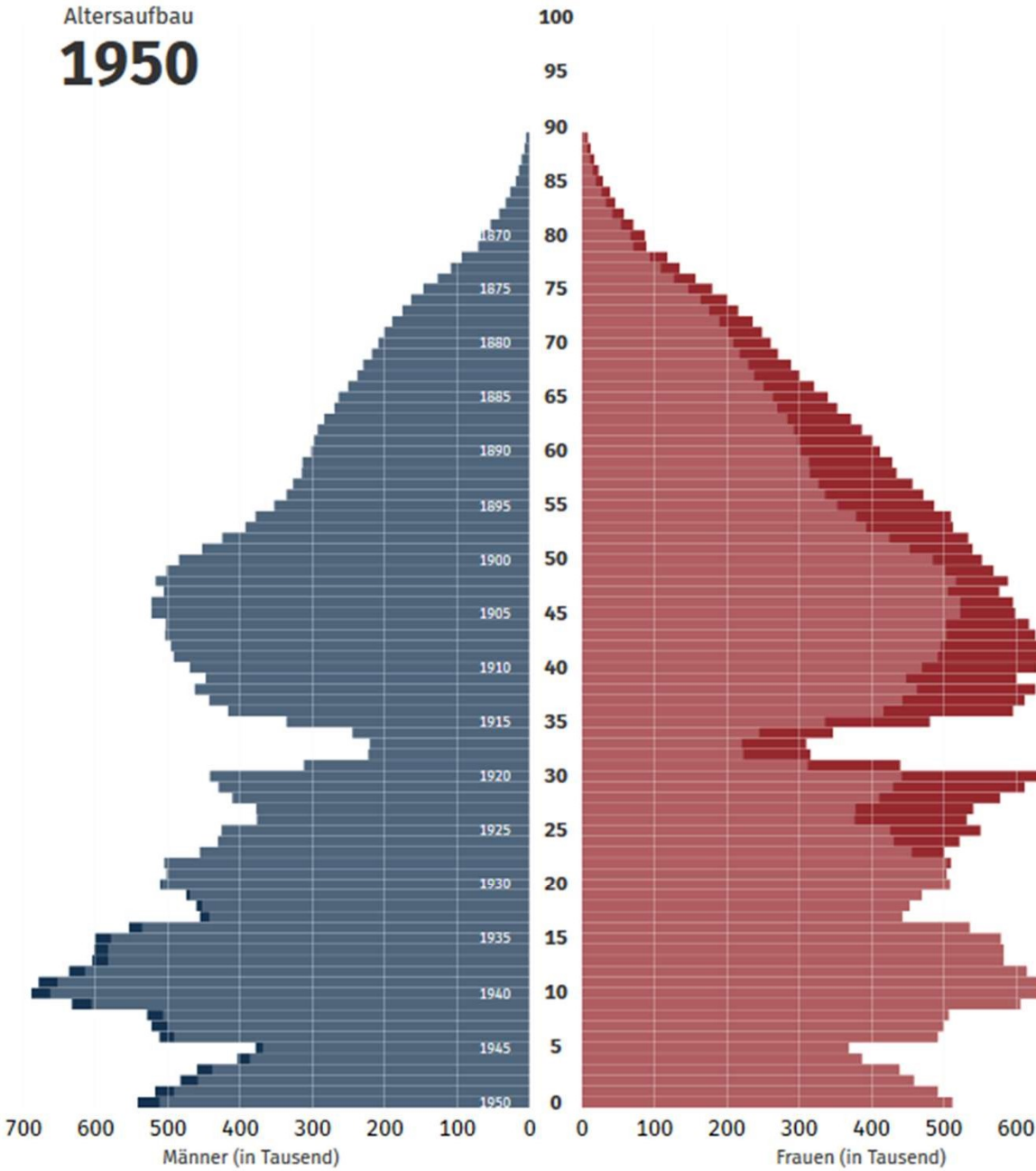


Midlife Crisis!

