NBER WORKING PAPER SERIES

THE ALLOCATION OF PUBLICLY-FUNDED BIOMEDICAL RESEARCH

Frank R. Lichtenberg

Working Paper 6601
http://www.nber.org/papers/w6601

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 1998

I am grateful to the American Enterprise Institute and to the Center for Economic Studies at the University of Munich for financial support. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

The Allocation of Publicly-Funded
Biomedical Research
Frank R. Lichtenberg
NBER Working Paper No. 6601
June 1998
JEL # H5, I1, O3

ABSTRACT

We develop a simple theoretical model of the allocation of public biomedical research expenditure, and present some empirical evidence about the determinants of this allocation. The structure of expenditure should depend upon the relative costs as well as the relative benefits of different kinds of research. Analysts of technical change typically have data on neither of these, but the measures of disease burden we use are indicative of the benefit of achieving advances against different diseases. We calculate distributions of government-funded biomedical research expenditure, by disease, from records of all research projects supported by the United States Public Health Service; to obtain a reasonably complete accounting of disease burden, we utilize data on both the dying (from the Vital Statistics-Mortality Detail file) and the living (from the National Health Interview Survey). We find a very strong positive relationship across diseases between total life-years lost before age 65 and public R&D expenditure. But the amount of publicly-funded research on a disease decreases with the share of life-years before age 65 lost to the disease that are lost by non-whites, perhaps because lack of scientific knowledge is a less important cause of premature mortality among non-whites than it is among whites. The number of research grants mentioning a chronic condition is completely uncorrelated with the number of people with the condition but very strongly positively related to the number of people whose activities are limited by that condition. There tends to be more research about chronic conditions that are prevalent among people living in low-income households, and that are prevalent among the young (under age 18) and the old (above age 75).

Frank R. Lichtenberg
Graduate School of Business
Columbia University
726 Uris Hall
3022 Broadway
New York, NY 10027
and NBER
frank.lichtenberg@columbia.edu

In the last century, the average health of the American people has improved dramatically. The mean life expectancy of Americans has increased almost 20 years, or two years per decade,[1] since the turn of the century. Just from 1979 to 1988, the age-adjusted mortality rate declined 7.2%.

An important part of this enormous progress in health (which is scarcely reflected in our national accounts) is probably due to large private and public investments in biomedical research. In 1993, health R&D accounted for 18% of total U.S. R&D expenditure. The contribution of at least some types of biomedical R&D to health progress may be difficult to establish. As some officials of the National Institutes of Health (NIH) have argued, much NIH-sponsored research is basic in nature and, although "scientific advances would not have been possible without continuing insight and understanding regarding the fundamental mechanisms of life and disease...basic research linkages to health care advances are complicated, long-term, and impossible to allocate clearly" (NIH (1993), p. 3).[2] But in a previous paper (Lichtenberg (1996)), I showed that reductions in mortality and hospitalization expenditure were significantly inversely related, across diseases, to the extent of pharmaceutical innovation, which results primarily from privately-funded biomedical research. It is not an exaggeration to say that biomedical research is a matter of life and death.

In this paper we develop a simple theoretical model of the allocation of public biomedical research expenditure, and present some empirical evidence about the determinants of this allocation. The implications of the theoretical model are consistent with government officials' descriptions of the allocation process: the structure of expenditure should depend upon research productivity (or "scientific opportunity") as well as on public health need, i.e. the societal and economic burden of the disease/condition. Although, we lack, at this point, useful indicators of research productivity (i.e., of the *cost* of achieving research advances), we have a number of measures of disease burden (i.e., of the *benefit* of achieving these advances). Analysts of technical change typically have data on neither the costs nor the benefits of technical advance. Failure to measure research productivity will not necessarily bias our estimates; if it does, it seems likely to bias them towards zero.

The paper is organized as follows. In the next section we develop the simple model of public research expenditure allocation.

We rely on three types and sources of data to estimate the parameters of the model: data on research activity derived from NIH's CRISP (Computerized Retrieval of Information on Scientific Projects) database, premature mortality data from the Vital Statistics-Mortality Detail

---

[1] "Buy ten, get two free," could be a fair, if crass, marketing slogan for U.S. health progress.

[2] NIH funded 31.5% of all public and private health R&D in 1993 and 81% of federal health R&D.

file, and data on chronic condition prevalence and severity from the National Health Interview Survey. These are discussed in Section 3. Preliminary estimates are presented in Section 4, and a summary is provided in Section 5.

2. A simple model of the determinants of research expenditure at the disease level

To motivate the discussion and develop a few intuitions, we will write down the simplest possible model of research funding allocation. This model is based on the following extremely strong assumptions (some of which are relaxed below): (1) there are only two diseases; (2) the number of people suffering from the two diseases, N_1 and N_2, is exogenous; (3) the average severity of the two diseases is identical; (4) the probability P_i of finding a cure for disease i (i = 1, 2) is a concave (deterministic) function of research funding for that disease, X_i: $P_i = X_i^{\alpha}$, where $0 < \alpha < 1$;[3] (5) the effect of funding on the probability of finding a cure is the same across diseases; and (6) the total research budget $X = X_1 + X_2$ is fixed.

Suppose that policymakers attempt to maximize the (expected) total number of people cured of both diseases subject to the budget constraint,[4] i.e. they choose X_1 to maximize

$$J^* = N_1 P_1 + N_2 P_2 = N_1 X_1^{\alpha} + N_2 X_2^{\alpha}$$
$$= N_1 X_1^{\alpha} + N_2 (X - X_1)^{\alpha} \qquad (1)$$

The first-order condition implies that relative funding of research on the two diseases should satisfy

$$\ln (X_1 / X_2) = [1 / (1 - \alpha)] \ln (N_1 / N_2) \qquad (2)$$

Research funding should increase with disease incidence: for example, $X_1 > X_2$ if $N_1 > N_2$. This is because the benefit of discovering a cure for the disease is proportional to its incidence, but the cost is independent of incidence. Moreover the elasticity of funding with respect to incidence

---

[3] Viscusi (1995, p.3) notes that "in the case of biomedical research, the typical outcome will be a change in societal risk levels induced by the biomedical research outcomes."

[4] We assume for simplicity that policymakers do not pay attention to *privately*-funded biomedical R&D, i.e. they are not merely trying to "fill gaps" in private research, nor do they consider the potential impact of public R&D on future private research activity. Toole (1997), however, presents evidence that suggests that public biomedical research may have a significant, albeit very delayed, impact on private drug discovery.

should exceed unity: if disease 1 is twice as prevalent as disease 2, research funding for disease 1 should be more than twice as great as research funding for disease 2.

One could generalize this model to the case of $I > 2$ diseases, to obtain $I - 1$ equilibrium conditions of the form

$$\ln X_i = \text{constant} + [1 / (1 - \alpha)] \ln N_i \qquad (3)$$

($i = 1,2,...,I-1$). Given cross-sectional or panel data on research funding and incidence by disease, one could estimate eq. (3) to test the hypothesis of diminishing returns to research funding at the disease level and to estimate the parameter α. But this simple model can and should be extended in at least two directions: we should allow for multiple indicators of incidence and for differences in research productivity (scientific opportunity) across diseases.

