CHAPTER 3

PhysPops --- The Doses in Some Massive Studies of Dose-Response

Definition of PhysPop

"PhysPop" is our abbreviation for "Number of Physicians per 100,000 Population." When pressed for space: PP. When really pressed for space: pp.

• Part 1. Purpose and Data for Our Dose-Response Studies

The titles of this chapter and of the book itself both emphasize the term "dose-response study." Such studies address questions like, "When all other things are equal, does the cancer mortality-rate rise as exposure from ionizing radiation rises?" Yes. That fact has been established for decades (Chapter 2, Part 4c). Additional proof, that ionizing radiation is a cause of Cancer, is not needed.

Then what is our interest in additional dose-response studies?

Our interest is in exploring Hypothesis-1, that specifically MEDICAL radiation --- which is readily controllable by humans --- is a highly important cause of Twentieth Century cancer-mortality in the United States. The new set of dose-response studies, contemplated for this monograph, might be able provide a basis for estimating the MAGNITUDE of that causal role. Is the role trivial, as so often claimed, or is it highly important? We decided to attempt dose-response studies based on the input described below.

1a. The Input-Data for Our Dose-Response Studies

The minimum requirements for a dose-response study include data on responses and doses, of course.

Cancer is the relevant response for testing our Hypothesis-1. Cancer mortality-rates per 100,000 population are available for the United States, by states and by the Nine Census Divisions. These rates provide our input-data on response (details in Chapter 4).

And what about input-data on dose? A fundamental premise of our studies is that the more physicians per 100,000 population, the more radiation procedures per 100,000 population will be ordered. Such procedures are initiated by a physician, even if someone else actually performs a procedure. Thus, we arrive at the premise that average radiation dose per capita FROM MEDICAL PROCEDURES during a specific year, is approximately proportional to the number of physicians per 100,000 population during the same year.

This common-sense premise is supported by the numerous authors of the 1988 and 1993 reports of the United Nations Committee on the Effects of Atomic Radiation. In their 1993 report, in Annex C on medical radiation exposures worldwide, they state (UNSCEAR 1993, pp.223-224/Para.10): "In the UNSCEAR 1988 Report, a good correlation was shown to exist between the number of xray examinations per unit of population and the number of physicians per unit of population." And they depict a linear correlation in Figure 1 (UNSCEAR 1993, p.347). The premise is also supported by the evidence already provided specifically for the USA in our Chapter 2, Part 3c, Point 1. The substantial increases, described there, in xray procedures per 1,000 population and in per capita sales of medical xray film, occurred during the period of a rapid increase in the number of physicians per 1,000 (or per 100,000) population --- namely, following federal enactment of medical entitlements in the mid-1960s.

Our input-data on dose will be numbers of physicians per 100,000 population, by Census Divisions (USA), as explained in Part 1b below. At the outset, we did not know at what year such records began to be kept, relative to the discovery of xrays in 1895. We were able to obtain data starting in 1921.

1b. PhysPop as a Surrogate for Medical Radiation Dose, by Census Divisions

Using the premise from Part 1a, we can state:

Average radiation dose (in rads) per capita from medical procedures = (k)(PhysPop), where k is a conversion-factor from physicians per 100,000 population into average number of rads received per capita during the same year. ("Rad" is defined in Appendix A.) We approximate that k has the same value nationwide at any one time. There is no requirement for the value of k to remain the same, decade after decade.

At any one time, in each of the Nine Census Divisions, the magnitude of average per capita dose from medical radiation is proportional to (k) times (PhysPop for that particular Census Division). Thus, we can (and we do) use the PhysPop values for individual Census Divisions as a surrogate for average per capita doses received in such Census Divisions. If a PhysPop value in the First Census Division is 1.43 times bigger than the PhysPop value in the Ninth Census Division, then the resulting average dose per person is 1.43 times higher in Census Region One than in Census Region Nine --- as a good approximation. PhysPop values reveal the RELATIVE size of average per capita doses in the Nine Census Divisions.

Our studies never require the quantification of k. Thus, our studies permit the possibility that average per capita dose could decrease in every Census Division, during a period when PhysPop values could simultaneously increase. For example, if PhysPop rose 2-fold while average dose-level of radiation per procedure fell 3-fold, then the average per capita dose would decrease to 2/3 of its earlier level: Dose = (k/3)(2PhysPop) = (2/3)(k)(PhysPop).

1c. Two Special Merits of Using PhysPop Values as Dose-Surrogates

Dose-response studies, based on the relative size of doses, of course can be fully as valid as studies based on absolute dose-values. Because the absolute doses from medical radiation in the past and present are highly uncertain and forever debatable (Chapter 2, Part 3), studies based on a reasonable approximation of the relative size of doses (PhysPop values) can be the MOST reliable. Indeed, one of the major scientific strengths of this monograph is its independence from anyone's estimates of absolute doses.

A second strength of the PhysPop method deserves some discussion:

Epidemiologic research on the health effects of ionizing radiation is sometimes characterized by input-data which are vulnerable to potentially biased, after-the-fact adjustments of dosage and responses, and after-the-fact exclusion of selected groups or cases as "unqualified" for retention in a study, or after-the-fact inclusion of "reserve" samples. Even retroactive shuffling of dose-cohorts --- after they have produced a dose-response --- is now a chronic practice in one of the world's most important radiation databases, the Atomic-Bomb Survivor Database (discussion in Chapter 2, Part 5c).

Such practices, as well as the fact that so much radiation research is funded by governments which are far from neutral about the hazards of ionizing radiation, necessarily create doubt about the trustworthiness of the raw databases themselves. The FIRST obligation of objective scientists is to seek assurance that they do not work with biased data which will produce misleading results. For example, few objective analysts on the smoking-issue would rely on data from a database sponsored by, and thus controlled by, the tobacco industry.

So, in the world of radiation epidemiology, the radiation studies which are presented in this monograph have a special foundation of credibility: The inputs for both dose (PhysPops) and response (Mortality-Rates) are data collected over decades by people with no conceivable intent or ability to bias the outcome of a radiation study. The data are public and not vulnerable to successful alteration.

• Part 2. Some Reasons for Expecting Our Dose-Response Concept to Fail

We were aware, at the outset, that the merits described in Part 1c could not eliminate the several reasons to bet AGAINST detecting any dose-response in such data. But researchers who demand a guarantee before they begin, rarely begin. Pessimism is paralytic. And sometimes irrational. It can be unreasonable to assume that all imaginable obstacles, to obtaining useful information, will actually materialize.

"Whatever you want to do, if you overanalyze it --- if you start looking for all the pluses and all the minuses --- you might never start." So spoke Dr. Herb Boyer, molecular biologist, and co-founder of Genentech Inc., a pioneering enterprise in the biotechnology world. The occasion: An interview on Genentech's 20th anniversary in 1996 (Boyer 1996).

Nonetheless, Part 2 will briefly describe some of the potential obstacles, as a guide to whether or not they materialized, and as a guide to some of our decisions.

2a. Inconsistent Studies on Natural Background Radiation

What made us ever imagine that a dose-response from medical irradiation might be DETECTABLE by geographical regions, when numerous attempts to find a dose-response from geographical differences in natural background radiation have been conflicting and non-definitive?

The idea probably occurred to us because of our 1981 analysis of the Frigerio paper, described in Chapter 2, Part 9b. Moreover, as a result of our work on the 1995 breast-cancer book, we had learned how MUCH medical irradiation has been used. So we thought that the average per capita dose per year from medical radiation might exceed the annual dose from natural background sources by enough to "show up." This thinking was related to the fact that medical x-rays are 2-fold to 4-fold more harmful (biologically) than the gamma rays from natural background sources, as discussed in Chapter 2, Part 7. So we decided to take the next step, and to examine the PhysPop data.

2b. The Necessity of DIFFERENCES in PhysPop Values by Census Divisions

Of course we did not know, until we obtained and studied the data in this chapter, that sufficient DIFFERENCES would exist in the doses (PhysPops) on a geographical basis. It is impossible to do a very useful dose-response study without appreciable differences in doses! Medical irradiation could be the paramount cause of the cancer-problem, and still we would obtain no hint of such a fact from our proposed dose-response study --- if the doses were about the same in all Nine Census Divisions.

A dose-response study typically plots, on a graph, a proposed cause on the horizontal x-axis, versus a proposed consequence on the vertical y-axis. If real-world evidence shows that a series of increments in the proposed cause, goes with a series of increments in the proposed consequence, the causal presumption is reasonable unless a better explanation can be demonstrated. The causal presumption is especially reasonable when the proposed cause (ionizing radiation) is already a PROVEN cause of the effect (excess cancer-mortality).

So our very first task was to find out if there would be any appreciable DIFFERENCES in the dose-input (PhysPops in the Nine Census Divisions) for our proposed study. In Part 5 of this chapter, we will discuss the range of differences we found in PhysPops, and how the range changed over the 1921-1992 period.

2c. Annual Radiation Dose versus Accumulated Radiation Dose

The chance, that a cell acquires a new (non-inherited) carcinogenic mutation due to ionizing radiation, is proportional to the cell's ACCUMULATED radiation dose. We knew at the outset that a PhysPop value for a single year would not be proportional to the average ACCUMULATED total dose in one census region, versus all other regions --- unless the regional PhysPop values retained their proportionality with EACH OTHER over time. This aspect of our studies is explored in Parts 6 and 7 (below).

PhysPops, as informative surrogates for ACCUMULATED doses, were threatened in yet another way. Even if PhysPop rankings in the various Census Divisions happened to remain stable long enough to produce some discernible differences in the radiation consequences, we needed to worry about the impact of the population's mobility.

The Potential Problem from Migration between Census Divisions

Whenever people move from a Census Division of higher PhysPop value to a Census Division of lower PhysPop value, they carry their cancer-risk with them. Because they mix their higher accumulated dose (and their higher risk of radiation-induced Cancer) with the new population's lower accumulated dose (and lower risk of radiation-induced Cancer), such migration necessarily degrades PhysPop as a measure of the relative magnitude of accumulated dose received by people dying within those two Census Divisions. And the same potential problem applies to migration from low PhysPop to high PhysPop Census Divisions. The concern would essentially vanish if all radiation-induced Cancers were delivered within 2 or 3 years after irradiation. The potential problem occurs because latency periods (delivery times) for radiation-induced Cancers are spread over decades, as discussed in Chapter 2, Part 8. As more decades pass, more people migrate between Census Divisions. By contrast, the migration which occurs WITHIN single Census Divisions creates no problem at all for our proposed dose-response studies.

2d. Distribution of the Combined Impact from Other Carcinogens

We knew at the outset, of course, that a dose-response to medical irradiation could be obscured in our proposed studies, if the combined force of carcinogens OTHER THAN MEDICAL IRRADIATION were to have an UNEQUAL impact on the cancer mortality-rates of the Nine Census Divisions. This is a nearly universal hazard in epidemiology. For example, in the Atomic-Bomb Survivor Study, one can (and must) assure that the groups receiving different doses of bomb-radiation are comparable in age-and-sex distribution. But there is no way to force the dose-cohorts to be comparable in their lifetime exposures to all non-bomb carcinogens (known and unknown) before and after August 1945. And in our own studies, there is no way to force the nine populations in the Nine Census Divisions to be comparable in their lifetime exposures to all non-xray carcinogens (known and unknown). In such studies (and many others), one hopes that providence has distributed the extraneous non-comparabilities in such a way that their combined carcinogenic force is nearly equal ("matched"in all dose-cohorts. Otherwise, these unequal impacts can distort the true dose-response between the two variables under study (Chapter 5, Part 7, and Chapter 48).