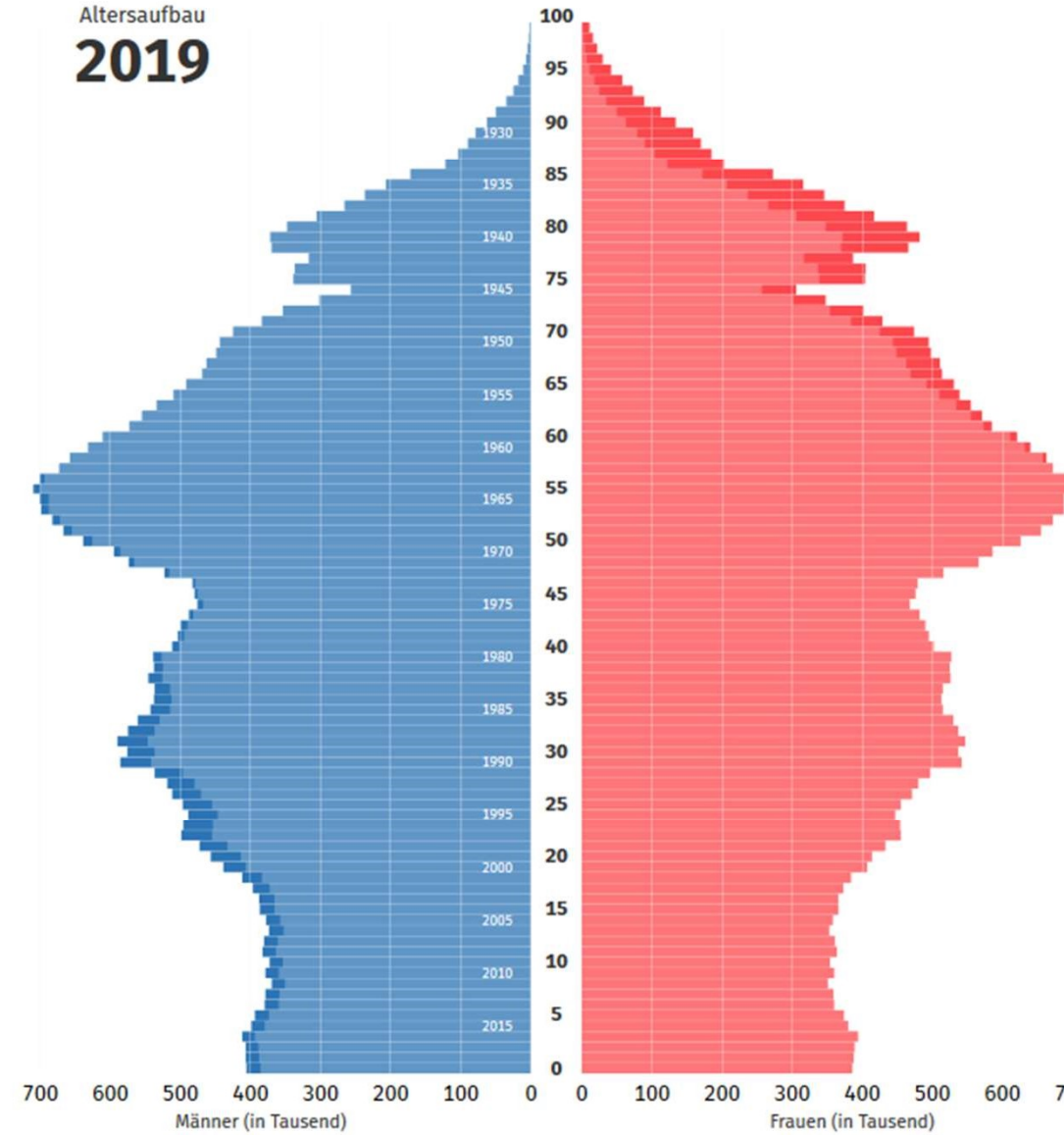
Selbstmorde pro 100 000 Lebende gleicher Altersklasse



