Volume Title: Measuring Capital in the New Economy

Volume Author/Editor: Carol Corrado, John Haltiwanger and Dan Sichel, editors

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-11612-3

Volume URL: http://www.nber.org/books/corr05-1

Conference Date: April 26-27, 2002

Publication Date: August 2005

Title: Pharmaceutical Knowledge-Capital Accumulation and Longevity

Author: Frank R. Lichtenberg

URL: http://www.nber.org/chapters/c10623

Chapter pages in book: (237 - 274)

Pharmaceutical Knowledge-Capital Accumulation and Longevity

Frank R. Lichtenberg

People value leisure time as well as goods, so longevity increase is an important part of economic growth, broadly defined. Research and development (R&D) is the principal source of economic growth, and the pharmaceutical industry is the most R&D-intensive sector of the economy. In this paper we assess the contribution of pharmaceutical R&D to longevity increase (hence to economic growth) by analyzing the relationship between Food and Drug Administration (FDA) approvals of new molecular entities and changes in the age distribution of deaths from all diseases, using longitudinal disease-level data.

Until the middle of the twentieth century, analyses of long-run macroeconomic performance were based on an aggregate production function of the form

(1) $$Y = F(K, N),$$

where Y is real GDP (the market value of goods and services produced), K is capital input, and N is labor input.

Capital is, of course, a produced means of production, which accumulates according to the perpetual inventory equation $K_{t+1} = (1 - \delta)K_t + I_t$, where δ is the depreciation rate and I denotes investment.

In the last fifty or sixty years, economists have recognized the inadequacies of this production function—its failure to account for important aspects of observed macroeconomic behavior—and have modified and

Frank R. Lichtenberg is the Courtney C. Brown Professor of Business at Columbia University Graduate School of Business and a research associate of the National Bureau of Economic Research.

extended it in several ways. The most important modifications have been expansions of the sets of both inputs and outputs accounted for.

Beginning in the 1950s, economists realized that some, if not most, of the growth in output could not be accounted for by growth in capital and labor. Some output growth was due to total factor productivity (TFP) growth—growth in output per unit of total input—which is not accounted for in equation (1). Growth in TFP was hypothesized to be due to technological progress. The production function could easily be modified to allow for the existence of technological progress:

$$(2) \qquad\qquad Y = AF(K, N),$$

where A is an index of the level of technology.

Solow (1956) demonstrated that, in the long run, the growth rate of per capita output is equal to the rate of technological progress (the growth rate of A).[1] In that paper, Solow assumed that technological progress was exogenous: it descends upon the economy like "manna from heaven," automatically and regardless of whatever else is going on in the economy (Jones 1998, pp. 32–33). But subsequent investigators have hypothesized and provided evidence that productivity growth and technological progress are endogenous—determined by investment in R&D. The dependence of technical progress on R&D is a key feature of recent theoretical ("endogenous growth") models (Romer 1990).

Griliches (1979) proposed the following model to incorporate endogenous (R&D-generated) technical change into the production function:

$$(3) \qquad\qquad Y = F(K, N, Z),$$

where Z is the stock of "knowledge capital."

Like physical capital, knowledge capital is a produced means of production, which accumulates according to the perpetual inventory equation $Z_{t+1} = (1 - \delta_Z)Z_t + RD_t$, where δ_z is the knowledge-capital depreciation rate and RD denotes R&D investment.

There are two ways in which one can use equation (3) to assess the contribution of knowledge capital to productivity growth. One is to examine the relationship (e.g., across industries) between TFP growth and the growth of Z. The other is to examine the relationship between TFP growth

---

1. As noted by Jones (1998, pp. 31 and 34), "there is *no* per capita income growth in the [basic Solow] model" (i.e., the model with no technological progress); "technological progress is *the* source of sustained per capita growth" (emphasis added). Empirical evidence about the relative contribution of TFP growth to output growth has been mixed. Solow estimated that about 90 percent of actual output growth was due to TFP growth; in 1967, Jorgenson and Griliches (1967) estimated that only about 10 percent was due to TFP growth; in 1973, Christensen and Jorgenson (1973) estimated that about half of actual output growth was due to TFP growth.

and "R&D intensity" (the ratio of R&D investment to output).[2] Under certain reasonable assumptions, the R&D intensity coefficient in the TFP regression is an estimate of the marginal product of knowledge capital and of the rate of return to investment in R&D.[3]

Numerous empirical studies (e.g., Griliches and Lichtenberg 1984; Lichtenberg and Siegel 1991) have provided strong support for the hypothesis that R&D has contributed significantly to growth in the market value of goods and services produced. But economists believe that the utility, or welfare, of individuals and nations depends not only on the goods and services they consume but also on the amount of (leisure) *time* they have. Leisure time as well as goods is an argument of the utility function. Becker defined an individual's "full income" as the value of goods consumed plus the value of leisure time "consumed." Let us define

(4) $$Y^* = G(Y, L),$$

where Y^* is "full income" (or utility) and L is leisure time.