• Part 3. PhysPop Data for the Nine Census Divisions, 1921-1992

Overlapping sources exist for data on the number of physicians per 100,000 population in the USA. They include the U.S. Government, the American Medical Association, and the American Hospital Association. (We did not happen to use any AHA publications.) We have found data back as far as 1921.

3a. The "Universal" PhysPop Table: Table 3-A (Four Pages)

Table 3-A is located, of course, after the text of this chapter. It is the Universal PhysPop Table covering the years 1921-1993, for the Nine Census Divisions of the USA. The word "universal" calls attention to the fact that the PhysPops are the same no matter what cause of death is compared with them. Thus, this single table is the origin of x-axis data for numerous chapters of this book.

The table covers general practitioners and specialists combined. The details are provided in Parts 3c and 3d. We did not find data for every calendar year between 1921 and 1992. The years for which we have found data are flagged "+" in the Universal PhysPop Table 3-A. The years for which we obtained values by interpolation, are unflagged.

The data on PhysPops are often presented state-by-state in various sources. In combining data from various states, to obtain the average PhysPop value for an entire Census Divisions, we weighted each state's PhysPop value by the contemporaneous size of the state's population (details in Part 3d).

3b. Which States Belong to Which Census Divisions?

Because we were searching for data on the Nine Census Divisions from 1895 onward, the fact that Alaska and Hawaii did not become states until after World War Two seemed like a probable complication. In view of their small populations, we decided at the outset to exclude Alaska and Hawaii from consideration. For consistency, we also excluded the District of Columbia, which is not a state and whose population has always been small, too. So, these three entities are omitted from our Universal PhysPop Table 3-A. Below, we list the states (total = 48) in each of the Nine Census Divisions (from PHS 1995, p.302, for example). Populations of each Census Division, by decades, are shown in our Table 3-B.

- EAST NORTH CENTRAL: Illinois, Indiana, Michigan, Ohio, Wisconsin. 5 states.
- EAST SOUTH CENTRAL: Alabama, Kentucky, Mississippi, Tennessee. 4 states.
- MIDDLE ATLANTIC: New Jersey, New York, Pennsylvania. 3 states.
- MOUNTAIN: Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, Wyoming. 8 states.
- NEW ENGLAND: Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont. 6 states.
- PACIFIC: California, Oregon, Washington. 3 states.
- SOUTH ATLANTIC: Delaware, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, West Virginia. 8 states.
- WEST NORTH CENTRAL: Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, South Dakota. 7 states.
- WEST SOUTH CENTRAL: Arkansas, Louisiana, Oklahoma, Texas. 4 states.

3c. Evolution of Four Categories of PhysPops

Over the years, the reports of PhysPop values gradually developed multiple categories. For instance, distinction is made between federal and nonfederal physicians. Federal physicians are those on active duty with the armed forces, the Public Health Service, the Veterans' Administration, the Indian Service, and other federal agencies (Pennell 1952, p.10). Some of the distinctions developed due to "manpower" forecasts, for wartime and for new federal programs such as Medicare and Medicaid, which were enacted in 1965. By 1965, PhysPop values came in four varieties, based on:

1. • Total physicians --- active + inactive, federal + nonfederal, in the USA and its possessions (Canal Zone prior to 1980, Pacific islands, Puerto Rico, Virgin Islands). Specified in AMA 1993, Table A-16, p.32.

2. • Total patient-care physicians, federal and nonfederal. Not available until 1965 (AMA 1993, Table A-16, p.32).

3. • Total nonfederal physicians --- active + inactive --- in the 50 states and D.C. Specified in AMA 1993, Table A-17, p.33.

4. • Nonfederal patient-care physicians, 50 states and D.C.

The relative sizes of these four types of PhysPops --- for the years 1965, 1970, 1975, 1980, 1985, and 1992 --- are graphed in AMA 1993, Figure 4, p.15. We have reproduced the AMA graph at the end of this chapter, as Figure 3-D. It is evident that the four types are very tightly correlated with each other.

3d. Sources Used for PhysPop Values from 1921 Onwards

Our choices for the Universal PhysPop Table 3-A were determined by what was available to us in the literature either by states or by Census Divisions. By states, the AMA tables offer only one "choice": Total nonfederal PhysPops. As of 1949, only a very small share of the total was inactive (see below). The fraction was low also in 1975, 1985, 1990 (see Part 3e).

PhysPops for 1921-1949

PhysPop data from 1921 through 1949 are from Pennell 1952 in our Reference List. This is Public Health Service Publication 263, Section 1, prepared by Maryland Y. Pennell and Marion E. Altenderfer, and entitled "The Health Manpower Source Book Section 1. Physicians." Its Table 2 (at page 14) is Physician-Population Ratios in the United States, by Region and State: 1921-1949. Pennell and Altenderfer based their table on the following sources in our Reference List: AMA 1950, Census Bureau 1951, and Linder 1947.

Pennell's Table 2 does not specify any subset of physicians. By comparing numbers in Pennell's Table 2 with numbers in Pennell's Tables 1 and 4, we can establish that the PhysPops in Table 2 are for total physicians (active + inactive, federal + nonfederal).

Separately, and only for the year 1949, Pennell provides the composition of the total 201,277 physicians: 179,041 active nonfederal + 12,536 federal + 9,700 retired.

PhysPops for 1959

PhysPop data for 1959 are from Stewart 1960 in our Reference List. This is Public Health Service Publication 263, Section 10, prepared by William H. Stewart and Maryland Y. Pennell, and entitled "The Health Manpower Source Book Section 10. Physicians' Age, Type of Practice, and Location." On page 26 is its table, Physician-Population Ratio in Each State, and Age of Physicians: Non-Federal Physicians per 100,000 Civilian Population, 1959. Stewart 1960 (p.1) bases the number and location of physicians (mid-1959) on data supplied to the Public Health Service by the American Medical Association, and rates per 100,000 population, by states, on mid-1959 population data from the Census Bureau (1959). We used 1959 population data from Grove 1968 (Table 74) in order to obtain population-weighted PhysPop values for the Nine Census Divisions.

PhysPops for 1963, 1965, 1970, 1975, 1980, 1981

PhysPop data for 1963, 1965, 1970, 1975, 1980, and 1981 all are taken from AMA 1982 in our Reference List, Table A-7, Non-Federal Physicians, Civilian Population, Physician-Population Ratios for Selected Years 1963-1981. That table provides all the data we need to calculate population-weighted PhysPop values for the Nine Census Divisions.

PhysPops for 1983

PhysPop data for 1983 are taken from AMA 1986 in our Reference List, Table A-9, Non-Federal Physicians, Civilian Population, and Physician/Population Ratios for Selected Years 1963-1985. That table provides all the data we need to calculate population-weighted PhysPop values for the Nine Census Divisions.

PhysPops for 1985, 1990, 1993

PhysPop data for 1985, 1990, and the start of 1993 are taken from AMA 1994 in our Reference List, Table A-18, Non-Federal Physicians, Civilian Population, Physician/Population Ratios for Selected Years 1970-1993. That table provides all the data we need to calculate population-weighted PhysPop values for the Nine Census Divisions.

3e. Difference between PhysPops from Table 3-A and from PHS 1995

In "Health, United States, 1995 (PHS 1995, pp.218-219), there is Table 97 which presents PhysPops for 1975, 1985, 1990, and 1994 by Census Divisions and by states. We note the word "active" in Table 97's title: Active Nonfederal Physicians and Doctors of Medicine in Patient Care per 10,000 Civilian Population ... 1975, 1985, 1990, and 1994. Of course, we multiply PhysPop values in Table 97 by ten, in order to convert them to the more customary "per 100,000" population. In our Universal PhysPop Table 3-A, the PhysPop values come from the combination of active plus inactive physicians. Not surprisingly, our PhysPops for 1975, 1985, and 1990 are higher than those presented in PHS 1995, Table 97.

Would analysts reach the same conclusions that we do, about the relationship of PhysPops with biological phenomena (such as cancer mortality-rates), if they used the ratios from Table 97, instead of the ratios from our Universal PhysPop Table 3-A?

The answer is yes. They would reach the same conclusions, because the correlations between the two sets of data are so very high. We demonstrate this by the three regression analyses which follow and which produce correlation coefficients (R) of 0.9916, 0.9855, and 0.9863. Our studies rely on the RELATIVE magnitudes rather than absolute magnitudes of the nine PhysPop values (Parts 1b and 1c), and such high correlations between Table 97 and Table 3-A mean that the relative magnitudes among the PhysPop values are extremely similar in Table 97 and Table 3-A. (Readers who are unfamiliar with linear regression analysis will find an introduction to the topic in Part 7 of this chapter, and more explanation in Chapter 5, Part 5).

Below, listed by the Nine Census Divisions, are the PhysPop values per 100,000 population from PHS 1995, Table 97 (including active doctors of osteopathy), and to the right of them, the values for the matching Census Divisions from our Universal PhysPop Table 3-A.

YEAR = 1975	Tab 97	Tab 3-A	YEAR = 1975	
Pacific	179	208	Regression Output:	
New England	191	215	Constant	-11.5257
West North Central	133	141	Std Err of Y Est	5.2312
Mid-Atlantic	195	213	R Squared	0.9832
East North Central	139	146	No. of Observations	9
Mountain	143	156	Degrees of Freedom	7
West South Central	119	128		
East South Central	105	117	X Coefficient(s)	1.1784
South Atlantic	140	156	Std Err of Coef.	0.0582
Unweighted Avg.	149.3	164.4	R = 0.9916	
Ratio (Tab3A/Tab97) =	1.10			
YEAR = 1985	Tab 97	Tab 3-A	YEAR = 1985	
Pacific	225	256	Regression Output:	
New England	267	293	Constant	-13.6604
West North Central	183	186	Std Err of Y Est	8.5884
Mid-Atlantic	261	276	R Squared	0.9712
East North Central	193	195	No. of Observations	9
Mountain	178	193	Degrees of Freedom	7
West South Central	164	171		
East South Central	150	162	X Coefficient(s)	1.1391
South Atlantic	197	216	Std Err of Coef.	0.0741
Unweighted Avg.	202.0	216.4	R = 0.9855	
Ratio (Tab3A/Tab97) =	1.07			
YEAR = 1990	Tab 97	Tab 3-A	YEAR = 1990	
Pacific	234	265	Regression Output:	
New England	290	320	Constant	-14.3549
West North Central	198	203	Std Err of Y Est	8.7876
Mid-Atlantic	284	298	R Squared	0.9729
East North Central	206	209	No. of Observations	9
Mountain	193	208	Degrees of Freedom	7
West South Central	178	184		
East South Central	168	182	X Coefficient(s)	1.1342
South Atlantic	217	234	Std Err of Coef.	0.0716

Table continues, next page

Unweighted Avg.	218.7	233.7	R =	0.9863
Ratio (Tab3A/Tab97) =	1.07			

• Part 4. Designation of the "High-5" and "Low-4" Census Divisions

In the Universal PhysPop Table 3-A, the PhysPop values for 1921 are presented in order of size, from the highest value in the Pacific Division (165.11 physicians per 100,000 population) to the lowest value in the South Atlantic Division (110.32 physicians per 100,000 population).