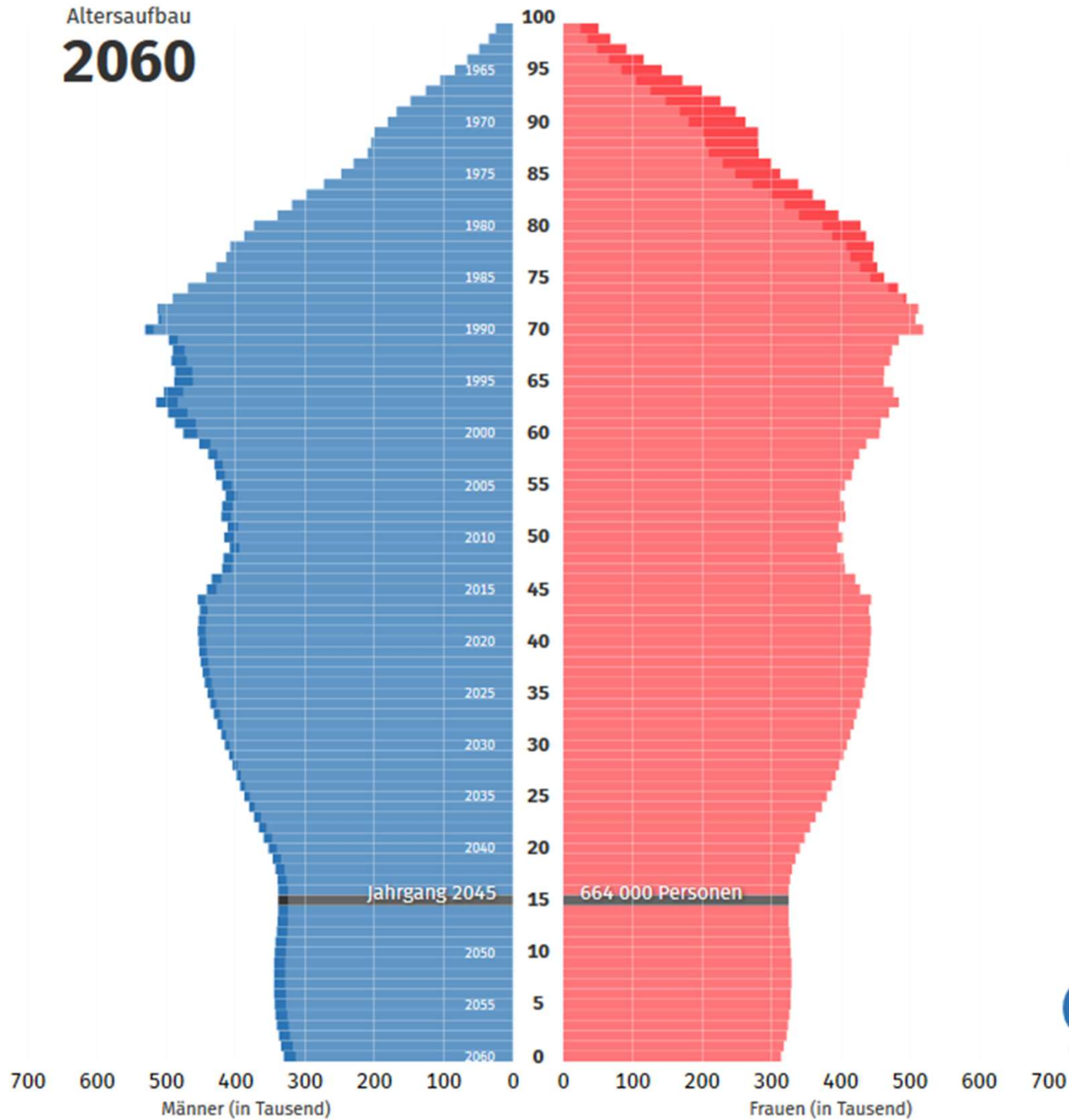
Altersaufbau 1950



Altersaufbau 2019



Altersaufbau 2060



?

In Abbildung 1.1 werden die Selbstmorde in Relation gesetzt zur Gesamtheit der Todesursachen in einer Altersgruppe. Es ist hier also der relative Anteil von Selbstmorden im altersspezifischen Todesursachenspektrum dargestellt. Jugendliche und jungen Erwachsene sterben vergleichsweise selten an einer Krankheit. Deshalb nimmt der Selbstmord im Todesursachenspektrum dieser Alterklassen einen relativ breiten Raum ein. Da die zum Tode führenden Krankheiten aber mit dem Alter exponentiell zunehmen, nimmt die relative Bedeutung des Selbstmords als Todesursache deutlich ab. Die Zunahme der Sterblichkeit mit dem Alter behindert also bei dieser Darstellung den Blick auf das altersspezifische Selbstmordrisiko. Das abweichende Bild in Abbildung 1.2 kommt dadurch zustande, daß man als Untersuchungspopulation nicht die Verstorbenen einer Altergruppe, sondern die insgesamt aufgetretenen Selbstmorde wählt: dargestellt ist also die Altersverteilung der Selbstmörder. Diese Verteilung spiegelt die Altersverteilung der Bevölkerung wieder, in der die mittleren und höheren Altersklassen stärker vertreten sind als die jüngeren. Es ist aber nicht verwunderlich und daher auch nicht sehr informativ, daß man in einer großen Bevölkerung mehr Selbstmorde findet als in einer kleinen. Abbildung 1.3 schließlich zeigt, wieviele von 100 000 Bundesbürgern eines bestimmten Alters Selbstmord verübten. Die Untersuchungspopulation besteht hier aus der Bevölkerung, die einer bestimmten Altersgruppe angehört. Dies ist die einzige Darstellungsform, die die altersspezifischen Selbstmordraten wiedergibt, also den relative Anteil von Bundesbürgern, die in einem bestimmten Alter Selbstmord begehen. Diese Raten sind unabhängig davon, wie stark eine Alterklasse in der Bevölkerung vertreten ist oder welche sonstigen Todesursachen es im jeweiligen Alter gibt. Abbildung 1.3 ist also die einzige der drei Varianten, die das alterabhängige Gefährdungspotential richtig darstellt.