Multiple indicators of incidence. As the director of NIH says, a given disease imposes a number of different kinds of burden on society, and "policy makers will need to consider the relative importance or weight to be placed on each criteri[on] when assessing the overall societal burden imposed by each disease." Suppose that the overall burden of a disease is perceived by policymakers to be a function of K attributes of the disease: $N_i = f(A1_i, A2_i, \ldots, AK_i)$ where, for example, A1 is the number of deaths, A2 is the number of bed-disability days, A3 is the number of hospital stays, and so forth. Further suppose that the functional form of this relationship is

$$\ln N_i = \beta_1 \ln A1_i + \beta_2 \ln A2_i + \ldots + \beta_K \ln AK_i \qquad (4)$$

where $\Sigma_k \beta_k = 1$. β_k reveals the relative "weight" assigned by policymakers to attribute k in the determination of overall disease burden. Substituting eq. (4) into eq. (3),

$$\ln X_i = \text{constant} + [1 / (1 - \alpha)]\{\beta_1 \ln A1_i + \beta_2 \ln A2_i + \ldots + \beta_K \ln AK_i\} \qquad (5)$$

Estimation of eq. (5) would provide estimates of these ("revealed preference") weights as well as of the technological parameter α. They would indicate the relative weight given to mortality and bed-disability days, for example.

Since disease outcome and incidence data are available by demographic group, we can also make inferences about weights associated with different demographic groups.[5] For

---

[5] NIH officials acknowledge that "research funding decisions will also reflect concerns about equity among groups of potential beneficiaries of the research as defined in terms of age, sex, and ethnic origin. Certain criteria favor one group over another. For example, mortality rates and measures of the impact on

example, let us define "adjusted" bed-disability days A2* = A2YOUNG + (1 + θ) A2OLD, where A2YOUNG and A2OLD denote bed-disability days of young and old people, respectively. If policymakers' evaluation of the marginal burden of the two groups' bed-disability days differs, θ will differ from zero. This parameter can be estimated by replacing A2 by A2* in eq. (5).

Differences in research productivity (scientific opportunity) across diseases. The preceding model is based on the assumption that the effect of funding on the probability of finding a cure is the same across diseases. This assumption is clearly unrealistic, and it is desirable to relax it.[6] We can modify the cure-probability equation to include a disease-specific research productivity parameter π_i: $P_i = \pi_i X_i^{\alpha}$. The objective function policymakers seek to maximize is now $J^* = N_1 P_1 + N_2 P_2 = N_1 \pi_1 X_1^{\alpha} + N_2 \pi_2 X_2^{\alpha}$, and the optimal expenditure on research on disease i is now

$$\ln X_i = \text{constant} + [1 / (1 - \alpha)] \ln N_i + [1 / (1 - \alpha)] \ln \pi_i \qquad (6)$$

The research-productivity parameters ((i) enter the objective function and the optimal expenditure equation in the same way as the disease incidence measures (N_i). Research expenditure should be an increasing function of scientific opportunity as well as of disease burden. This implication is consistent with the views expressed by government officials: "It is vital that the allocation of medical research dollars takes into account several factors, including scientific opportunity, public health need, gaps in knowledge, as well as societal and economic burden of the disease/condition."[7]

We believe that the CRISP data can eventually be exploited to obtain indicators of (changes in) the relative productivity of research on different diseases. The data will enable us to determine, for example, the extent to which research related to a given disease tends to be concentrated in rapidly growing and advancing scientific fields (e.g. molecular genetics) as opposed to mature fields. They will also allow us to quantify the extent to which research on a

---

functioning may favor the elderly whereas measures of economic impact, such as lost productivity, would favor younger citizens." [NIH Director Varmus' responses to questions from Senator Slade Gordon, Labor, HHS, Education Subcommittee Heating, NIH appropriations for FY 1996, May 18, 1995]

[6] Henderson and Cockburn (1996) have studied the determinants of research productivity of pharmaceutical firms, using patents and scientific papers as measures of research output.

[7] OSP, NIH Response to Congressional Questions, June 1996. Garber and Romer (1995) also argue that "federal policy toward research and development should respond to scientific advances, technology trends, and changes in the political and social environment."

disease utilizes innovative research techniques (e.g. protein engineering), and how much the distribution of techniques has changed over time.

At present, however, we must treat π_i as unobservable. If research productivity is uncorrelated across diseases with disease burden, i.e. if differences in supply (or cost of achieving progress) are uncorrelated with differences in demand (or benefits of achieving progress), estimation of eq. (5) will yield an unbiased estimate of the relationship between research expenditure and disease burden. It is possible, however, that N and π are negatively correlated: the diseases that impose the heaviest burden do so, in part, because of the low productivity of past research on those diseases (which should also have resulted in relatively low research funding on them). If this is the case, then the omission of π_i from the research expenditure equation would bias the estimated coefficient on $\ln N_i$ towards zero. In particular, although the theory implies that the coefficient on $\ln N_i$ should be greater than one, we should not be surprised if we obtain estimated coefficients smaller than one, i.e. if we fail to observe this kind of 'increasing returns.'

In future research, we hope to directly estimate the contribution of medical research expenditure to subsequent progress against disease, by analyzing the correlation across diseases between research investment and indicators of progress, such as reductions in potential life years lost.[8] We recognize, however, that heterogeneous, unobserved research productivity is likely to lead to *overestimates* of the average return to research expenditure. Diseases receiving the greatest research funding are presumably those for which research productivity is highest. The slope of the relationship across diseases between research funding and progress exceeds the mean of the slopes of the disease-specific relationships.[9]

---

[8] The existing evidence on the contribution of medical research expenditure to subsequent progress against disease is rather limited. NIH has produced estimates of cost savings from 34 "examples" of health care advances resulting from NIH support for applied research and clinical trials. Most focus on a single innovation such as a new vaccine, a new diagnostic test, or a particular therapy. But these case studies are not necessarily a random sample of all NIH-sponsored research, so they may not reveal the "aggregate or average" effect of this research on costs. It is possible, for example, that the distribution of cost-savings is highly skewed to the right--a few programs confer large cost savings, but the majority confer few—and that the specific examples chosen tend to be concentrated in the upper tail of the distribution. Mushkin (1979) attempted to determine econometrically the contribution of biomedical research to reductions in mortality and morbidity. But most of her analysis was in an aggregate time-series framework, and was based on fairly crude measures of biomedical research, such as the number of biomedical PhD's lagged ten years.

[9] The reasoning underlying this is the same as that underlying Gary Chamberlin's argument that estimation of production functions using data for a cross-section of firms will result in overestimates of the returns to factors of production, e.g. labor. Firms with exogenously higher productivity (due, e.g., to greater managerial ability) will emply more workers.

3. Data sources and methods

3.1. Data on government-funded research funding, by disease

We have calculated distributions of government-funded biomedical research expenditure, by disease, from records of research grants contained in NIH's CRISP (Computerized Retrieval of Information on Scientific Projects) system. The CRISP database includes records of all research ventures supported by the United States Public Health Service since 1972. In fiscal year 1995, there were records of 63,289 grants whose total value was $10.1 billion. Most of this research falls within the broad category of extramural projects: grants, contracts, and cooperative agreements conducted primarily by investigators at universities, hospitals, and other research institutions. The projects are funded by NIH and the Substance Abuse and Mental Health Services Administration. A very small number of these research grants are funded by the Centers for Disease Control, the Food and Drug Administration (FDA), the Health Resources and Services Administration, and the Agency for Health Care Policy and Research. CRISP also contains information on intramural research programs conducted by scientists employed by the FDA and the various institutes of the NIH.