We can (and do) retain the 1921 sequence of the Census Divisions, even though the PhysPop values do not remain ranked in that order during all subsequent years. Use of the 1921 sequence leads to two terms used in Table 3-A (and used also in our tables of mortality rates): High-5 and Low-4.

Definition of "High-Five" and "Low-Four" Census Divisions

• The term "High-5" always refers to the first Five Census Divisions listed in the Universal PhysPop Table for 1921: Pacific, New England, West North Central, Mid-Atlantic, East North Central. Since PhysPop values are surrogates for average per capita dose from medical radiation (Part 1b), the term High-5 refers to the Census Divisions with the highest average doses per capita from medical irradiation in 1921. Our shortest abbreviation is Hi5.

• The term "Low-4" always refers to the last Four Census Divisions listed in the Universal PhysPop Table for 1921: Mountain, West South Central, East South Central, South Atlantic. Since PhysPop values are surrogates for average per capita dose from medical radiation (Part 1b), the term Low-4 refers to the Census Divisions with the lowest average doses per capita from medical irradiation in 1921. Our shortest abbreviation is Lo4.

A Point to Keep in Mind, and the Next Question

A point to keep in mind is that High-5 and Low-4 are two Census-Division sets whose members were determined by their PHYSPOP rankings in 1921, not by their cancer mortality-rates in 1921. What happens to High-5 and Low-4 PhysPop values, as the interval after 1921 grows ever longer? We will explore that issue in Part 5, below.

• Part 5. Dose-Differences: What Does the Evidence Show?

For PhysPop, which is the dose-surrogate in our dose-response studies, there are pages of entries in the Universal Table. Do these entries reflect sufficient DIFFERENCES in dose among the Nine Census Divisions --- and are differences maintained long enough in their rank order --- to produce detectable differences in cancer consequences?

To facilitate getting a grasp on the issue of PhysPop differences and their duration, we calculated average values for the High-5 and Low-4 Divisions in each column. In the Universal PhysPop Table 3-A, the nine main entries for the Nine Census Division are weighted averages (Part 3a), but the High-5 and Low-4 averages (located beneath the main entries) are not population-weighted. They are provided just as approximations which can supply an overview for each particular year, and for changes over time. Table 3-A also shows the ratio of Hi5/Lo4.

5a. Revelations about PhysPop Behavior, 1921 through 1990: Figure 3-A

Figure 3-A presents two graphs which plot annual PhysPop behavior from 1921 through 1990, in terms of High-5 and Low-4 groupings. These graphs provide a visual overview. No values from the graphs are ever used in calculations. Therefore, readers need not worry at all about some minor differences between the graphs and Table 3-A. The graphs reflect our early exploration --- before we had every final PhysPop value of Table 3-A --- of a question which would determine whether or not to proceed with the project: Was there a persistent dose-difference between the Census Divisions?

The upper graph plots the annual High-5 and Low-4 averages, separately, 1921-1990. It is clear that they are relatively flat until almost 1970, when both of them take off like rockets to much higher values. The steep rise in PhysPop values occurred after the 1965 enactment of Medicare and other federal programs.

The lower graph plots the annual RATIOs of average High-5 PhysPop over average Low-4 PhysPop. The ratio tells us how many-fold larger High-5 PhysPop is, compared with Low-4. At a value of 1.0, of course, their magnitudes would be equal. The graph produces some very important information.

First, because the ratios never fall to 1.0, it immediately assures us that average annual High-5 doses always remained higher than the average annual Low-4 doses. So there has been an annual dose-difference, from 1921 to 1990.

Second, the graph of Hi5/Lo4 PhysPop ratios shows us that there was an extended period of relative PhysPop stability from about 1933 through 1968. From Table 3-A (which has the final PhysPop values), we know that the ratio of High-5 PhysPop over Low-4 PhysPop was 1.37 in 1933; then the ratio rose to a maximum value of 1.46 in 1940; by 1968 the ratio had returned to 1.37. In other words, between 1933 and 1968, the range for the Hi5/Lo4 ratio stayed within the limits of 1.37 and 1.46.

5b. What the Ratios Fail to Show

The Hi5/Lo4 ratios obscure the full magnitude of the differences between PhysPops. Although the maximum Hi5/Lo4 ratio is 1.46, the ratio comes from averages. Two examples illustrate the point. In 1921, when the Hi5/Lo4 ratio was only 1.18, the ratio of Pacific Division over South Atlantic was (165.11 / 110.32), or 1.50. In 1950, when the Hi5/Lo4 ratio was 1.44, the ratio of the Mid-Atlantic Division over East South Central was (168.81 / 83.25) = 2.03.

In addition, the Hi5/Lo4 ratios are crude enough to obscure shifts of PhysPop rank WITHIN both the High-5 and the Low-4 groups. Therefore, Box 1 provides a separate study of changes in PhysPop ranking for the 1921-1990 period. What emerges from Part 2 of Box 1 is that there is remarkable stability in PhysPop ranking, when the Nine Census Divisions are viewed as three "Trios": TopTrio, MidTrio, LowTrio. For example, in the 1931-1990 period, only two of the Nine Divisions (West North Central and South Atlantic) ever "migrate" from their 1940 Trio into another Trio. (Details in Box 1.)

• Part 6. "Lockstep" --- The Ideal Relation among All Sets of PhysPops

The formal definition of "lockstep" is: A method of marching in such close file that the corresponding legs of the marchers must keep step precisely.

We are going to bend the term, so that "lockstep" refers to a set of PROPORTIONS (ratios) whose values persist unchanged through time. For example, PhysPop "lockstep" would mean that the proportions observed among the nine PhysPop values in 1921, and the proportions observed in every subsequent year, are the same. PhysPop "lockstep" would mean that the RELATIVE magnitudes are constant among the nine PhysPop values, even when the absolute values rise or fall. Part 6b will provide an illustration.

Box 1 already demonstrates that PERFECT "lockstep" for PhysPop values does not occur, for perfection would tolerate no changes in Hi5/Lo4 ratios over time (see Part 7d) and no changes in rank order of the Divisions over time.

6a. The Ideal Data for Our Proposed Study

Researchers always wish for "better data." Under ideal circumstances for our inquiry, no migration of populations from one Census Division to another would have occurred after 1895, and for the entire century, the PhysPops of the Nine Census Divisions would have retained a fixed proportionality with each other: "Lockstep."

Under such circumstances, the nine average doses of medical radiation, accumulated by any particular year, would always be in that fixed proportion to each other --- regardless of their absolute values in rads. And the nine irradiated populations would gradually DELIVER the consequent radiation-induced cancers in the same Census Division where they were irradiated --- in proportion to dose (Chapter 5, Part 5d). The changing age-distribution of the population since 1895 would not distort that expectation because cancer mortality-rates, by Census Divisions, are age-adjusted to a fixed year (Chapter 4).

A Note about "Ideal" Data

In this chapter, and later, we sometimes refer to "ideal" data or circumstances. We feel impelled to emphasize, for students who may not have done any research yet themselves, that the term "ideal" does not imply any bias or passion. To imagine conditions "exactly as one would desire" (see below), unclouded by real-world perturbations, can be so crucial to elucidating a topic that it is a regular feature of science.

For example, chemists and physicists refer to "an ideal gas," "ideal conditions," and "the perfect gas law." We quote from Mahan's "University Chemistry" text (Mahan 1975, p.43-44): "The expression PV = nRT is obeyed by all gases in the limit of low densities and high temperatures --- 'ideal' conditions under which the forces between molecules are of minimum importance. Consequently, [PV = nRT] is known as the perfect gas law, or the ideal gas equation of state."

In other words, the law is valid under ideal conditions, but does not make perfect predictions under real-world conditions. This use of the word "ideal" is in full harmony with the dictionary definition which says that "ideal" means: "Existing as an idea, model, or archetype ...; thought of as perfect or a perfect model; exactly as one would desire ...; having the nature of an idea or conception; identifying or illustrating an idea or conception" (Webster 1954, p.720).

6b. Figure 3-B: Retention of Perfect Proportionality ("Lockstep")

With respect to evaluating retention (or non-retention) of PhysPop proportionalities through time, we will use Figures 3-B and 3-C as illustrations. (The term "linear regression," in the titles of these two figures, may be unfamiliar to some readers. But the point of Part 6b can be understood without any understanding of linear regression.) Figures 3-B and 3-C each compare the set of 1921 PhysPops with the set of 1940 PhysPops, but in different ways. When readers understand the two figures, they will understand our Table 3-C, which shows with great simplicity how 21 sets of PhysPops, from 1921 through 1993, compare with EVERY OTHER set of PhysPops through most of this century.

In Figure 3-B, Column B presents the 1921 PhysPop values and Column C presents the 1940 PhysPop values, from the Universal PhysPop Table 3-A. The numerical values for South Atlantic changed from 110.32 to 100.74. The ratio (1940 / 1921) is 100.74 / 110.32, or 0.9131617.

If the nine PhysPops had the same proportions with each other in 1940 as they had with each other in 1921, we could multiply every 1921 PhysPop value by 0.9131617 to discover what the values would have been in 1940. We put these "ideal" values in Column D of Figure 3-B. Column D entries = (Column B entries times 0.9131617).

Some Consequences of Retaining PERFECT Proportionality

The D-Column values in Figure 3-B are the "ideal" values which we would have preferred to find in 1940. We would have preferred them to the REAL values in Column C, because every value in Column D still stands in the same proportion to every other value in Column D, as every value in Column B stands to every other value in Column B. We can demonstrate this "lockstepping" for any two Census Divisions. For example:

- (WNoCent 1940 Ideal / Mountain 1940 Ideal) = (128.69 / 123.62) = 1.041.
- (WNoCent 1921 / Mountain 1921) = (140.93 / 135.38) = 1.041.

And because the sets of real 1921 data and ideal 1940 data have the same internal proportionalities, it is also true that cross-ratios for every pair must be the same. Example:

- (NewEngl 1940 Ideal / NewEngl 1921 Real) = (129.89 / 142.24) = 0.913.
- (WSoCentral 1940 Ideal / WSoCentral 1921 Real) = (114.28 / 125.15) = 0.913.

And because the sets of real 1921 data and ideal 1940 data have the same internal proportionalities (we have created perfect "lockstep"), the two sets of data have a perfect linear correlation WITH EACH OTHER. Part 7 discusses "perfect linear correlation" and linear regression analysis, for readers who are unfamiliar with these terms.