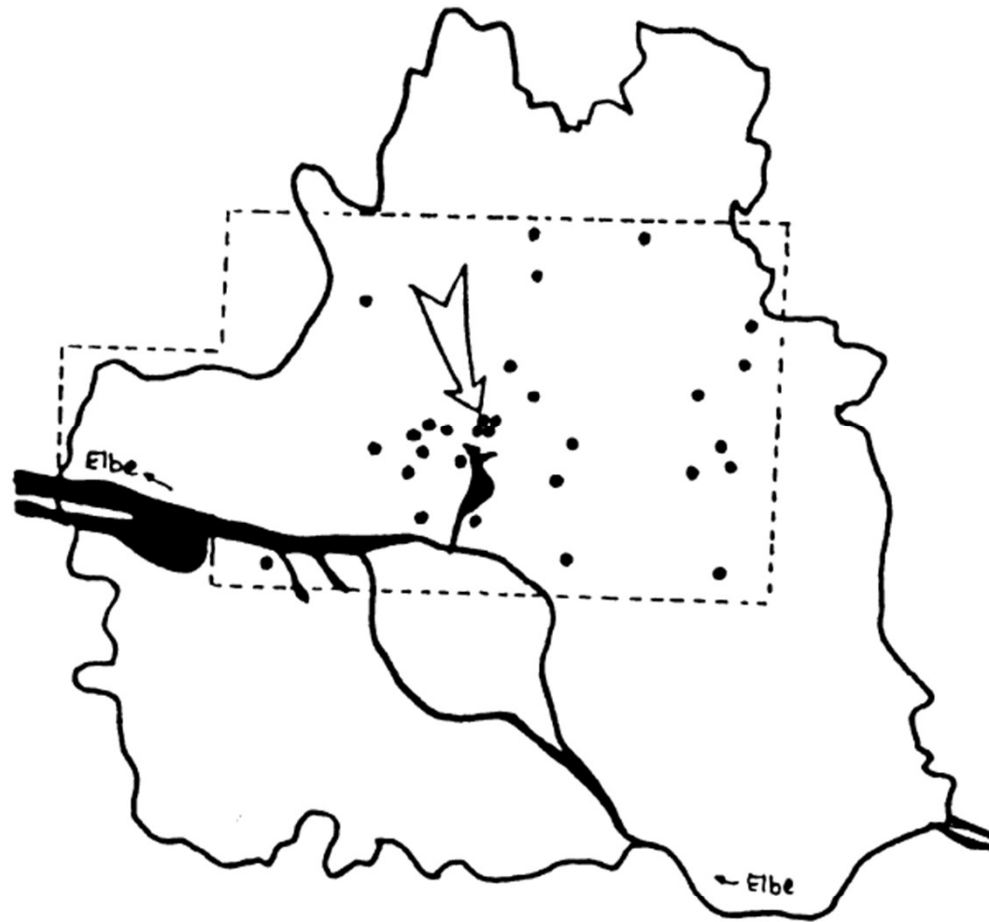


Abbildung 6: Ergebnis der Simulation eines Leukämieszenarios mit Hilfe des Hamburger Telefonbuches und Stadtplans sowie mehrerer Würfel. Die Punkte stellen die dreißig simulierten Fälle dar. Die gestrichelte Linie gibt die Grenze des verwendeten Stadtplans an. Erstellt in unserer Vorlesung im Wintersemester 1995/96.

Irene Guggenmoos-Holzmann
Klaus-Dieter Wernecke

Medizinische Statistik

Mit 71 Tabellen und 61 Abbildungen

Beispiel: Die therapeutische Illusion [82].

In einer Allgemeinpraxis wurden konsekutiv 200 Patienten, bei denen sich trotz vorliegender Beschwerden keine definitive Diagnose stellen ließ, nach dem Zufallsprinzip einer von zwei „Behandlungen“ zugeführt: ihnen wurde entweder mitgeteilt, daß keine Anhaltspunkte für eine Erkrankung bestünden und daß deshalb eine Behandlung nicht nötig sei; oder man nannte ihnen eine symptomatische Diagnose – meist die, die sie selbst vorgeschlagen hatten – und verordnete eine entsprechende Medikation.

Alle Patienten wurden gebeten, in einer Woche wieder vorzusprechen, wenn die Beschwerden nicht zurückgingen. Einen Monat später wurden alle befragt, ob die Beschwerden verschwunden waren und ob zur Besserung der Beschwerden eine weitergehende Behandlung notwendig war.

Häufigkeit einer Zweitkonsultation innerhalb eines Monats

	mit Behandlung
Patienten, die nicht wiederkamen	72
Patienten, die wegen derselben Beschwerden wiederkamen	18
Patienten, die wegen anderer Beschwerden wiederkamen	10
insgesamt	100

Angaben über eine Besserung der Beschwerden

	mit Behandlung
Patienten, deren Beschwerden sich gebessert haben	55
Patienten mit weitergehender Behandlung, nachdem sich die Beschwerden nicht gebessert haben	30
Patienten ohne weitergehende Behandlung, nachdem sich die Beschwerden nicht gebessert haben	15
insgesamt	100

Häufigkeit einer Zweitkonsultation innerhalb eines Monats

	mit Behandlung	ohne Behandlung
Patienten, die nicht wiederkamen	72	71
Patienten, die wegen derselben Beschwerden wiederkamen	18	15
Patienten, die wegen anderer Beschwerden wiederkamen	10	14
insgesamt	100	100

vergleichende Untersuchung wäre die Versuchung groß, die Besserung der Beschwerden gänzlich der Medikation zuzuschreiben. Dabei hat diese Medikation offenbar überhaupt nichts bewirkt.