Each record reports the name of the investigator, the name and address of his or her organization (e.g. university and department), the title (and in many cases an abstract) of the project, the administering organization (e.g. National Cancer Institute), the award amount (including both direct and indirect costs), the type of award, and a number of (generally about 15) indexing terms assigned by Technical Information Specialists in the Research Documentation Section, Information Systems Branch, of NIH's Division of Research Grants. The indexing process is governed by the CRISP Thesaurus, which is the "controlled vocabulary used to assign indexing terms for the CRISP System, and to retrieve subject-related information from it."

The number of distinct indexing terms in the CRISP Thesaurus is quite large (about 9000), but most of these terms are organized into a small number of hierarchical classification schemes, including one for diseases. Table 1 illustrates the disease classification; it is similar to the International Classification of Diseases, the system used for reporting of diagnoses in most health-related data. There are 35 disease categories at the highest level of aggregation. Within each of these is a series of more specific disease categories. Space limitations prevent us from displaying the entire "tree structure" of diseases (which includes about 2900 items), but to illustrate the classification system we show the second level classification of "nervous disorders" and a branch leading to a "fifth level" disease (with no further subcategories), lymphocytic choriomeningitis.

This disease classification scheme enables us to compute distributions of research grants and dollars by disease, at various levels of aggregation.[10] How accurate will these distributions be? Recently the Office of the Director of NIH prepared a report that included estimates of NIH FY1994 research support by disease. These figures, based on data provided by NIH institutes, centers, and divisions (ICDs), "reflect NIH-wide resources devoted to research on the listed diseases...[and] generally do not correspond to budget figures for the ICD identifying the cost data."[11] For 16 randomly-selected diseases, we compared FY1994 funding as reported there with the number of FY1995 grants citing the disease contained in the FY1995 CRISP database.

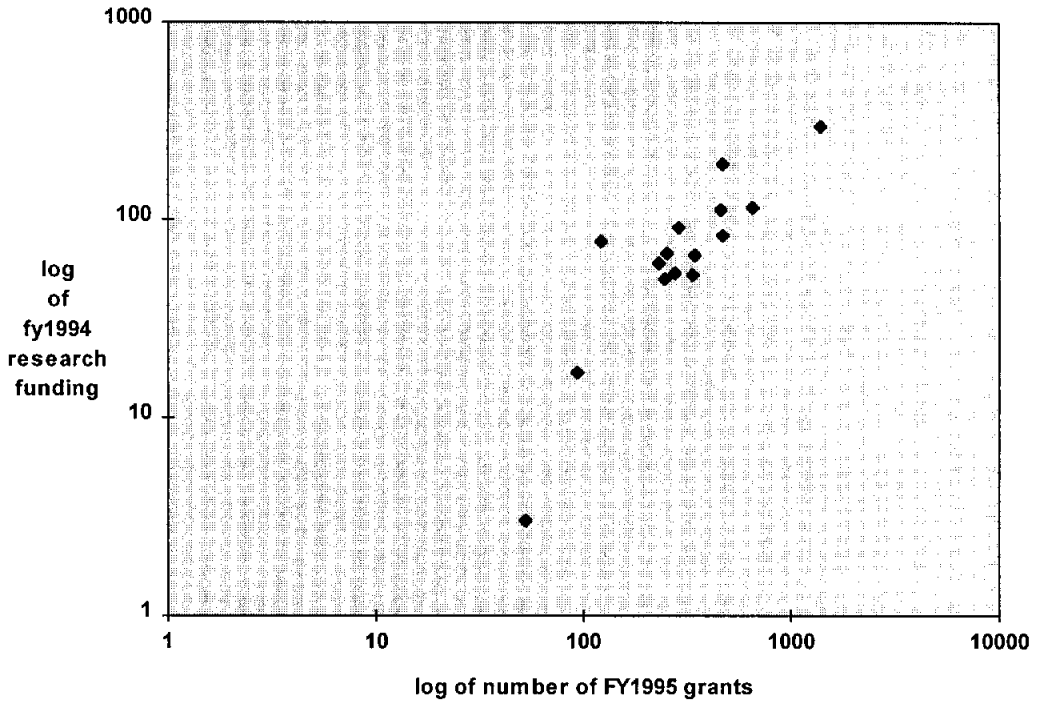
The raw data are reported below:

Disease/disorder	**FY1995 grants**	**FY1994 funds**
diabetes	1390	292
epilepsy	338	52
asthma	345	66
arthritis	476	191
atherosclerosis	650	116
schizophrenia	458	111
MS	123	78
obesity	474	83
osteoporosis	288	92
Parkinson's	253	68
psoriasis	53	3
sickle cell anemia	278	54
suicide	94	17
TB	248	50
pneumonia & influenza	230	60

[10] Data on the disease-distribution of *private* R&D sponsored by pharmaceutical firms are available from the Pharmaceutical Research and Manufacturers Association's Annual Survey of companies. Unfortunately, the private R&D data are disaggregated into only about seven broad categories. Figure 1 shows the percentage distributions of both private and government R&D, by these categories. Public R&D seems to be more concentrated on digestive/genitourinary and neoplasm/endocrine/metabolic diseases, and less concentrated on infective/parasitic, nervous system, and cardiovascular diseases than private R&D.

[11] "Disease-Specific Estimates of Direct and Indirect Costs of Illness and NIH Support," November, 1995, Table 1.

Relationship Between Estimated NIH Research Funding, by Disease, and Number of NIH Grants Citing Disease



A scatter plot of the logarithms of these two variables is shown above; their correlation coefficient is .91. Despite differences in timing and unit of measurement, the two estimates of relative research support by disease are quite similar, suggesting that the CRISP data provide accurate statistics.

As NIH officials observe, much NIH-sponsored research is basic in nature and, although "scientific advances would not have been possible without continuing insight and understanding regarding the fundamental mechanisms of life and disease...basic research linkages to health care advances are complicated, long-term, and impossible to allocate clearly" (NIH (1993), p. 3). Therefore, many research grants do not refer to *any* disease (even though the research may ultimately lead to breakthroughs in the treatment of that disease). In other words, the grants fall into two categories: those that have been assigned to at least one disease and those that have not been assigned.[12] Our estimates of research activity by disease are based only on grants that have been assigned.[13] Due to the logarithmic specification of eq. (6), the validity of our parameter

[12] This distinction resembles the distinction made in industrial R&D between basic and applied research.
[13] When two or more diseases are cited by a grant, we assign the *entire* amount of funding for the grant to *each* of the diseases cited.

estimates does not require us to reliably measure the *absolute level* of research funding, by disease; their validity is predicated only on reliable measurement of *relative* research funding, or activity. If one is willing to assume that the disease distribution of unassigned grants is similar to that of assigned grants, then our estimates could be regarded as applicable to all public biomedical research.

3.2. Data on Disease Burden, Prevalence, and Incidence

As indicated in eq. (4) above, rather than treating disease burden N as a scalar, we regard it as an index of a number of disease attributes. Data on these attributes are obtained from two sources: the Vital Statistics-Mortality Detail file, a virtually complete census of deaths in the United States, and the National Health Interview Survey, a continuing nationwide survey of households for which a probability sample of the civilian noninstitutionalized[14] population of the United States is interviewed by the U.S. Bureau of the Census regarding the health and other characteristics of each member of the household. (The sample for the years 1990–92 was composed of 142,638 households containing 368,075 persons.)

Our use of these two data sources reflects our belief that to obtain a reasonably complete accounting for disease burden, one must consider data on both the dying and the living. Analysis based on only one source will almost surely be subject to considerable sample selection bias.