• Part 7. "Lockstep" --- Reality-Checks by Regression Analysis

The technique of data regression is a branch of mathematics which can evaluate the correlation between two sets of values (for example, a set called "x" and a set called "y"). Regression analysis will be covered in considerably more detail in Chapter 5 (Parts 5, 6, 7). For now, we need touch on only a few aspects of regression analysis.

7a. Equation of Best Fit, Line of Best Fit, and the R-Squared Value

In linear regression analysis, the input data are a finite set of x-values and the corresponding y-values --- as shown in Columns B and D of Figure 3-B. The output includes three values of interest to us here: The X-Coefficient, the Constant, and the R-squared value.

Equation of Best Fit: How It Relates to Part 6b

In linear regression analysis, the equation of best fit is the equation for a straight line: (y) = (m * x) + (c). Note: * is the symbol for multiplication in this book. The regression output (boxed in Figure 3-B) provides the values for "m" (the X-Coefficient) and for "c" (the Constant). Users of the equation can then specify additional values for "x" (values additional to the regression's input values) and calculate what the corresponding values for "y" would be if (repeat, if) there were a PERFECT correlation between the x-values and the y-values.

Example: If x = 80 (a value NOT in Col.B), what would the matching y-value be? We use the equation for a straight line: y = (X-Coefficient * x) + (Constant). The boxed output in Figure 3-B tells us the X-Coefficient = 0.91316 --- a number already seen in Part 6b. And the output tells us that the Constant = zero. So, when x = 80, y = 73, because: y = (0.91316 * 73) + zero.

This example is only what we already demonstrated in Part 6b --- except 80 is an ADDITIONAL value of x not used in Part 6b. The X-Coefficient in Part 6b is 0.91316 --- we just didn't give it the formal name there. And because we made x and y directly proportional in Part 6b --- when we said (y = 0.91316 * x) --- then zero is the only possible value for the Constant. Thus, it is no surprise at all that the regression output produced zero as the value of the Constant.

Line of Best Fit, and Graphing

In making x,y graphs, it is customary to measure the x-values along the horizontal axis, and to measure the y-values along the vertical axis.

The line of best fit (the straight line seen in Figure 3-B) simply depicts a long series of x,y pairs, calculated by using the equation of best fit. The point, which depicts y=73 when x=80, is part of the straight line in Figure 3-B.

The R-Squared Value: A Key Measure of Correlation

Regression output also provides a value for R-squared, which is the output of real interest in this chapter. The R-squared value is a measure of how closely the x-input and the y-input are correlated. Only a PERFECT correlation has an R-squared value of 1.00. Imperfect correlations produce R-squared values between 1.00 and zero. The value "R" --- also called the correlation coefficient (Part 3e) --- is the square root of R-squared.

Since we insured in Part 6b that our x,y pairs are perfectly proportional to each other, they are also perfectly correlated with each other. And thus it is no surprise at all that the regression output in Figure 3-B produces an R-squared value of 1.00. When R-squared = 1.00, every pair of x,y values sits right upon the line of best fit, with no scatter. In Figure 3-B, the nine boxy symbols are indeed upon the line of best fit.

7b. Figure 3-C: Degradation of Perfect Proportionality

Figure 3-C moves from the "ideal" world, depicted in Figure 3-B, into the real world. Figure 3-C shows no "ideal" values. It shows only real-world input-data: The PhysPop set of 1921 and the PhysPop set of 1940.

The two graphs in Figure 3-C look quite different from the graph in Figure 3-B. The nine boxy symbols show some SCATTER around the lines of best fit. The scatter reflects the inferior correlation compared with the "ideal" correlation (R-sq = 0.58 here, compared with 1.00 in Figure 3-B).

Thus, R-squared is an evaluation of how much the 1940 PhysPop values have strayed from the proportions which they had with each other in 1921. Quite obviously, the 1921 and 1940 sets of PhysPops are not in perfect "lockstep."

A Point about Correlations

If regression analysis is employed to study a cause-effect relationship, it is customary to designate the proposed cause as the x-axis variable. However, the correlation between two sets of numbers is whatever it is, independent of human choices to call one set "x" and the other set "y." Figure 3-C demonstrates this point by reversing the designations of the two sets of PhysPops. The R-squared values in both figures turn out the same, as they must. However, other things have changed --- such as the Constant (the value of the y-intercept) and the X-Coefficient (the slope of the best-fit line).

7c. Table 3-C: How Sets of PhysPops Correlate through Time

Table 3-C is "How Sets of PhysPops Correlate through Time." At its top are 21 sets of PhysPop values. They are the input data for approximately 200 separate regression analyses, whose R-squared values are reported in the body of Table 3-C.

Because Table 3-C (like Figure 3-A) was part of our early exploration, we had not yet obtained all the PhysPop sources which we subsequently obtained. So, not every PhysPop value in Table 3-C is an exact match for the corresponding final value in Table 3-A. The differences often come from a mixture in Table 3-C of the four different types of PhysPop values described in Part 3c. The purpose of Table 3-C was to ascertain if PhysPops were hopelessly deviant from "lockstepping" --- and since the four types of PhysPop values are so highly correlated with each other, Table 3-C is not misleading. When we undertook our subsequent dose-response studies, we used PhysPop values only from Table 3-A --- as readers can verify for themselves.

Due to Table 3-C's early origin, it does not put the Census Divisions in the same sequence as Table 3-A. Of course, the sequence has no impact whatsoever on the regression output, as long as the x and y sets of PhysPops are in the SAME sequence with respect to Census Divisions.

The Grid of R-Squared Values

Beneath the raw PhysPop data is a grid of R-squared values. For instance, where the COLUMN for 1980 intersects the ROW for 1934, the R-squared value of 0.72 comes from the regression output when the 1980 PhysPops (directly above) are regressed on the 1934 PhysPops (in a column far to the right). The R-squared value would be the same if we had regressed the 1934 PhysPops (as the y-set) on the 1980 PhysPops (as the x-set), as pointed out in Part 7b.

Readers can quickly orient themselves in Table 3-C by knowing that, when the PhysPops of 1921 are regressed upon the PhysPops of 1921, there has to be a PERFECT correlation --- and it shows up as an R-squared value of 1.00 where the COLUMN for 1921 intersects the ROW for 1921.

Because Table 3-C describes every comparison between two sets of PhysPops by an R-squared value, everyone can readily see the decrement in "lockstepping" over any chosen interval of time. The approximately 200 regression analyses are not shown.

7d. Consistency between Figure 3-A and Table 3-C

If successive PhysPop sets had retained a fixed proportionality ("lockstep") over time, the Hi5/Lo4 ratio depicted in Figure 3-A would be perfectly flat. The ratio would be the same, year after year. We can illustrate this quickly.

The Hi5/Lo4 PhysPop ratio for 1921 is 1.18 --- provided in the Universal Table 3-A. The "ideal" 1940 values from Figure 3-B (Column D) reflect perfect "lockstep" with the 1921 values. We compute the Hi5 average PhysPop as 131.792. The Lo4 average PhysPop is 112.00. The Hi5/Lo4 PhysPop ratio is (131.792 / 112.00), or 1.18 for the "ideal" entries too. Change in Hi5/Lo4 PhysPop ratios, over time, reflects DEVIATION from "lockstep."

In Figure 3-A and in the text (Part 5a), we pointed to the period of 1933 through 1968 as a period when the Hi5/Lo4 PhysPop ratio was relatively constant. This means that Table 3-C should show high R-squared values during this same period, in the vertical column for 1967. It does. The lowest R-squared value is 0.82, at the intersect of the 1967 column with the 1934 row.

• Part 8. Two Crucial Aspects of PhysPop History

Earlier in this chapter (Part 2b), we pointed out that our proposed dose-response studies require the existence of appreciable DIFFERENCES in PhysPops (our dose-surrogates) among the Nine Census Divisions. PhysPop history might NOT have delivered differences. It is just happenstance that such differences occurred (Parts 5a and 5b).

It is also happenstance that, during the years after the introduction of medical radiation, chaos did NOT characterize the relationships between successive sets of PhysPops. Chaos would have prevented PhysPops from representing relative ACCUMULATED dose-differences in the Nine Census Divisions. Although the R-squared value (in Table 3-C) of 0.58 between the 1921 PhysPops and the 1940 PhysPops is not "great," it's far from being a value of 0.02. By 1927-1929, the correlations with 1940 become very respectable. And for the entire stretch from 1933-1967, successive sets of PhysPops were close to retaining "lockstep" proportionality with each other (Table 3-C, Figure 3-A).

If PhysPop history had not met the requirements for dose-response studies, it might have been forever impossible for anyone to detect the particular consequences which are uncovered in this book from the introduction of radiation into medicine.

>>>>>>>>>>

Box 1 of Chap. 3

Summary of PhysPop Values by Decades, and Their Ranking by Census Divisions.

Related text = Part 5b.

• **Part 1. Census Divisions, in our permanent order, with corresponding PhysPop values.**

From Table 3-A.	1921 PhysPop	1931 PhysPop	1940 PhysPop	1950 PhysPop	1960 PhysPop	1970 PhysPop	1980 PhysPop	1990 PhysPop
Pacific	165.11	159.97	159.72	148.60	158.74	183.83	235.84	265.09
New England	142.24	142.35	161.55	162.51	164.37	186.51	254.37	319.88
West North Central	140.93	126.50	123.14	120.06	111.25	123.77	165.86	202.78
Mid-Atlantic	137.29	140.82	169.76	168.71	162.65	192.00	237.41	297.79
East North Central	136.06	128.59	133.36	123.69	114.56	127.17	169.79	208.54
Mountain	135.38	118.89	119.89	119.38	112.93	137.27	177.76	208.20
West South Central	125.15	105.95	103.94	101.34	101.65	113.20	153.18	184.34
East South Central	119.76	96.73	85.83	83.05	88.00	100.89	139.51	182.42
South Atlantic	110.32	99.59	100.74	99.07	105.36	130.70	187.22	234.48
Average ALL	134.70	124.38	128.66	125.16	124.39	143.93	191.22	233.72
Average High-Five	144.33	139.65	149.51	144.71	142.31	162.66	212.65	258.82
Average Low-Four	122.65	105.29	102.60	100.71	101.99	120.52	164.42	202.36
Ratio (Hi5/Lo4)	1.18	1.33	1.46	1.44	1.40	1.35	1.29	1.28