Ohne eine

Angaben über eine Besserung der Beschwerden

	mit Behandlung	ohne Behandlung
Patienten, deren Beschwerden sich gebessert haben	55	61
Patienten mit weitergehender Behandlung, nachdem sich die Beschwerden nicht gebessert haben	30	29
Patienten ohne weitergehende Behandlung, nachdem sich die Beschwerden nicht gebessert haben	15	10
insgesamt	100	100

Beispiel: Wenn es heute schlecht geht, kann es morgen nur besser werden.

In einer Studie zur Behandlung der Divertikulitis untersuchten Brodribb et al. [14] die Wirkung von Weizenkleie unter anderem hinsichtlich ihres Effektes auf die Dauer der Magen-Darm-Passage. Die Passagezeit wurde bei 39 Patienten vor Behandlungsbeginn und sechs Monate nach regelmäßiger

Einnahme von Weizenkleie bestimmt. Die mittlere Passagezeit war zu beiden Zeitpunkten die gleiche. In einer weitergehenden Analyse stellten die Autoren jedoch fest, daß sich deutliche Unterschiede ergaben, wenn man die Patienten danach einteilte, wie ihre Passagezeit zu Beginn der Behandlung war. Nach Behandlungen verlängerten sich die Passagezeiten bei all den Patienten, die Anfangs kurze Zeiten gehabt hatten; bei anfänglich langen Passagezeiten fand sich eine deutliche Verkürzung; und bei mittleren Passagezeiten waren keine auffälligen Änderungen erkennbar. Die Autoren schlossen daraus, daß Weizenkleie bei Divertikulitispatienten die Dauer der Magen-Darm-Passage „normalisiert“.