3.2.1 Premature Mortality data

The measure of disease burden we computed from the mortality file is potential life-years lost before age 65, by disease.[15] The latter is defined as the summation of (65 - age-at-death) for decedents under 65. This is a standard measure of disease burden, or (lack of)

---

[14] It should be pointed out that the restriction of the NHIS to the civilian population not confined to institutions affects the estimated prevalence of chronic conditions. Omission of the institutionalized population reduces the prevalence estimates, especially for the elderly, because the proportion of persons in institutions who have chronic conditions is high. These estimates do not indicate the prevalence in the total population.

[15] Demographic information on the death certificate is provided by the funeral director based on information supplied by an informant. Medical certification of cause of death is provided by a physician, medical examiner, or coroner.

progress against disease, in health statistics. It has the drawback of giving no weight at all to deaths of people aged 65 and over.

3.2.2 Data on Prevalence of Selected Chronic Conditions

Collins (1997) presents statistics on the prevalence of selected chronic conditions in the U. S. during 1990–92 by age, sex, race, family income, and geographic region, derived from data collected in the National Health Interview Survey (NHIS). He also reports the percent of selected conditions that cause activity limitation, the percent for which a physician was consulted, and the percent that caused hospitalization.

All information collected during the survey is from responsible family members residing in the household. Methodological studies have shown that chronic conditions are generally underreported in interview surveys. Respondents in health interviews tend to report conditions of which they are aware and about which they are willing to report to the interviewer. Reporting is better for conditions that have made a significant impact on affected individuals and their families. Conditions that are severe or costly, or are being treated, tend to be better reported than conditions having less impact. Methodological studies have also indicated that inclusion of a checklist of descriptive condition titles as part of the interview questionnaire increases the probability that a respondent will recognize the terms and report those of which the respondent is aware.

The current procedure for collecting information on chronic conditions was established in 1978. Currently, six categorical lists of selected chronic conditions are included in the NHIS questionnaire: circulatory conditions; respiratory conditions; digestive conditions; impairments and conditions of the nervous system and sense organs; conditions of the skin and subcutaneous tissue and of the musculoskeletal system and connective tissue; and endocrine, nutritional, and metabolic diseases and immunity disorders, diseases of the blood and blood-forming organs, and conditions of the genitourinary system. Each family in the NHIS is questioned on only one of these six lists, selected on a predetermined basis. Therefore, each list is administered to only one-sixth of the total NHIS sample each year. For some items, responses are based on the following question: ''During the past 12 months did anyone in the family (read names) have . . .?'' For others, responses are based on the question ''Does anyone in the family (read names) now have . . .?'' For the rest, responses are based on the question ''Has anyone in the family (read names) ever had . . .?'' Estimates for days of disability caused by chronic conditions are based on the number of disability days reported for the 2 weeks before interview.

The survey includes data only on persons living in the household at the time of interview. Thus the experience of persons who died prior to the time of interview is excluded

from the data. Also excluded is the experience of persons who were institutionalized or who were members of the Armed Forces at the time of the household interview.

In these data, ''prevalence'' is defined as the average number of some item existing during a specified interval of time--usually referred to as ''period prevalence''—rather than the number of some item existing at a given point in time--usually referred to as ''point prevalence.'' Chronic conditions are defined as conditions that either were first noticed 3 months or more before the date of interviews, or belong to a group of conditions considered chronic regardless of when they began.

The data presented represent the prevalence of conditions, not the prevalence of persons with a chronic condition. However, for most conditions, the condition prevalence and the person prevalence are almost identical.[16]

4. Preliminary estimates

4.1 Premature mortality

The first measure of disease burden we analyze is potential life-years lost before age 65 (PLYL). Data on PLYL in 1980 and government research funding, in 1982, for 14 major disease categories, are shown, in descending PLYL order, in Table 2. Diseases of the circulatory system and neoplasms are, by far, the diseases with the largest tolls in terms of premature death. While the research funding for these two diseases is among the highest for all diseases, R&D funding for two other diseases with much smaller burdens exceed the funding for the first two diseases, in one case by a large amount. Nevertheless, as the scatter plot in Figure 2 and the following regression indicate, there is a very strong positive relationship across the entire sample between life-years lost and public R&D expenditure (t-statistics in parentheses):

$$\ln(RD82) = \underset{(0.34)}{-0.464} + \underset{(3.19)}{0.355}\ln(LYL80) + e \qquad R^2=.459,\ N = 14$$

[16] There are some instances in which large variations are present; these occur for two different reasons. The first is that a prevalence estimate of a condition may include more than one of the specified checklist items or a checklist item and a specified ''other condition'' item that falls into the same ICD category as the checklist item. The second reason is that some prevalence categories shown are a combination of other categories and, as a result, a person may have more than one of the conditions that are added to form the combined category. The concept of condition prevalence is generally used in NHIS because specific health indexes such as limitation of activity and disability days can be ascribed to specific conditions. In addition, prosthetic and pharmaceutical treatment modes are more condition specific than person specific.

Life-years lost in 1980 explains almost half of the variation across diseases in 1982 research expenditure. However, contrary to the implication of our simple theoretical model of research allocation, the coefficient on ln(LYL80) is significantly less than one. As argued above, this may be due to a negative correlation between the regressor and the omitted research-productivity variable.

Life-years lost can be classified by sex, race, educational attainment, and other characteristics, so we can investigate whether premature mortality among certain demographic groups tends to be associated with especially high government research funding. Sixty percent of life years lost before age 65 are lost by males, and 25% are lost by non-whites (who comprise about 10% of the population), reflecting the lower life expectancy of these two groups. The proportion of life-years lost by men and by non-whites varies considerably across diseases. Whites account for 81% of life-years lost to neoplasms but for only 53% of those due to diseases of the blood and blood-forming organs. Men account for 81% of life-years lost to infectious and parasitic diseases but for only 28% of life-years lost to musculoskeletal and connective-tissue diseases.

The matrix of correlation coefficients for four variables-- ln(RD82), ln(LYL80), and the fractions of life-years lost to men (%MALE) and to whites (%WHITE)—are reported below (p-values are shown below the correlation coefficients).

	ln(RD82)	ln(LYL80)	%MALE
ln(LYL80)	0.67714		
	0.0078		
%MALE	0.55375	0.56093	
	0.0399	0.0369	
%WHITE	0.75643	0.88477	0.50235
	0.0017	0.0001	0.0672

Public R&D investment is significantly positively correlated with the fractions of life-years lost to men and (especially) to whites, as well as with the total number of life-years lost. Indeed, %WHITE is more strongly correlated with R&D than total life-years lost is. (A scatter plot of ln(RD82) against %WHITE is shown in Figure 3.) But as the second column of coefficients reveals, both %MALE and %WHITE are significantly positively correlated with total life-years lost: the diseases associated with the greatest number of premature deaths are those for which men and whites account for the greatest fractions of life-years lost. We therefore need to determine whether %MALE and %WHITE have significant effects on public R&D, controlling

for total life-years lost (although our ability to determine this will be hampered by multicollinearity). The appropriate regressions are:

$$\ln(RD82) = \underset{(0.12)}{-0.164} + \underset{(2.08)}{0.280}\ \ln(LYL80) + \underset{(0.99)}{1.11}\ \%MALE + e$$

R^2=.503 N = 14

$$\ln(RD82) = \underset{(0.64)}{-0.813} + \underset{(0.09)}{0.019}\ \ln(LYL80) + \underset{(1.71)}{6.56}\ \%WHITE + e$$