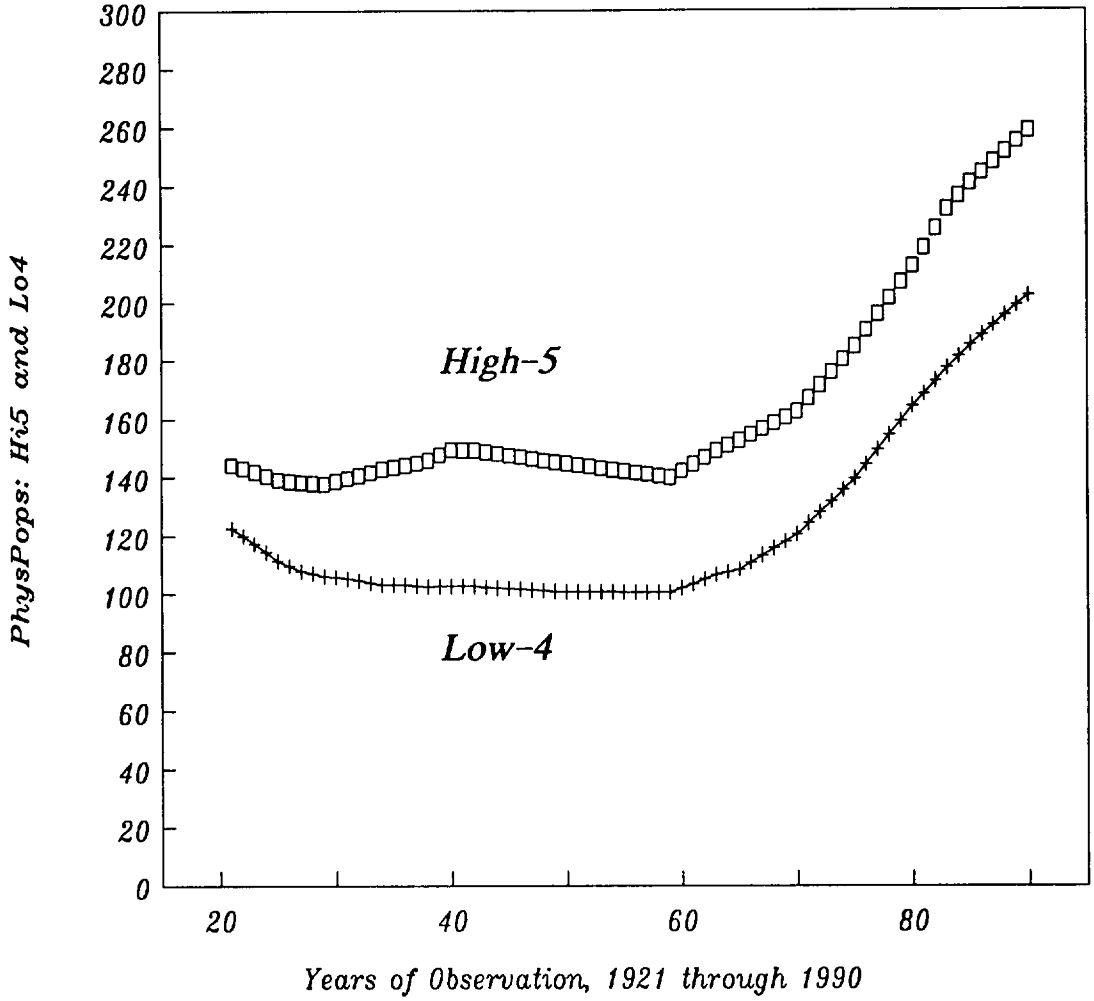
• **Part 2. Census Divisions, in shifting order, sorted by descending PhysPop values.**

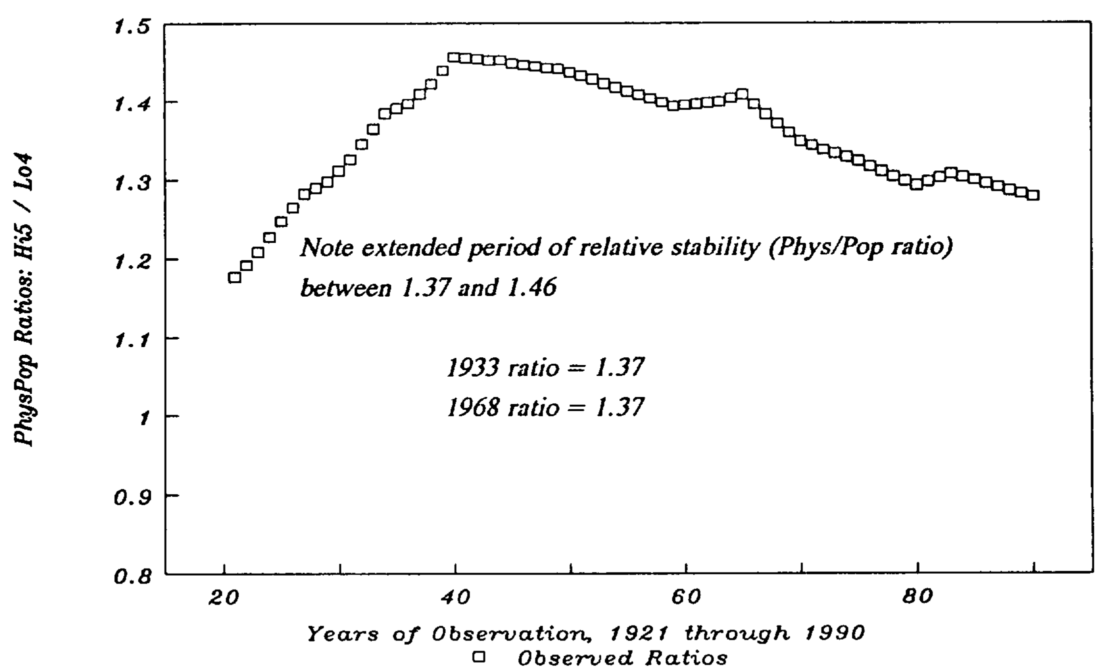
	1921 PhysPop	1931 PhysPop	1940 PhysPop	1950 PhysPop	1960 PhysPop	1970 PhysPop	1980 PhysPop	1990 PhysPop
Top Trio	Pac NewEng WNoCen	Pac NewEng MidAtl	MidAtl NewEng Pac	MidAtl NewEng Pac	NewEng MidAtl Pac	MidAtl NewEng Pac	NewEng MidAtl Pac	NewEng MidAtl Pac
Mid Trio	MidAtl ENoCen Mtn	ENoCen WNoCen Mtn	ENoCen WNoCen Mtn	ENoCen WNoCen Mtn	ENoCen Mtn WNoCen	Mtn SoAtl ENoCen	SoAtl Mtn ENoCen	SoAtl ENoCen Mtn
Low Trio	WSoCen ESoCen SoAtl	WSoCen SoAtl ESoCen	WSoCen SoAtl ESoCen	WSoCen SoAtl ESoCen	SoAtl WSoCen ESoCen	WNoCen WSoCen ESoCen	WNoCen WSoCen ESoCen	WNoCen WSoCen ESoCen

Above, in Part 2, where the Nine Census Divisions are sorted by descending PhysPop values, we have labeled them as three "Trios": Top, Mid, Low --- reflecting, RELATIVELY, the highest to lowest average per capita dosage from medical radiation.

During the 1931-1990 period, only two of the nine Divisions (West North Central and South Atlantic) ever "migrated" from their 1931 Trio into an adjacent Trio. Measured in terms of Trios, remarkable stability occurred for sixty years in PhysPop ranking. When the overview includes the 1921 values, then the Mid-Atlantic Division becomes a migrant too, and West North Central makes an additional move.

Figure 3-A.
Behavior of Hi5 and Lo4 PhysPops through Time.





Related text = Part 5a.

Figure 3-B.

Complete "Lockstepping" of PhysPop Proportions over Time.

Linear Regression of Two Perfectly Correlated and Perfectly Proportional Sets of Data.

Related text = Part 6b.

• For the regression analysis below, the x-variable input is Col.B: Actual PhysPop values for 1921. The y-variable input is Col.D: Ideal (synthetic) PhysPop values for 1940. Output for this regression is shown to the right. Both input and output are depicted by the graph. For discussion of regression analysis and its depiction, please consult Part 7 of the text and the Index.

• The nine boxy symbols on the graph depict the nine pairs of input from Columns B and D. The line of best fit depicts the output for this perfect correlation. (R-squared = 1.00). All nine boxes sit right on the line, with no scatter. Boxes overlap when input-pairs have similar values. Because perfect proportionality exists between Columns B and D, the Constant = 0 in the best-fit equation. The line of best-fit goes right through the origin (y = 0, when x = 0).

Col.A Census Division	Col.B Real 1921 Phys Pops	Col.C Real 1940 Phys Pops	Col.D Ideal 1940 Phys Pops
Pacific	165.11	159.72	150.77
New England	142.24	161.55	129.89
West North Central	140.93	123.14	128.69
Mid Atlantic	137.29	169.76	125.37
East North Central	136.06	133.36	124.24
Mountain	135.38	119.89	123.62
West South Central	125.15	103.94	114.28
East South Central	119.76	85.83	109.36
South Atlantic	110.32	100.74	100.74

Data Regression

Regression of Ideal 1940 PhysPops (Col.D) upon Real 1921 PhysPops (Col.B)

Regression Output:	
Constant	0.000000000
Std Err of Y Est	0.0000014617
R Squared	1.000000
No. of Observations	9
Degrees of Freedom	7
X Coefficient	0.9131617114
Std Err of Coefficient	0.0000000332

Regression of "Ideal" PhysPop 1940 upon Real PhysPop 1921

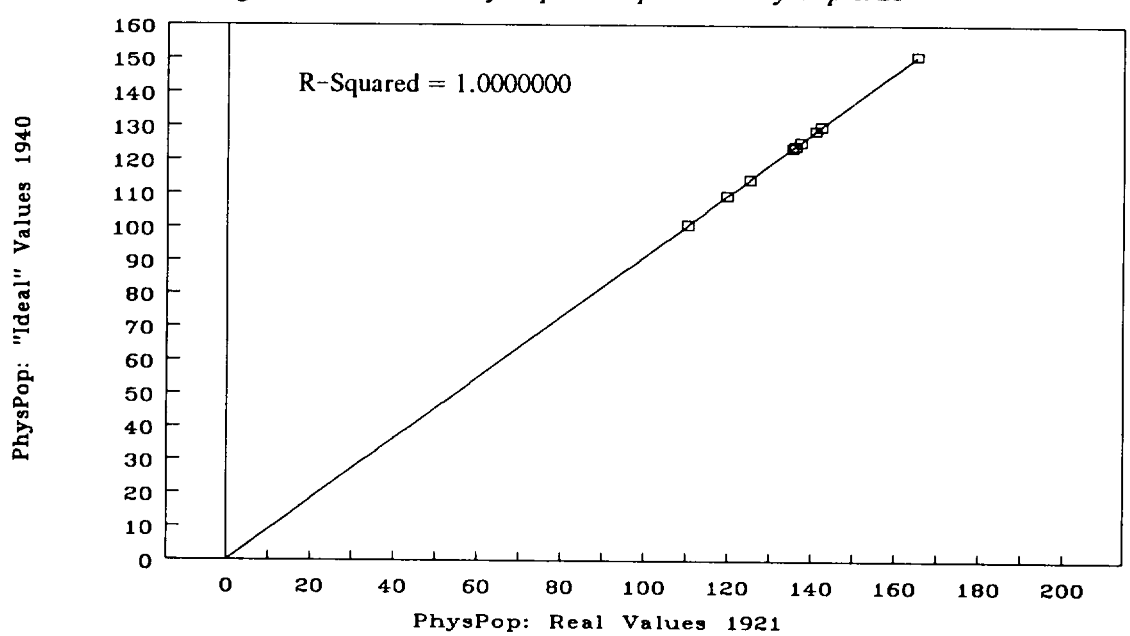


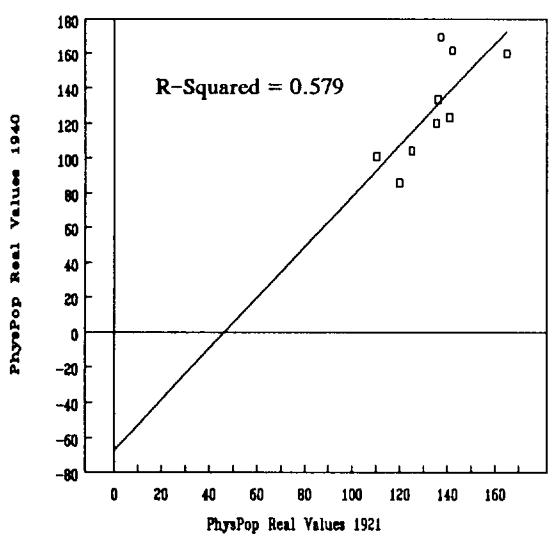
Figure 3-C.
Imperfect Retention of PhysPop Proportions over Time.
Linear Regressions with 1921 and 1940 PhysPops from Table 3-A.