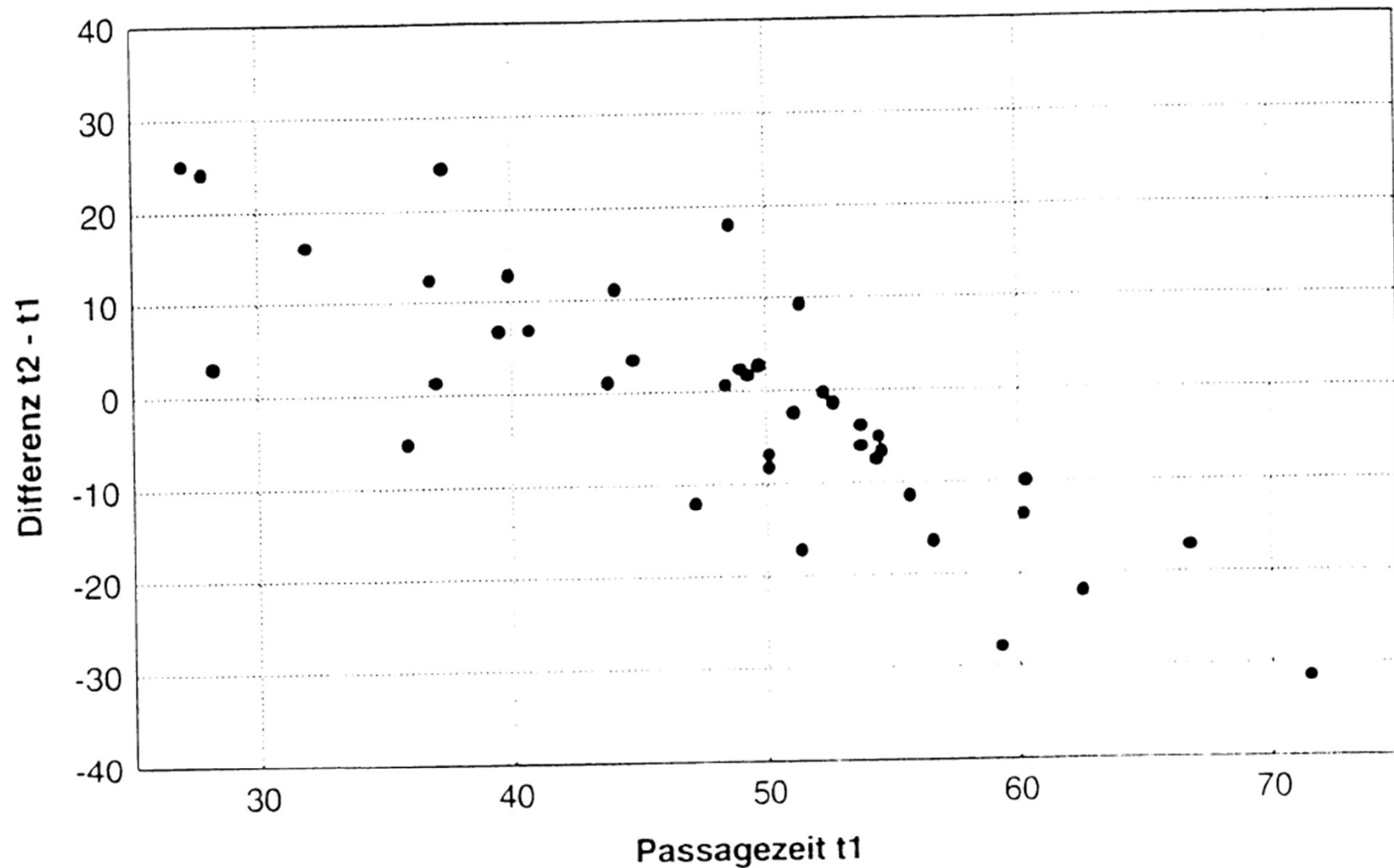


Abbildung 1.6. Magen-Darm-Passagezeiten II – Abhängigkeit der Differenz der Passagezeiten vom Ausgangswert

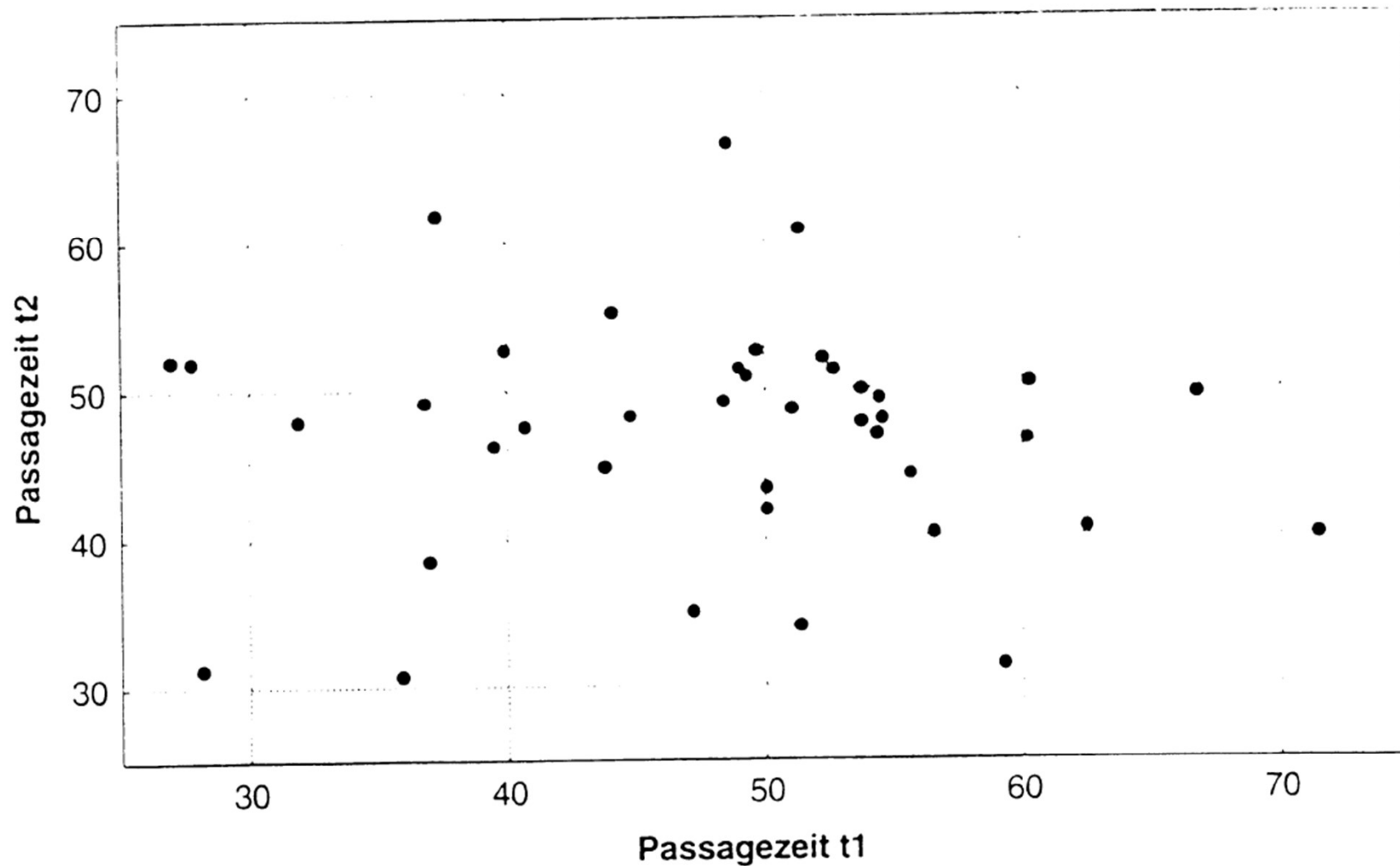


Abbildung 1.5. Magen-Darm-Passagezeiten I – Korrelation der im Abstand von 6 Monaten bestimmten Dauer der Magen-Darmpassagezeiten bei 40 Patienten