R^2=.573 N = 14

The coefficient on %MALE is insignificant and the inclusion of this variable only slightly reduces the coefficient on ln(LYL80). In contrast, the coefficient on %WHITE is marginally significant, even in the presence of the other regressor, which becomes completely insignificant when %WHIITE is included. We also estimated an alternative functional form of the relationship RD82=f(LYL80, %WHITE):

$$\ln(RD82) = \underset{(1.14)}{2.29} + \underset{(1.90)}{1.35}\ \ln(LYL80 * \%WHITE) - \underset{(1.44)}{1.30}\ \ln(LYL80*(1 - \%WHITE)) + e$$

R^2=.560 N = 14

These estimates indicate that research expenditure is positively correlated with life-years lost by whites but not with life-years lost by non-whites; the coefficient on the latter is negative, but its p-value is only .18. The two coefficients are virtually equal in magnitude and opposite in sign; if one imposes that restriction (which is not nearly rejected by the data), the estimates are:

$$\ln(RD82) = \underset{(8.40)}{2.72} + \underset{(3.89)}{1.47}\ \ln(\%WHITE / (1 - \%WHITE)) + e$$

R^2=.558 N = 14

The data are highly consistent with the hypothesis that the amount of publicly-funded research on a disease decreases with the share of life-years before age 65 lost to the disease that are lost by non-whites. A possible explanation for this finding is that lack of scientific knowledge is a less important cause of premature mortality among non-whites than it is among whites. Non-white premature mortality may be due, to a greater extent, to poor diet, reduced utilization of medical care, or other factors. In other words, it is plausible that the health status of non-whites tends to be well below the frontier of medical knowledge, whereas the health status

of whites tends to be on, or closer to, the frontier. The purpose of biomedical research is to shift the frontier outward, and the allocation or "direction" of research should depend (more) on the distribution of the disease burden of those on, or close to, the frontier. If cures for diseases that impose a heavy toll on minorities have already been found, then the productivity of further research on those diseases may be quite low.

The relative lack of research on diseases borne disproportionately by minorities may also be due to other reasons and may not be efficient. It may reflect the relatively low representation of minorities among the ranks of biomedical scientists. The National Science Foundation monitors the participation of women and minorities in science and engineering and has adopted some policies to increase their participation.

4.2 Prevalence and severity of chronic conditions in the (living) population

Table 3 presents data on the number of FY1995 research grants mentioning chronic conditions surveyed in the National Health Interview Survey and the number of people having, and limited in activity by, these conditions.[17] The condition mentioned in the most (1807) research grants is diabetes. About 7 million Americans suffer from diabetes, according to this household survey; about one third of them are limited in activity by this condition. Although arthritis is far more prevalent, afflicting 32 million Americans, the number of research grants mentioning it (609) is much smaller.

Table 4 presents correlation coefficients of the logarithms of these variables and related measures of condition severity. This table indicates that the number of research grants mentioning a chronic condition is completely uncorrelated with the number of people with the condition and with the number who have seen a physician about that condition. Research activity is weakly positively related (p-value = .08) to the number of people who have been hospitalized for a condition. It is very strongly positively related (p-value = .0003) to the number of people whose activities are limited by that condition. Somewhat surprisingly, research activity is significantly positively correlated with the *proportion* of people who have seen a doctor or been hospitalized, as well as those whose activities are limited.[18]

---

[17] In this section the measure of public research activity we use is the number of grants rather than the dollar value of those grants. For technical reasons, the former is much easier to compute. Substitution of the former for the latter will not affect our results if the average size of grants is uncorrelated across conditions with the number of grants. In the future we plan to compute the distribution of dollars by condition and to integrate the premature mortality and chronic-condition prevalence analyses.

[18] This is particularly surprising since, as the second column of the Table indicates, these proportions are significantly inversely related to prevalence *per se*: conditions that are more prevalent tend to be less

The determinants of the number of research grants citing chronic conditions are further analyzed in Table 5. The first column presents the regression of ln(NGRANTS95) on a measure of condition prevalence (ln(N)) and severity (%LA). As one might expect given the simple correlations in the previous table, only the severity measure has a significant positive effect on research activity. In the second column, we estimate an alternative functional form of the relationship; the regressors are the logarithms of the number of people with the condition whose activities are (N * %LA) and are not (N * (1 - %LA)) limited by the condition. The coefficient on the former is positive and highly significant, indicating that the amount of public research about a chronic condition increases with the number of people whose activities are limited by that condition.[19] Moreover, the amount of public research is significantly *inversely* related to the number of people who have a condition but whose activities are not limited by it. This could conceivably signify that, the greater the number of people who have a condition but are not seriously affected by it, the greater the odds that an adequate treatment for the condition already exists, and the less worthy that condition is of further research. This inverse relation becomes insignificant, however, when we include (in column 3) measures of the income- and age-distribution of persons reporting the condition. This regression indicates that there tends to be more research about chronic conditions that are prevalent among people living in low-income (below $10,000) households, and that are prevalent among the young (under age 18) and the old (above age 75). This suggests that the poor, the young, and the very old may derive disproportionately large benefits from government-sponsored biomedical research. In the previous section we reported that the amount of publicly-funded research on a disease decreases with the share of life-years before age 65 lost to the disease that are lost by non-whites. Since non-whites are more likely to be poor than whites, it is surprising that chronic conditions prevalent among the poor tend to be more intensively researched.

5. Summary

We have developed a simple theoretical model of the allocation of public biomedical research expenditure, and presented some empirical evidence about the determinants of this allocation. The implications of the theoretical model are consistent with government officials' descriptions of the allocation process: the structure of expenditure should depend upon research

---

severe (i.e., associated with lower probabilities of hospitalization, activity limitation, and physician consultation).

[19] As in the analysis of premature mortality, however, the elasticity is significantly less than unity.

productivity (or "scientific opportunity") as well as on public health need, i.e. the societal and economic burden of the disease/condition.

Although, we lack, at this point, useful indicators of research productivity (i.e., of the *cost* of achieving research advances), we have a number of measures of disease burden (i.e., of the *benefit* of achieving these advances). Analysts of technological change typically have data on neither the costs nor the benefits of technical advance. Failure to measure research productivity will not necessarily bias our estimates; if it does, it seems likely to bias them towards zero.

We calculated distributions of government-funded biomedical research expenditure, by disease, from records of all research projects supported by the United States Public Health Service; in fiscal year 1995, there were records of 63,289 projects whose total value was $10.1 billion. Some research expenditure cannot be assigned to specific diseases, in some cases because the research being conducted is basic in nature. The distribution of research expenditure by disease that we constructed is quite similar to one calculated by NIH based on data provided by NIH institutes, centers, and divisions (ICDs) designed to "reflect NIH-wide resources devoted to research on the listed diseases" (as opposed to budget figures for the ICD identifying the cost data).

We performed an empirical examination of the relationship of public research expenditure to a number of measures of disease burden. To avoid "sample selection bias," and to obtain a reasonably complete accounting of disease burden, we utilized data on both the dying (from the Vital Statistics-Mortality Detail file) and the living (from the National Health Interview Survey).

The mortality-related measure of disease burden we use is life-years lost before age 65. We found a very strong positive relationship across diseases between total life-years lost and public R&D expenditure (although the slope of this relationship was smaller than that implied by the theory, perhaps due to failure to measure research productivity). Further analysis indicated that research expenditure is positively correlated with life-years lost by whites but not with life-years lost by non-whites. In other words, the amount of publicly-funded research on a disease decreases with the share of life-years before age 65 lost to the disease that are lost by non-whites. A possible explanation for this finding is that lack of scientific knowledge is a less important cause of premature mortality among non-whites than it is among whites.