• For the first regression analysis below, the x-variable input is the set of real PhysPop values for 1921. The y-variable input is the PhysPop set of 1940. For the second regression analysis, we switch. The x-input is 1940 and the y-input is 1921. BOTH regressions produce R-squared = 0.58, because the correlation between two fixed sets of numbers is fixed. The leftside graph depicts the first regression, and the rightside graph depicts the second. Because the correlation is not perfect, the nine boxy symbols do not all sit exactly upon the line of best fit. There is some scatter around the line.

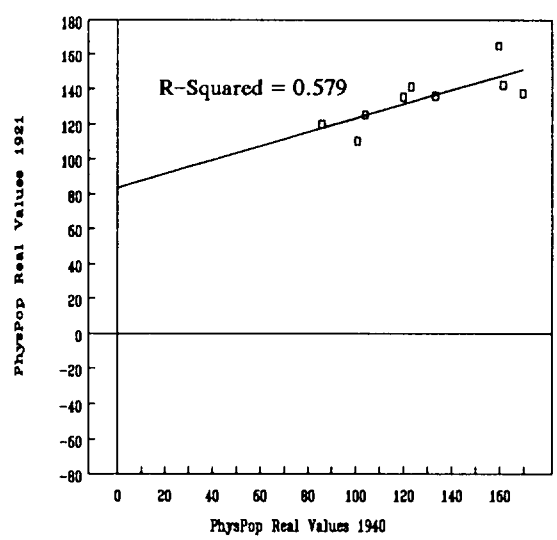
Census Division	1921 Real PhysPops "x"	1940 Real PhysPops "y"	Data Regression	Regression Output:
Pacific	165.11	159.72	Constant	-67.425
New England	142.24	161.55	Std Err of Y Est	20.641
West North Central	140.93	123.14	R Squared	0.579
Mid-Atlantic	137.29	169.76	No. of Observations	9
East North Central	136.06	133.36	Degrees of Freedom	7
Mountain	135.38	119.89		
West South Central	125.15	103.94	X Coefficient	1.4558
East South Central	119.76	85.83	Std Err of Coef.	0.4688
South Atlantic	110.32	100.74	X-Coeff. / S.E. =	3.1050

Census Division	1940 Real PhysPops "x"	1921 Real PhysPops "y"	Data Regression	Regression Output:
Pacific	159.72	165.11	Constant	83.491
New England	161.55	142.24	Std Err of Y Est	10.792
West North Central	123.14	140.93	R Squared	0.579
Mid-Atlantic	169.76	137.29	No. of Observations	9
East North Central	133.36	136.06	Degrees of Freedom	7
Mountain	119.89	135.38		
West South Central	103.94	125.15	X Coefficient	0.3980
East South Central	85.83	119.76	Std Err of Coef.	0.1282
South Atlantic	100.74	110.32	X-Coeff. / S.E. =	3.1050

Regression of PhysPop 1940 (Y-axis) upon PhysPop 1921 (X-axis)



Regression of PhysPop 1921 (Y-axis) upon PhysPop 1940 (X-axis)



Related text = Part 7b.

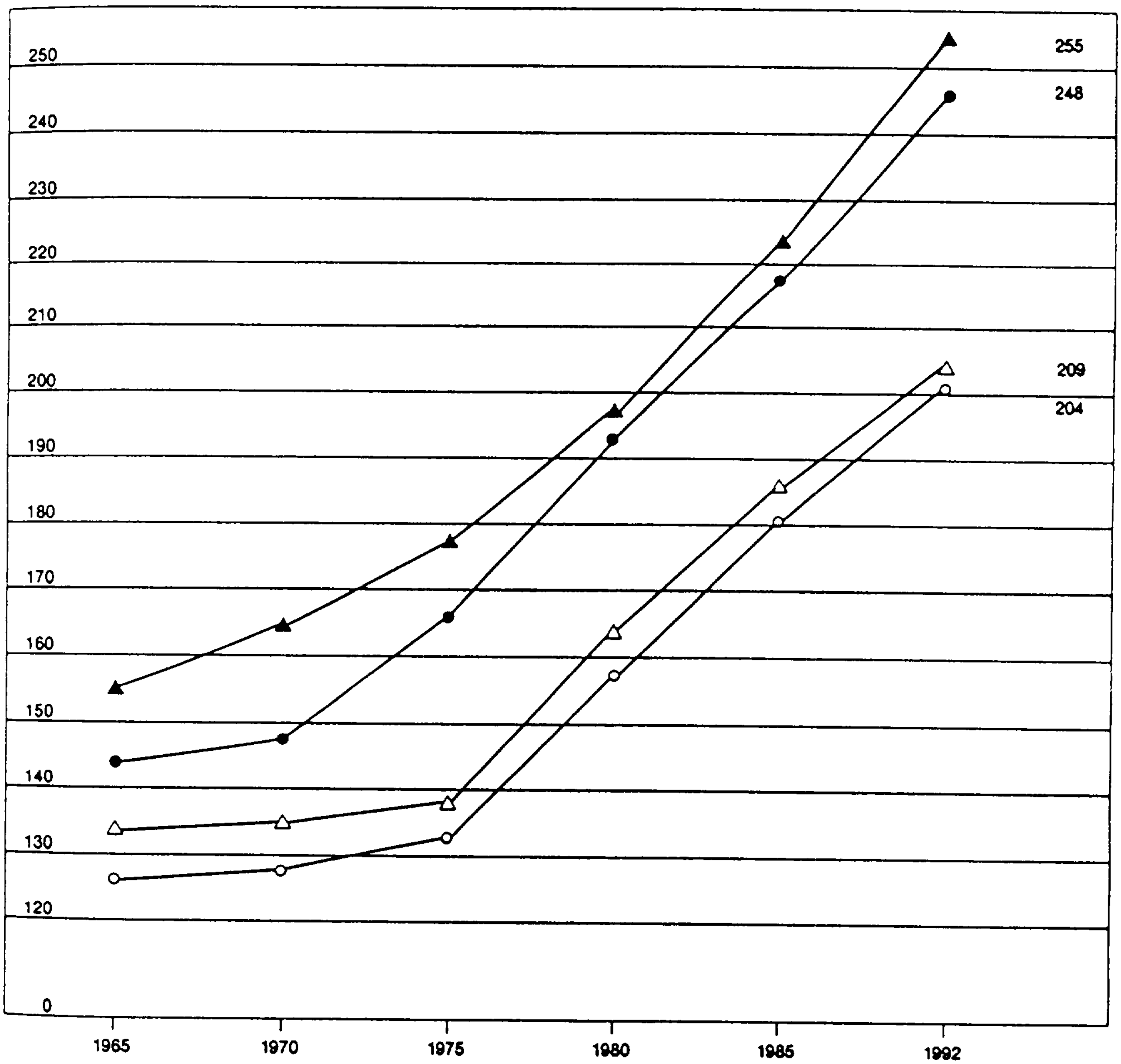
Figure 3-D.
Comparison of Four Types of PhysPop Values.

This figure is reproduced from p.15 of AMA 1993 in our Reference List: Physician Characteristics and Distribution in the U.S., 1993 Edition, by Roback + Randolph + Seidman of the American Medical Association, Department of Physician Data Services.

Related text = Part 3c.

Figure 4

Trends in Physician/Population Ratios for Selected Years 1965–1992



▲ Total Phys. per 100,000 Total Pop. in *U.S. & Possessions*
△ Total Patient Care Phys. per 100,000 Total Pop. in *U.S. & Possessions*
● Non-Federal Phys. per 100,000 Civilian Pop. in *U.S. only*
○ Non-Federal Patient Care Phys. per 100,000 Civ. Pop. in *U.S. only*

Sources: Table A-16 and Table A-17.

Universal PhysPop Table

First page of four

Related text = Parts 3a + 3b

- PhysPop values are numbers of physicians per 100,000 population. Entries are for general practitioners and specialists combined --- 1921 through 1993 (details in text). Sources of the data are provided in the text, Part 3d. The years which are flagged with a "+" sign present prime data. Entries for the unflagged years have been interpolated.

- The particular states belonging to each Census Division are listed in the text, Part 3b. PhysPop entries for the Nine Census Divisions have been weighted by state populations, whereas the three rows of averages are non-weighted. High-5 and Low-4 are defined in the text, Part 4.

- This single table is the source of data for numerous chapters of this book. The term "universal" in the table's title emphasizes that the PhysPops are the same, regardless of which cause of death is compared with them.