Abbildung 1.5 und Abbildung 1.6 zeigen, wie diese Ergebnisse zustande kommen. Nehmen wir an, die Passagezeiten der Patienten streuen um einen Mittelwert. Für jeden Patienten variieren sie zufällig über die Zeit, aber sie halten sich in gewissen physiologischen Grenzen. Ein Zusammenhang zwischen der ersten und der zweiten Messung ist also nicht erkennbar (Abbildung 1.5). Wenn man jedoch die Differenzen zwischen der zweiten und der ersten Messung ins Auge faßt, dann ist wegen der physiologischen Möglichkeiten klar, daß anfänglich hohe Passagezeiten eher abnehmen als zunehmen, während anfänglich lange Passagezeiten eher zunehmen als abnehmen (Abbildung 1.6). Weil dies aber aus logischen Gründen immer so sein muß, kann man nicht daraus folgern, daß bei der zweiten Messung eine „Normalisierung“ eingetreten ist. Viel eher sollte man aus der insgesamt im Mittelwert nicht veränderten Passagezeit schließen, daß Weizenkleie gar keine Wirkung auf die Darmpassage hat.

Beispiel: Wer fleißig eine Pille nimmt, lebt bestimmt.

In einer Studie, die den positiven Effekt von Clofibrat bei koronarer Herzkrankheit einer Prüfung unterzog, zeigte sich, daß die Sterberate der Patienten nach 5 Jahren unter Clofibrat die gleiche war wie unter Placebo. Um zu erkunden, ob der fehlende Effekt vielleicht auf eine zu niedrige Dosierung von Clofibrat zurückzuführen war, bestimmte man zunächst die Compliance der Patienten anhand der geschätzten Zahl der Tabletten, die tatsächlich eingenommen wurden, bezogen auf die Zahl der Tabletten, die hätten eingenommen werden sollen. Dann untersuchte man, ob die Compliance eine Auswirkung auf die Sterberate hat. Man fand, daß die Sterberate in der Clofibratgruppe bei niedriger Compliance um 10% höher war als bei hoher Compliance. Damit liegt der Schluß nahe, daß ein positiver Effekt von Clofibrat nur bei zuverlässiger Einnahme und einer entsprechend hohen Dosis zu erzielen ist.

	Clofibrat	
	Fallzahl	%Mortalität
	1103	20.0
Compliance		
< 80%	357	24.6
≥ 80%	708	15.0

Da die Studie doppelblind war, lagen Information über die Compliance für alle Patienten, also auch für die Placebogruppe vor. Überraschenderweise zeigte sich auch in der Placebogruppe, daß Patienten mit hoher Compliance eine geringere Mortalität hatten [35] (Tabelle 2.2). Gute Compliance ist offenbar eine spezielle Art des Placebo-Effektes. Das Beispiel zeigt, wie irreführend es sein kann, für den Nachweis einer Dosis-Wirkungs-Beziehung die Compliance als Determinante der Dosis zu nehmen.

Tabelle 2.2. 5-Jahres-Mortalität in der Clofibrat- und in der Placebogruppe

	Behandlungsgruppen			
	Clofibrat		Placebo	
	Fallzahl	%Mortalität	Fallzahl	%Mortalität
	1103	20.0	2789	20.9
Compliance				
< 80 %	357	24.6	882	28.2
≥ 80 %	708	15.0	1813	15.1

Compliance ist ein Problem bei allen Medikamentenprüfungen. Bredford Hill, der 1948 die erste randomisierte kontrollierte Studie auf den Weg brachte, berichtet das Gespräch eines Patienten mit dem Arzt, der gerade einen doppelblinden Crossover-Versuch durchführt.

Patient: „Warum bekomme ich denn in dieser Woche andere Tabletten als in der letzten Woche?“

Arzt: „Woher wissen Sie denn, daß das andere Tabletten sind?“

Patient: „Letzte Woche, als ich sie in's Klo geworfen habe, sind sie geschwommen. Diese Woche gehen sie unter.“

Beispiel: Der heilsame Effekt des Zigarettenrauchens.