Disease prevalence and severity data for the (living) population provide additional indicators of disease burden. We found that the number of research grants mentioning a chronic condition is completely uncorrelated with the number of people with the condition and with the number who have seen a physician about that condition. Research activity is weakly positively related to the number of people who have been hospitalized for a condition, and very strongly

positively related to the number of people whose activities are limited by that condition. Moreover, there tends to be more research about chronic conditions that are prevalent among people living in low-income households, and that are prevalent among the young (under age 18) and the old (above age 75).

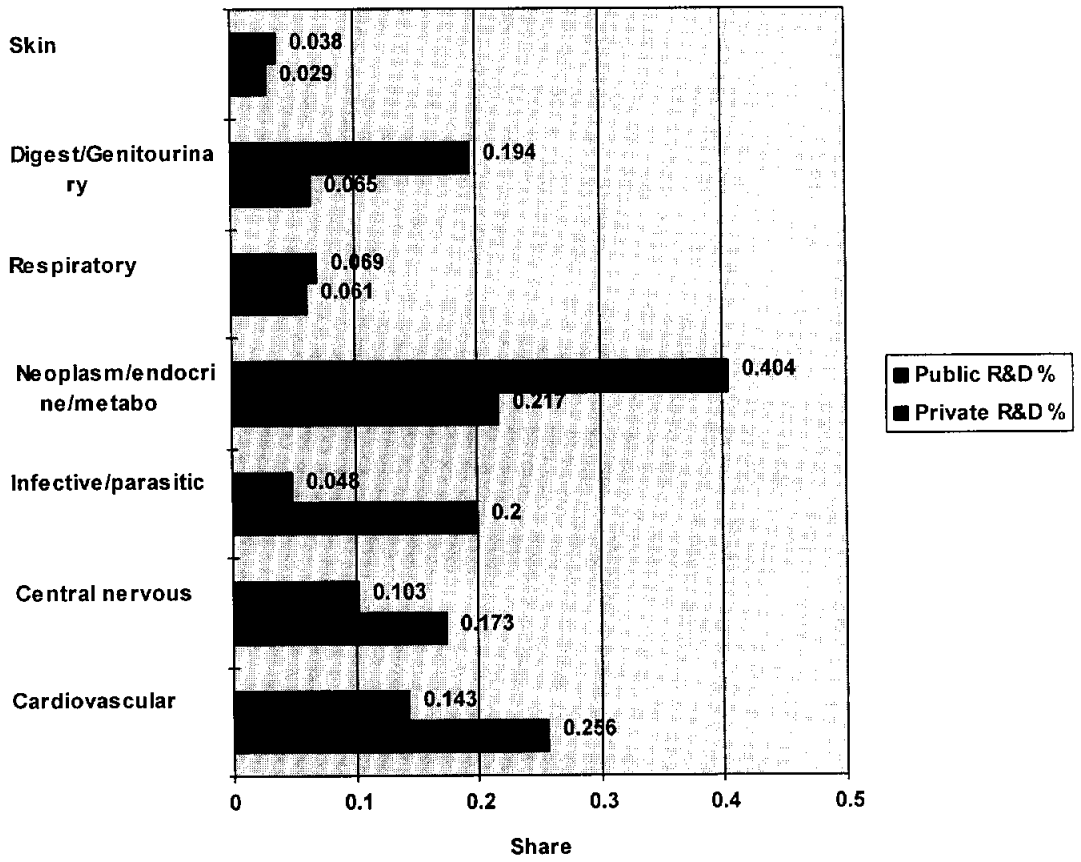
Table 1

Classification systems for diseases used in CRISP database

blood disorder 04273600
calcium disorder 05316510
cardiovascular disorder 05710209
communicable disease 07156766
communication disorder 15792995
congenital disorder 07231051
connective tissue disorder 07297208
digestive disorder 40000163
ear disorder 09775187
endocrine disorder 10255693
enzyme deficiency 40010049
eye disorder 11148096
genetic disorder 12547727
hernia 09445779
immunopathology 15604280
infection 40000216
injury 15823104
lymphatic disorder 04277757
mental disorder 24836609
metabolism disorder 18462030
musculoskeletal disorder 40000257
neoplasm /cancer 20000173
nervous disorder 20422001
- autonomic disorder 20423808
  - central nervous system disorder 20424612
    - brain disorder 04850499
    - cataplexy 20573270
    - central nervous system neoplasm 20125421
    - degenerative motor system disease 20573603
    - encephalomyelitis 20424989
    - gliosis 20422145
    - hemiplegia 20573642
    - meningitis 20425301
      - infectious meningitis
        - bacterial meningitis 20425411
        - viral meningitis 20425450
          - lymphocytic choriomeningitis 20425332

[other disorders]

Figure 1
Shares of Public and Private Health R&D Allocated to Major Disease Categories in 1982



Sources:
Public R&D: 1982 CRISP file.
Private R&D: PhRMA Annual Survey

Table 2
Life-Years Lost Before Age 65 in 1980, and Public R&D Expenditure in 1982, 14 Major Disease Categories

Disease/Disorder (ICD9 codes)	Life-Years Lost Before Age 65 in 1980	Public R&D Expenditure in 1982 (millions of $)
Diseases of the Circulatory System (390-459)	2043559	117
Neoplasms (140-239)	1860531	113
Congenital Anomalies (740-759)	760820	37
Diseases of the Digestive System (520-579)	503531	27
Diseases of the Respiratory System (460-519)	434770	47
Diseases of the Nervous System and Sense Organs (320-389)	294239	118
Endocrine, Nutritional, and Metabolic Diseases and Immunity Disorders	223015	168
Infectious and Parasitic Diseases (001-139)	162568	33
Mental Disorders (290-319)	124407	70
Diseases of the Genitourinary System (580-629)	86015	73
Diseases of the Blood and Blood-Forming Organs (280-289)	49814	16
Diseases of the Musculoskeletal System and Connective Tissue	37403	25
Complications of Pregnancy, Childbirth, and the Puerperium (	12536	8
Diseases of the Skin and Subcutaneous Tissue (680-709)	7848	21

Figure 2
The Relationship Between Government Research Funding, by disease, in 1982, and Life-Years Lost Before Age 65, by disease, in 1980