Census Division	1921+	1922	1923+	1924	1925+	1926	1927+	1928	1929+	1930
Pacific	165.11	164.09	163.06	162.36	161.67	159.75	157.83	157.24	156.64	158.30
New England	142.24	139.82	137.39	137.85	138.31	137.91	137.50	137.98	138.46	140.40
West North Central	140.93	139.62	138.31	136.11	133.92	132.73	131.54	130.13	128.72	127.61
Mid-Atlantic	137.29	138.11	138.92	136.64	134.36	136.38	138.40	138.45	138.49	139.65
East North Central	136.06	133.94	131.82	129.68	127.54	126.86	126.18	126.35	126.51	127.55
Mountain	135.38	132.95	130.51	126.40	122.30	120.52	118.75	118.72	118.68	118.79
West South Central	125.15	122.16	119.16	116.00	112.83	110.54	108.25	106.92	105.60	105.77
East South Central	119.76	116.46	113.16	110.19	107.22	104.64	102.07	100.74	99.41	98.07
South-Atlantic	110.32	108.56	106.79	105.20	103.61	102.87	102.13	101.50	100.86	100.23
Average ALL	134.70	132.85	131.01	128.94	126.86	125.80	124.74	124.22	123.71	124.04
Average High-Five	144.33	143.11	141.90	140.53	139.16	138.73	138.29	138.03	137.76	138.70
Average Low-Four	122.65	120.03	117.41	114.45	111.49	109.64	107.80	106.97	106.14	105.72
Ratio (High/Low)	1.18	1.19	1.21	1.23	1.25	1.27	1.28	1.29	1.30	1.31

Census Division	1931+	1932	1933	1934+	1935	1936+	1937	1938+	1939	1940+
Pacific	159.97	160.01	160.05	160.09	159.26	158.44	158.03	157.62	158.64	159.72
New England	142.35	144.43	146.51	148.60	149.39	150.18	152.13	154.08	157.82	161.55
West North Central	126.50	126.32	126.14	125.96	126.05	126.14	125.54	124.95	124.06	123.14
Mid-Atlantic	140.82	143.75	146.69	149.62	152.33	155.05	157.87	160.69	165.19	169.76
East North Central	128.59	128.84	129.10	129.36	129.89	130.42	131.20	131.98	132.66	133.36
Mountain	118.89	118.32	117.74	117.16	118.48	119.80	119.84	119.88	119.95	119.89
West South Central	105.95	105.53	105.11	104.68	104.10	103.52	103.15	102.79	103.37	103.94
East South Central	96.73	95.15	93.58	92.00	90.97	89.94	89.07	88.21	87.03	85.83
South-Atlantic	99.59	99.20	98.80	98.41	98.78	99.16	99.21	99.26	100.06	100.74
Average ALL	124.38	124.62	124.86	125.10	125.47	125.85	126.23	126.61	127.64	128.66
Average High-Five	139.65	140.67	141.70	142.72	143.38	144.04	144.96	145.87	147.68	149.51
Average Low-Four	105.29	104.55	103.81	103.06	103.08	103.10	102.82	102.53	102.60	102.60
Ratio (High/Low)	1.33	1.35	1.37	1.38	1.39	1.40	1.41	1.42	1.44	1.46

Universal PhysPop Table

Second page of four

- PhysPop values are numbers of physicians per 100,000 population. Entries are for general practitioners and specialists combined --- 1921 through 1993 (details in text). Sources of the data are provided in the text, Part 3d. The years which are flagged with a "+" sign present prime data. Entries for the unflagged years have been interpolated.

- The particular states belonging to each Census Division are listed in the text, Part 3b. PhysPop entries for the Nine Census Divisions have been weighted by state populations, whereas the three rows of averages are non-weighted. High-5 and Low-4 are defined in the text, Part 4.

- This single table is the source of data for numerous chapters of this book. The term "universal" in the table's title emphasizes that the PhysPops are the same, regardless of which cause of death is compared with them.

Census Division	1941	1942+	1943	1944	1945	1946	1947	1948	1949+	1950
Pacific	152.84	145.95	146.22	146.48	146.74	147.00	147.27	147.53	147.79	148.60
New England	162.77	163.99	163.77	163.55	163.33	163.11	162.88	162.66	162.44	162.51
West North Central	125.09	127.05	126.21	125.38	124.54	123.76	122.87	122.04	121.20	120.06
Mid-Atlantic	172.19	174.63	173.93	173.23	172.53	171.83	171.13	170.43	169.73	168.71
East North Central	134.12	134.89	133.48	132.06	130.65	129.24	127.82	126.41	125.00	123.69
Mountain	118.18	116.46	117.01	117.55	118.09	118.64	119.18	119.72	120.27	119.38
West South Central	104.41	104.88	104.42	103.31	103.52	103.06	102.61	102.16	101.40	101.34
East South Central	86.16	86.49	85.94	85.39	84.84	84.29	83.74	83.19	82.64	83.05
South-Atlantic	101.71	102.68	102.10	101.53	100.95	100.38	99.80	99.22	98.65	99.07
Average ALL	128.61	128.56	128.12	127.61	127.24	126.81	126.37	125.93	125.46	125.16
Average High-Five	149.40	149.30	148.72	148.14	147.56	146.99	146.39	145.81	145.23	144.71
Average Low-Four	102.62	102.63	102.37	101.95	101.85	101.59	101.33	101.07	100.74	100.71
Ratio (High/Low)	1.46	1.45	1.45	1.45	1.45	1.45	1.44	1.44	1.44	1.44

Census Division	1951	1952	1953	1954	1955	1956	1957	1958	1959+	1960
Pacific	149.40	150.21	151.01	151.82	152.62	153.43	154.23	155.04	155.84	158.74
New England	162.59	162.66	162.74	162.81	162.88	162.96	163.03	163.11	163.18	164.37
West North Central	118.09	117.77	116.62	115.48	114.34	113.19	112.05	110.90	109.76	111.25
Mid-Atlantic	167.68	166.66	165.81	164.62	163.59	162.57	161.55	160.52	159.50	162.65
East North Central	122.37	121.06	119.75	118.44	117.12	115.81	114.50	113.18	111.87	114.56
Mountain	118.51	117.64	116.76	115.88	115.00	114.12	113.25	112.37	111.49	112.93
West South Central	101.28	101.21	101.15	101.09	101.03	100.97	100.90	100.84	100.78	101.65
East South Central	83.46	83.86	84.27	84.68	85.09	85.50	85.90	86.31	86.72	88.00
South-Atlantic	99.49	99.91	100.33	100.75	101.16	101.58	102.00	102.42	102.84	105.36
Average ALL	124.76	124.55	124.27	123.95	123.65	123.35	123.05	122.74	122.44	124.39
Average High-Five	144.03	143.67	143.19	142.63	142.11	141.59	141.07	140.55	140.03	142.31
Average Low-Four	100.69	100.66	100.63	100.60	100.57	100.54	100.51	100.49	100.46	101.99
Ratio (High/Low)	1.43	1.43	1.42	1.42	1.41	1.41	1.40	1.40	1.39	1.40

Universal PhysPop Table

Third page of four

- PhysPop values are numbers of physicians per 100,000 population. Entries are for general practitioners and specialists combined --- 1921 through 1993 (details in text). Sources of the data are provided in the text, Part 3d. The years which are flagged with a "+" sign present prime data. Entries for the unflagged years have been interpolated.

- The particular states belonging to each Census Division are listed in the text, Part 3b. PhysPop entries for the Nine Census Divisions have been weighted by state populations, whereas the three rows of averages are non-weighted. High-5 and Low-4 are defined in the text, Part 4.

- This single table is the source of data for numerous chapters of this book. The term "universal" in the table's title emphasizes that the PhysPops are the same, regardless of which cause of death is compared with them.

Census Division	1961	1962	1963+	1964	1965+	1966	1967	1968	1969	1970+
Pacific	161.64	164.55	167.45	167.54	167.62	170.86	174.10	177.35	180.59	183.83
New England	165.56	166.75	167.94	170.52	173.09	175.77	178.46	181.14	183.83	186.51
West North Central	112.74	114.24	115.73	118.25	120.76	121.36	121.96	122.57	123.17	123.77
Mid-Atlantic	165.80	168.94	172.09	175.22	178.34	181.07	183.80	186.54	189.27	192.00
East North Central	117.25	119.94	122.63	123.16	123.69	124.39	125.08	125.78	126.47	127.17
Mountain	114.37	115.81	117.25	117.26	117.26	121.26	125.26	129.27	133.27	137.27
West South Central	102.52	103.38	104.25	104.28	104.31	106.09	107.87	109.64	111.42	113.20
East South Central	89.28	90.57	91.85	92.98	94.11	95.47	96.82	98.18	99.53	100.89
South-Atlantic	107.88	110.39	112.91	115.41	117.91	120.47	123.03	125.58	128.14	130.70
Average ALL	126.34	128.29	130.23	131.62	133.01	135.19	137.38	139.56	141.74	143.93
Average High-Five	144.60	146.88	149.17	150.93	152.70	154.69	156.68	158.67	160.66	162.66
Average Low-Four	103.51	105.04	106.57	107.48	108.40	110.82	113.24	115.67	118.09	120.52
Ratio (High/Low)	1.40	1.40	1.40	1.40	1.41	1.40	1.38	1.37	1.36	1.35

Census Division	1971	1972	1973	1974	1975+	1976	1977	1978	1979	1980+
Pacific	188.70	193.57	198.45	203.32	208.19	213.72	219.25	224.78	230.31	235.84
New England	192.25	197.99	203.72	209.46	215.20	223.03	230.87	238.70	246.54	254.37
West North Central	127.20	130.63	134.06	137.49	140.92	145.91	150.90	155.88	160.87	165.86
Mid-Atlantic	196.24	200.47	204.71	208.94	213.18	218.03	222.87	227.72	232.56	237.41
East North Central	130.94	134.72	138.49	142.27	146.04	150.79	155.54	160.29	165.04	169.79
Mountain	141.06	144.85	148.65	152.44	156.23	160.54	164.84	169.15	173.45	177.76
West South Central	116.19	119.18	122.18	125.17	128.16	133.16	138.17	143.17	148.18	153.18
East South Central	104.18	107.47	110.75	114.04	117.33	121.77	126.20	130.64	135.07	139.51
South-Atlantic	135.78	140.86	145.94	151.02	156.10	162.32	168.55	174.77	181.00	187.22
Average ALL	148.06	152.19	156.33	160.46	164.59	169.92	175.24	180.57	185.89	191.22
Average High-Five	167.07	171.48	175.89	180.30	184.71	190.30	195.89	201.47	207.06	212.65
Average Low-Four	124.30	128.09	131.88	135.67	139.45	144.45	149.44	154.43	159.42	164.42
Ratio (High/Low)	1.34	1.34	1.33	1.33	1.32	1.32	1.31	1.30	1.30	1.29

Universal PhysPop Table

Fourth page of four

- PhysPop values are numbers of physicians per 100,000 population. Entries are for general practitioners and specialists combined --- 1921 through 1993 (details in text). Sources of the data are provided in the text, Part 3d. The years which are flagged with a "+" sign present prime data. Entries for the unflagged years have been interpolated.

- The particular states belonging to each Census Division are listed in the text, Part 3b. PhysPop entries for the Nine Census Divisions have been weighted by state populations, whereas the three rows of averages are non-weighted. High-5 and Low-4 are defined in the text, Part 4.

- This single table is the source of data for numerous chapters of this book. The term "universal" in the table's title emphasizes that the PhysPops are the same, regardless of which cause of death is compared with them.