1989 begründete der Surgeon General, der oberste Gesundheitsbeamte der USA, seine strikte Anti-Tabak-Politik mit einer Hochrechnung, derzufolge 20% aller Todesfälle in den Vereinigten Staaten auf das Rauchen zurückzuführen sind. Diese Hochrechnung basierte auf dem Cancer Prevention Survey II, einer Verlaufsuntersuchung an einer Million Amerikanern. Die Mortalitätsraten der Nichtraucher in dieser Studie wurden auf die amerikanische Bevölkerung hochgerechnet und zeigten den genannten Unterschied zur Mortalität der Gesamtbevölkerung [66]. Um die Validität dieser Schätzung zu überprüfen, verwendeten Sterling et al. [80] die gleiche Datenbasis, den Cancer Prevention Survey II, rechneten aber nicht beobachtete Mortalität der Nichtraucher, sondern die Mortalität der Raucher hoch auf die Gesamtbevölkerung der USA. Und sie stellten fest, daß über 14% aller Todesfälle zu verhindern wären, wenn alle Amerikaner Zigaretten rauchten. Also wäre nicht nur das Rauchen, sondern auch das Nichtrauchen schädlich?

Zurückzuführen ist dieses paradoxe Ergebnis darauf, daß beide Hochrechnungen von der Untersuchungspopulation des Cancer Prevention Survey ausgehen. Die Hochrechnungen zeigen, daß sowohl die Raucher als auch die Nichtraucher, also alle Teilnehmer in dieser Untersuchung eine niedrigere Mortalität haben als die Gesamtbevölkerung. Das aber bedeutet, daß die Untersuchungspopulation nicht repräsentativ ist. Einschlägige Untersuchungen zeigen, daß es in den USA überwiegend Angehörige der Mittelschicht sind, die an epidemiologischen Studien teilnehmen. Die unteren sozialen Schichten, die traditionell eine höhere Mortalität als die Mittelschicht haben, sind also in der Untersuchungspopulation unterrepräsentiert. Damit werden aber Hochrechnungen auf die Gesamtbevölkerung fraglich – auch wenn innerhalb der Untersuchungspopulation die Schädlichkeit des Rauchens durchaus erkennbar ist.

Beispiel: Diagnostischer Fehlschluß beim toxischen Schock Syndrom.

Nachdem die Kontroverse über den beobachteten Zusammenhang zwischen dem Gebrauch einer bestimmten Tampon-Marke während der Menstruation und dem toxischem Schock-Syndrom immer heftiger geworden war, führten Mary Harvey et al. [36] eine Studie durch, in der sie 218 Ärzten wiederholt eine Serie von 6 Fallgeschichten von Patienten mit mehr oder weniger abgesichertem toxischen Schock-Syndrom vorlegten. Die Fallgeschichten unterschieden sich bei der Wiederholung lediglich dadurch, daß in zufälliger Weise die Angaben „männlich“, oder „weiblich“, oder „weiblich, Menstruation“, oder „weiblich, Menstruation, Tampon-Gebrauch“ ausgetauscht waren. Die Ärzte wurden gebeten, für jede Fallgeschichte eine Diagnose zu stellen und dabei unter 22 möglichen Diagnosen zu wählen. Die größten Unterschiede in den Diagnosen traten auf bei Paaren von Fallgeschichten, die in lediglich einer ihrer beiden Versionen die Angaben „Menstruation“ oder „Tampongebrauch“ enthielten (Tabelle 2.1). Die potentielle Ursache diente also als Signalgeber, um die angenommene Wirkung zu konstatieren. Die Autoren demonstrieren damit, daß der augenscheinliche Zusammenhang zwischen Tampons und toxischem Schock-Syndrom allein durch Vorurteile bei der Diagnostik zustandegekommen sein könnte. Da derartige Vorurteile nur schwer zu beherrschen sind, bleibt es unter diesen Umständen schwierig, eine wissenschaftlich valide Untersuchung über die tatsächliche Ursache des toxischem Schock-Syndrom durchzuführen.

Tabelle 2.1. Anteil der Ärzte ($N = 218$), die eine Diagnose „toxisches Schock-Syndrom“ (TSS) stellen, in Abhängigkeit von der Charakteristik der Fallgeschichte

Diagnose zur Fallgeschichte	Charakteristik Version 1	% Ärzte, die TSS diagnostizieren	Charakteristik Version 2	% Ärzte, die TSS diagnostizieren
sicheres TSS	weiblich Menstruation Tampon	85 %	männlich	23 %
wahrscheinliches TSS	weiblich Menstruation Tampon	51 %	männlich	8 %
wahrscheinliches TSS	weiblich Menstruation Tampon	47 %	weiblich	9 %
wahrscheinliches TSS	weiblich Menstruation	38 %	männlich	9 %
wahrscheinliches TSS	weiblich	18 %	männlich	10 %
Kawasaki-Krankheit	weiblich Menstruation Tampon	21 %	männlich	3 %