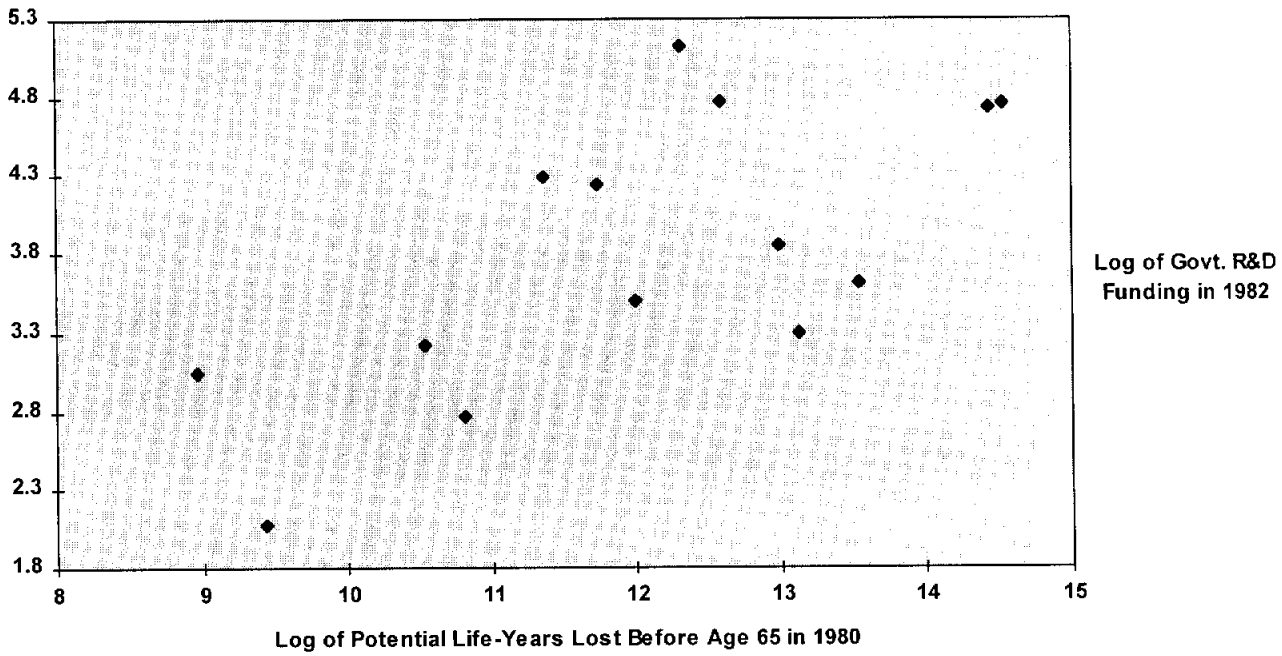


Figure 3

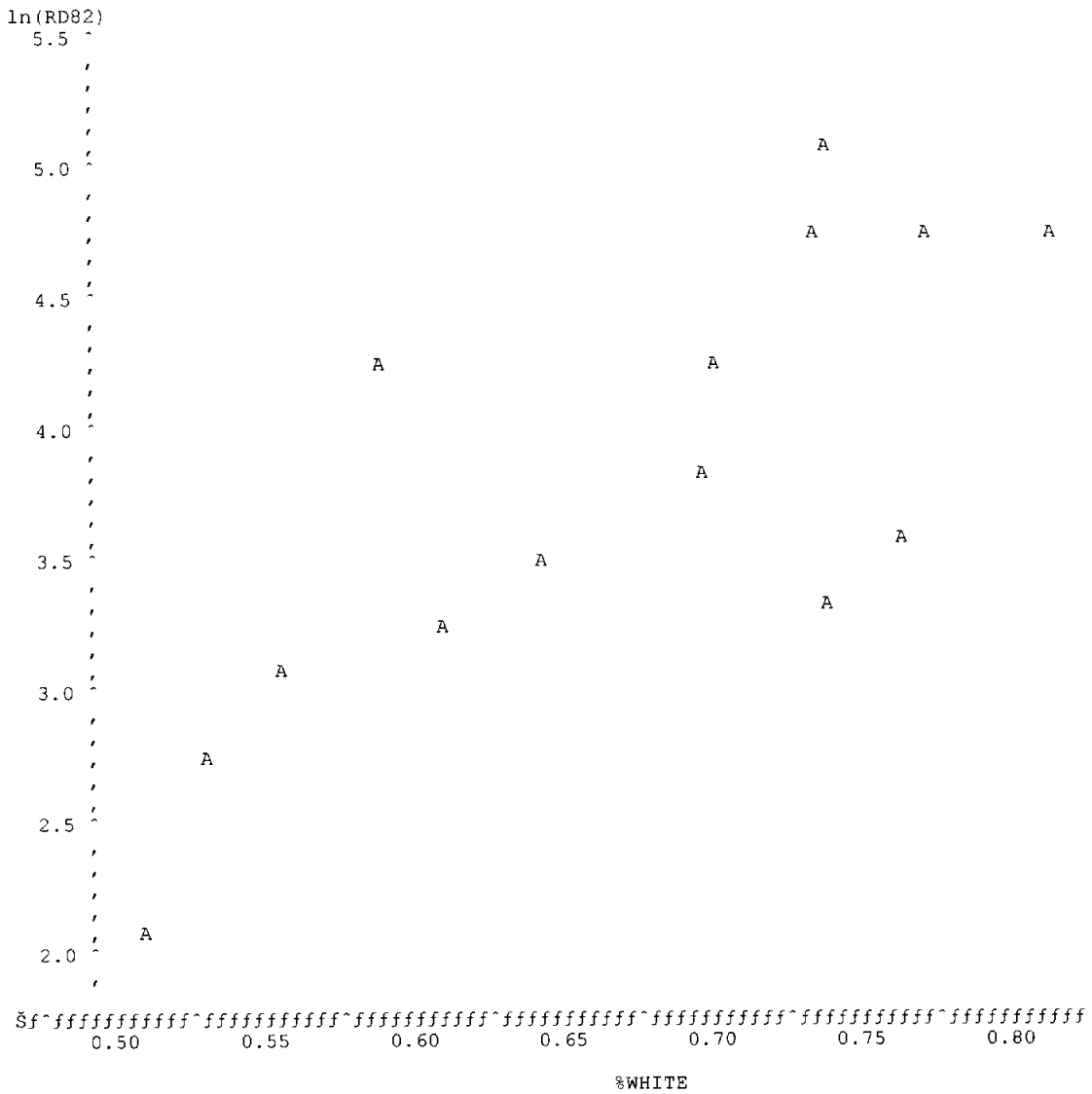

Plot of ln(RD82)*%WHITE
ln(RD82)
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
0.50
0.55
0.60
0.65
0.70
0.75
0.80
%WHITE

Table 3

Number of FY1995 Research Grants Citing, and Number of People Reporting and Limited in Activity by, Major Chronic Conditions

Key:
NGRANT: Number of FY1995 grants mentioning condition
N: Average number of people (in thousands) in 1990-92 reporting that they have the condition
NLA: Average number of people (in thousands) in 1990-92 reporting that their activities are limited by the condition

NGRANT	N	NLA	Chronic condition
1807	6962	2415.81	Diabetes
1540	27600	2925.60	High blood pressure (hypertension)
671	1293	373.68	Diseases of retina
609	31788	6739.06	Arthritis
593	1513	78.68	Diseases of prostate
573	3739	157.04	Anemias
493	11482	2503.08	Asthma
425	1243	551.89	Epilepsy
402	1562	1366.75	Mental retardation
315	8169	1290.70	Blindness and other visual impairments
293	766	130.22	Liver diseases including cirrhosis
288	3002	1077.72	Cerebrovascular disease
282	23266	1279.63	Deafness and other hearing impairments
258	7732	2435.58	Ischemic heart disease
241	180	124.92	Multiple sclerosis
241	2725	555.90	Speech impairments
216	802	190.07	Malignant neoplasm of breast
203	6416	391.38	Cataracts
195	2433	326.02	Glaucoma
119	2333	160.98	Enteritis and colitis
118	741	132.64	Congenital heart disease
118	834	45.04	Disease of the esophagus
103	322	199.96	Malignant neoplasms of stomach intestines colon and rectum
99	218	132.11	Malignant neoplasms of lung bronchus and other respiratory
90	1984	17.86	Menstrual disorders
85	2378	49.94	Psoriasis
84	1861	820.70	Emphysema
84	1325	46.38	Kidney infections
84	1911	133.77	Tachycardia or rapid heart
83	7868	503.55	Heart rhythm disorders
83	344	76.02	Malignant neoplasm of prostate
80	4201	327.68	Ulcer gastric duodenal and/or peptic
74	2269	88.49	Malignant neoplasms of the skin
61	2074	199.10	Hardening of arteries
59	3121	240.32	Gastric ulcer
54	73	1.97	Benign neoplasm of breast
53	217	6.94	Cleft palate
50	3003	30.03	Gastritis and duodenitis
49	468	24.80	Peptic ulcer
47	518	23.83	Neuritis or neuralgia unspecified
46	258	191.95	Cerebral palsy
44	9273	139.10	Dermatitis
42	1009	34.31	Kidney stones