Census Division	1981+	1982	1983+	1984	1985+	1986	1987	1988	1989	1990+
Pacific	241.07	245.83	250.59	253.18	255.78	257.64	259.50	261.37	263.23	265.09
New England	261.79	270.07	278.35	285.44	292.52	298.00	303.47	308.94	314.41	319.88
West North Central	170.49	175.13	179.76	183.06	186.36	189.65	192.93	196.21	199.50	202.78
Mid-Atlantic	245.75	255.00	264.24	270.03	275.83	280.22	284.61	289.01	293.40	297.79
East North Central	174.96	180.94	186.91	190.82	194.72	197.49	200.25	203.01	205.78	208.54
Mountain	182.02	184.91	187.80	190.17	192.53	195.67	198.80	201.93	205.07	208.20
West South Central	156.72	160.32	163.92	167.48	171.04	173.70	176.36	179.02	181.68	184.34
East South Central	144.39	148.87	153.34	157.67	162.00	166.09	170.17	174.25	178.34	182.42
South-Atlantic	191.23	197.83	204.43	210.15	215.86	219.59	223.31	227.03	230.76	234.48
Average ALL	196.49	202.10	207.70	212.00	216.30	219.78	223.27	226.75	230.24	233.72
Average High-Five	218.81	225.39	231.97	236.51	241.04	244.60	248.15	251.71	255.26	258.82
Average Low-Four	168.59	172.98	177.37	181.37	185.36	188.76	192.16	195.56	198.96	202.36
Ratio (High/Low)	1.30	1.30	1.31	1.30	1.30	1.30	1.29	1.29	1.28	1.28

Census Division	1991	1992+	1993+ *
Pacific	266.57	268.05	269.50
New England	327.11	334.35	343.80
West North Central	209.48	216.17	219.00
Mid-Atlantic	307.67	317.56	323.60
East North Central	215.02	221.50	225.40
Mountain	211.23	214.26	218.30
West South Central	189.43	194.53	195.40
East South Central	188.38	194.33	196.70
South-Atlantic	239.45	244.41	247.80
Average ALL	239.37	245.02	248.83
Average High-Five	265.17	271.53	276.26
Average Low-Four	207.12	211.88	214.55
Ratio (High/Low)	1.28	1.28	1.29

* 1993 entries are for January 1, 1993, from Roback 1994.

Table 3-B

Population Sizes of the Census Divisions: 1910 through 1990

Census Division	1910 Pop	1910 Fraction	1920 Pop	1920 Fraction	1930 Pop	1930 Fraction
Pacific	4,192,304	0.0457	5,566,851	0.0529	8,194,433	0.0670
New England	6,652,675	0.0725	7,400,909	0.0703	8,268,680	0.0676
West North Central	11,637,921	0.1269	12,544,249	0.1192	13,296,915	0.1086
MidAtlantic	19,315,892	0.2105	22,261,144	0.2115	26,260,750	0.2146
East North Central	18,250,621	0.1989	21,475,543	0.2040	25,297,185	0.2067
Mountain	2,633,517	0.0287	3,336,101	0.0317	3,702,789	0.0303
West South Central	8,784,534	0.0958	10,242,224	0.0973	12,176,830	0.0995
East South Central	8,409,901	0.0917	8,893,307	0.0845	9,887,214	0.0808
South Atlantic	11,864,826	0.1293	13,552,701	0.1287	15,306,720	0.1251
	91,742,191	1.0000	105,273,029	1.0000	122,391,516	1.0000

Census Division	1940 Pop	1940 Fraction	1950 Pop	1950 Fraction	1960 Pop	1960 Fraction
Pacific	9,733,262	0.0739	14,486,527	0.0961	21,198,044	0.1182
New England	8,437,290	0.0641	9,314,453	0.0618	10,509,367	0.0586
West North Central	13,516,990	0.1027	14,061,394	0.0933	15,394,115	0.0858
MidAtlantic	27,539,487	0.2092	30,163,533	0.2002	34,168,452	0.1905
East North Central	26,626,342	0.2022	30,399,368	0.2017	36,225,024	0.2020
Mountain	4,150,003	0.0315	5,074,998	0.0337	6,855,060	0.0382
West South Central	13,064,525	0.0992	14,537,572	0.0965	16,951,255	0.0945
East South Central	10,778,225	0.0819	11,477,181	0.0762	12,050,126	0.0672
South Atlantic	17,823,151	0.1354	21,182,335	0.1406	25,971,732	0.1448
	131,669,275	1.0000	150,697,361	1.0000	179,323,175	1.0000

Census Division	1970 Pop	1970 Fraction	1980 Pop	1980 Fraction	1990 Pop	1990 Fraction
Pacific	26,087,000	0.1293	31,523,000	0.1398	37,837,000	0.1535
New England	11,781,000	0.0584	12,322,000	0.0546	12,998,000	0.0527
West North Central	16,240,000	0.0805	17,124,000	0.0759	17,777,000	0.0721
MidAtlantic	37,149,000	0.1842	36,770,000	0.1630	37,660,000	0.1527
East North Central	40,212,000	0.1993	41,636,000	0.1846	42,232,000	0.1713
Mountain	8,230,000	0.0408	11,319,000	0.0502	13,398,000	0.0543
West South Central	19,132,000	0.0948	23,669,000	0.1049	26,797,000	0.1087
East South Central	12,723,000	0.0631	14,573,000	0.0646	15,313,000	0.0621
South Atlantic	30,169,000	0.1496	36,621,000	0.1624	42,540,000	0.1725
	201,723,000	1.0000	225,557,000	1.0000	246,552,000	1.0000

Some sources provided entries to the last digit, but no one should take seriously any such implied accuracy of census-taking. Sources: For 1910, 1920, 1930: World Almanac 1991, p.553. For 1940, 1950, 1960: Grove 1968, Table 74. For 1970, 1980: Roback 1990. For 1990: Roback 1994. Entries above exclude no one by color or "race."

Related text = Part 3b.

Table 3-C. How Sets of Physpops Correlate through Time.

The 9 Census Div	1993 Phys/ Pop	1992 Phys/ Pop	1990 Phys/ Pop	1985 Phys/ Pop	1980 Phys/ Pop	1975 Phys/ Pop	1967 Phys/ Pop	1965 Phys/ Pop	1963 Phys/ Pop	1949 Phys/ Pop	1942 Phys/ Pop	1940 Phys/ Pop	1938 Phys/ Pop	1936 Phys/ Pop	1934 Phys/ Pop	1931 Phys/ Pop	1929 Phys/ Pop	1927 Phys/ Pop	1925 Phys/ Pop	1923 Phys/ Pop	1921 Phys/ Pop
New England	343.8	334.3	319.9	292.5	254.2	215.1	174.5	168.6	167.1	162.4	164.0	161.6	154.1	150.2	148.6	142.3	138.5	137.5	138.3	137.4	142.2
Middle Atlantic	323.6	317.6	297.8	275.8	237.3	213.1	178.2	173.1	168.7	169.7	174.6	169.8	160.7	155.1	149.6	140.8	138.5	138.4	134.5	138.9	137.3
East North Centr	225.4	221.5	208.6	194.7	169.8	145.9	124.5	121.3	118.2	125.0	134.9	133.4	132.0	130.4	129.4	128.6	126.5	126.2	127.5	131.8	136.1
West North Centr	219.0	216.2	202.8	186.4	165.8	140.8	119.1	116.2	114.0	121.2	127.1	123.1	125.0	126.1	126.0	126.5	128.7	131.5	133.9	138.3	140.9
South Atlantic	247.8	244.4	234.5	215.9	187.0	156.0	122.7	118.4	113.0	98.7	102.7	100.7	99.3	99.2	98.4	99.6	100.9	102.1	103.6	106.8	110.3
East South Centr	196.7	194.3	182.4	162.0	139.7	117.4	93.4	90.5	89.2	83.2	86.5	85.8	88.2	89.9	92.0	96.7	99.4	102.1	107.2	113.2	119.8
West South Centr	195.4	194.5	184.3	171.0	153.3	128.0	106.2	103.4	102.5	102.2	104.9	103.9	102.8	103.5	104.7	106.0	105.6	108.2	112.8	119.2	125.2
Mountain	218.3	214.3	208.2	192.5	177.5	155.9	125.1	121.0	117.8	119.7	116.1	119.9	119.9	119.8	117.2	118.9	118.7	118.7	122.3	130.5	135.4
Pacific	269.5	268.0	265.1	255.8	236.2	208.1	167.3	161.4	159.6	147.5	146.0	159.7	157.6	158.4	160.1	160.0	156.6	157.8	161.7	163.1	165.1

Correlation of Each Phys/Pop with All Other Phys/Pops (Measured in R-Squared).

Year	1993	1992	1990	1985	1980	1975	1967	1965	1963	1949	1942	1940	1938	1936	1934	1931	1929	1927	1925	1923	1921
Phys/Pop 21	0.15	0.16	0.20	0.25	0.34	0.38	0.41	0.41	0.44	0.48	0.43	0.58	0.65	0.72	0.77	0.87	0.88	0.90	0.96	0.98	1.00
Phys/Pop 23	0.20	0.21	0.25	0.31	0.40	0.45	0.49	0.49	0.52	0.56	0.51	0.67	0.78	0.83	0.84	0.91	0.94	0.95	0.98	1.00	
Phys/Pop 25	0.28	0.29	0.33	0.40	0.49	0.53	0.56	0.56	0.59	0.61	0.57	0.71	0.77	0.83	0.88	0.95	0.97	0.98	1.00		
Phys/Pop 27	0.38	0.39	0.43	0.49	0.58	0.62	0.67	0.67	0.69	0.72	0.69	0.81	0.87	0.92	0.95	0.98	0.99	1.00			
Phys/Pop 29	0.42	0.43	0.47	0.53	0.61	0.66	0.71	0.71	0.73	0.76	0.73	0.85	0.90	0.94	0.97	0.99	1.00				
Phys/Pop 31	0.45	0.46	0.50	0.57	0.65	0.69	0.74	0.74	0.76	0.79	0.76	0.88	0.92	0.96	0.98	1.00					
Phys/Pop 34	0.56	0.57	0.60	0.66	0.72	0.77	0.82	0.83	0.85	0.89	0.87	0.95	0.98	0.99	1.00						
Phys/Pop 36	0.60	0.61	0.63	0.69	0.74	0.79	0.85	0.86	0.87	0.93	0.91	0.98	0.99	1.00							
Phys/Pop 38	0.65	0.65	0.67	0.72	0.77	0.81	0.88	0.89	0.90	0.96	0.94	0.99	1.00								
Phys/Pop 40	0.71	0.72	0.73	0.78	0.81	0.85	0.91	0.92	0.93	0.98	0.96	1.00									
Phys/Pop 42	0.76	0.76	0.74	0.77	0.76	0.79	0.88	0.89	0.89	0.98	1.00										
Phys/Pop 49	0.77	0.78	0.78	0.81	0.82	0.85	0.92	0.93	0.93	1.00											
Phys/Pop 63	0.87	0.88	0.90	0.94	0.96	0.98	1.00	1.00	1.00												
Phys/Pop 65	0.87	0.88	0.91	0.94	0.96	0.98	1.00	1.00													
Phys/Pop 67	0.87	0.89	0.91	0.95	0.96	0.99	1.00														
Phys/Pop 75	0.88	0.89	0.93	0.96	0.99	1.00															
Phys/Pop 80	0.90	0.91	0.95	0.98	1.00																
Phys/Pop 85	0.96	0.97	0.99	1.00																	
Phys/Pop 90	0.99	0.99	1.00																		
Phys/Pop 92	0.99	1.00																			
Phys/Pop 93	1.00																				
Years	1993	1992	1990	1985	1980	1975	1967	1965	1963	1949	1942	1940	1938	1936	1934	1931	1929	1927	1925	1923	1921

Table 3-C: The 21 sets of Phys/Pop values for 1921-1993 match Table 3-A, with some exceptions (please see text of Part 7c). The rows of R-squared values come from regressing one PhysPop set upon another PhysPop set. The intersection of a column with a row reveals which two sets produced the R-squared value.

Related text = Part 7c.