NGRANT	N	NLA	CONDNAME
35	332	80.01	Chronic ulcer of skin
33	226	35.03	Aneurysm
32	1616	40.40	Bladder infections
31	613	63.75	Duodenal ulcer
31	184	4.05	Inflammatory disease of female genital organs
31	221	36.02	Malignant neoplasm of female genital organs
25	4302	30.11	Constipation

24	862	18.10	Benign neoplasms of the skin
22	12884	257.68	Chronic bronchitis
21	324	84.89	Pneumoconiosis and asbestosis
20	478	26.77	Goiter
20	9992	339.73	Migraine headache
16	2029	306.38	Rheumatic fever with or without heart disease
15	4904	.	Acne
14	2167	208.03	Gout
14	454	77.18	Rheumatism unspecified
13	746	37.30	Benign neoplasm of female genital organs
13	33736	168.68	Chronic sinusitis
13	7144	57.15	Tinnitus
11	2697	5.39	Color blindness
10	24060	336.84	Hay fever or allergic rhinitis without asthma
9	1068	117.48	Gallbladder stones
7	4976	2249.15	Intervertebral disc disorders
5	2058	133.77	Sciatica (including lumbago)
4	4768	467.26	Hernia of abdominal cavity
3	2633	252.77	Bone spur or tendinitis not otherwise specified
3	7403	162.87	Varicose veins of lower extremities
2	18144	4064.26	Deformities or orthopedic impairments of back
2	727	125.04	Phlebitis thrombophlebitis
2	690	11.04	Pleurisy
1	1508	.	Chronic laryngitis
1	4276	94.07	Heart murmurs
1	9441	37.76	Hemorrhoids
1	805	3.22	Nasal polyps
1	1686	45.52	Spastic colon
0	2907	26.16	Bunions
0	4674	247.72	Bursitis not elsewhere classified
0	2836	11.34	Chronic disease of tonsils and adenoids
0	4731	23.66	Corns and calluses
0	5078	619.52	Curvature or other deformity of back or spine
0	1646	3.29	Deviated nasal septum
0	1999	85.96	Diverticula of intestines
0	3698	44.38	Flat feet
0	6437	70.81	Indigestion and other functional disorders of the stomach
0	6078	18.23	Ingrown nails
0	1219	0.00	Noninflammatory disease of female genital organs
0	1249	2.50	Sebaceous skin cyst

Table 4

Correlations Between Research Activity and Prevalence/Severity of Chronic Conditions (Pearson Correlation Coefficients / Prob > |R| under Ho: Rho=0 / Number of Observations)

	LGRANTS	LN	LNLA	LNHOSP
LGRANTS log(no. of research grants)	1.00000 0.0 78	0.03858 0.7374 78	0.39969 0.0003 76	0.19966 0.0817 77
LN log(no. of people w. condition)	0.03858 0.7374 78	1.00000 0.0 126	0.54194 0.0001 123	0.61135 0.0001 125
LNLA log(no. w. limited activity)	0.39969 0.0003 76	0.54194 0.0001 123	1.00000 0.0 123	0.73564 0.0001 122
LNHOSP log(no. hospitalized)	0.19966 0.0817 77	0.61135 0.0001 125	0.73564 0.0001 122	1.00000 0.0 125
LNPHYS log (no. seeing physician)	0.07243 0.5286 78	0.99309 0.0001 126	0.59053 0.0001 123	0.66237 0.0001 125
LA % w. limited activity	0.34507 0.0023 76	-0.26328 0.0031 124	0.49119 0.0001 123	0.11785 0.1942 123
HOSP % hospitalized	0.21568 0.0596 77	-0.45766 0.0001 125	0.09308 0.3079 122	0.28156 0.0015 125
PHYS % seeing physician	0.32167 0.0041 78	-0.42512 0.0001 126	0.17895 0.0477 123	0.18110 0.0433 125

Table 5
Determinants of Number of FY1995 Research Grants
Mentioning Chronic Conditions (N = 54)

	Eq. 1	Eq. 2	Eq. 3
Variable			
ln(N)	0.142 (0.73)		
%LA	4.45 (2.77)		
ln(N * %LA)		0.651 (4.12)	0.369 (2.39)
ln(N * (1 - %LA))		-0.436 (2.17)	-0.167 (0.88)
%INCOME<$10K			8.61 (2.17)
%AGE<18			5.67 (2.63)
%AGE>75			7.30 (2.73)
Intercept	2.09 (1.31)	3.82 (2.81)	0.267 (0.17)
R^2	0.1317	0.2460	0.4470

The dependent variable is the log of the number of FY1995 grants.
N: Average number of people (in thousands) in 1990-92 reporting that they have the condition
%LA: Fraction of people reporting that their activities are limited by the condition
%INCOME<10$K: Fraction of people with household income < $10K
%AGE<18: Fraction of people under 18 years of age
%AGE>75: Fraction of people over 75 years of age

References

Adams, James D. (1990), "Fundamental Stocks of Knowledge and Productivity Growth," *Journal of Political Economy* 98 (4), August, 673-702.

Collins, J. G. (1997), *Prevalence of selected chronic conditions: United States, 1990–1992.* National Center for Health Statistics. Vital Health Stat 10 (194).

Cutler, David (1995), "Technology, Health Costs, and the NIH," September, unpub. paper prepared for NIH Roundtable on Economics, October 19.

Garber, Alan, and Paul Romer (1995), "Evaluating the Federal role in financing health-related research," unpub. paper prepared for NIH Roundtable on Economics, October 19.

Griliches, Zvi, and Frank Lichtenberg (1984), "R&D and Productivity at the Industry Level: Is There Still a Relationship?," in *R&D, Patents, and Productivity*, Zvi Griliches, ed. (Chicago: Univ. of Chicago Press).

Henderson, Rebecca, and Iain Cockburn (1996), "Scale, scope, and spillovers: The determinants of research productivity in drug discovery," *Rand Journal of Economics* 27 (1), Spring, 32-59.

Lichtenberg, Frank (1996a), "Do (More and Better) Drugs Keep People Out of Hospitals?," *American Economic Review* 86, May, 384-8.

---------- (1996b), "The Effect of Pharmaceutical Utilization and Innovation on Hospitalization and Mortality," *NBER Working Paper No. 5418*, January.

Mushkin, Selma (1979), *Biomedical Research: Costs and Benefits*, Ballinger.

National Center for Health Statistics (1995), *Health, United States, 1994*, (Hyattsville, MD, Public Health Service).

National Institutes of Health (1993), *Cost Savings Resulting from NIH Research Support*, Second Edition, NIH Publication No. 93-3109, September.

----------, "Stories of Discovery: NIH's Contributions to Progress Against Disease," unpub. paper, no date.

---------- (1995), "Disease-Specific Estimates of Direct and Indirect Costs of Illness and NIH Support," November.

Toole, Andrew (1997),

Viscusi, W. Kip (1995), "Valuing the health consequences of biomedical research," unpub. paper prepared for NIH Roundtable on Economics, October 19.