A simple linear approximation of this function is

$$Y^* = Y + p_L L,$$

where p_L is the shadow price of leisure time (relative to the price of goods). Suppose, for simplicity, that p_L remains constant over time. Then

$$\Delta Y^* = \Delta Y + p_L \Delta L.$$

The change in full income is the change in gross domestic product (GDP) plus the change in the value of leisure time consumed. During the last century, longevity increase has been an important source of increase in the average person's leisure time over the course of the life cycle. Nordhaus (2003) estimated that "to a first approximation, the economic value of increases in longevity over the twentieth century is about as large as the value of measured growth in nonhealth goods and services." In other words, his estimates imply that $\Delta Y \approx p_L \Delta L$.

Due to the importance of leisure time in general, and longevity in particular, to economic well-being, we propose replacing GDP in the production function by "full income":

(5) $$Y^* = G(Y, L) = F(K, N, Z).$$

We hypothesize that R&D-generated increases in the stock of knowledge capital (Z) may have a positive impact on *both* components of full income: leisure time (via longevity) and consumption of goods and services.

---

2. The second approach does not require a long history of R&D investment or an estimate of the initial knowledge-capital stock.

3. Since capital and labor engaged in R&D are already included in K and N—they are "double counted"—the R&D intensity coefficient is an estimate of the *excess* return to R&D—the difference between the return to R&D and the return on ordinary investment.

According to the National Science Foundation (NSF), in 1996 16 percent of U.S. R&D was associated primarily with the life sciences; this share increased from 12 percent in 1985.

In the next section we discuss the measurement of pharmaceutical knowledge-capital accumulation. In section 7.2 we postulate an econometric model of the effect of pharmaceutical knowledge-capital accumulation on the age distribution of deaths. Measurement of changes in the age distribution of deaths, by cause of death, is discussed in section 7.3. Empirical results are reported in section 7.4, and section 7.5 presents a summary and conclusions.

7.1   Measurement of Pharmaceutical Knowledge-Capital Accumulation

The basic hypothesis we wish to investigate is that pharmaceutical R&D investment has increased the longevity of Americans, as shown in panel A of figure 7.1. For a variety of reasons, however, we didn't think that the best way to test this hypothesis is to perform an econometric analysis of the relationship between pharmaceutical R&D investment and longevity. There are two other indicators of pharmaceutical R&D investment that are potentially more fruitful to analyze than pharmaceutical R&D investment itself: pharmaceutical patents and FDA new drug approvals. We will argue that pharmaceutical patents are subject to most of the same econometric limitations as pharmaceutical R&D investment but that FDA new drug approval data provide an excellent opportunity to (indirectly) examine the R&D-longevity relationship.

New drug approvals by the FDA may be interpreted as an "intermediate good" in the R&D-longevity relationship, as shown in panel B of figure 7.1.[4] To explain the relationship between R&D investment and new drug approvals, and why the latter is a superior indicator for explaining changes in longevity, it is useful to briefly describe the process of drug development.

The FDA's depiction of the new drug development timeline is shown in figure 7.2. There are three main phases of drug development up until the time of new drug approval. The first phase is preclinical testing and R&D, including testing in animals. According to the FDA, the average duration of this phase is eighteen months. In order to proceed to the second stage, the drug sponsor must submit, and receive FDA approval of, an investigational new drug (IND) application. Upon approval of the IND, the sponsor may begin clinical R&D (human trials).

As table 7.1 indicates, there are three phases of clinical R&D. According to the FDA, the average duration of the three phases combined is five years.

4. "FDA estimates that, on average, it takes eight-and-a-half years to study and test a new drug before the agency can approve it for the general public. That includes early laboratory and animal testing, as well as later clinical trials using human subjects. Drug companies spend $359 million, on average, to develop a new drug, according to a 1993 report by the Congressional Office of Technology Assessment" (FDA 1999, p. 15).

Figure A:

Figure B:

Fig. 7.1   *A,* **Pharmaceutical R&D investment and longevity;** *B,* **Pharmaceutical R&D investment, FDA new drug approvals, and longevity**

After completing clinical R&D, the drug sponsor can submit a new drug application (NDA) to the FDA. For decades, the regulation and control of new drugs in the United States has been based on the NDA. Since 1938, every new drug has been the subject of an approved NDA before U.S. commercialization. The data gathered during the animal studies and human clinical trials of an IND become part of the NDA. According to the FDA, the average duration of the NDA review process is two years.

The FDA says that of 100 drugs for which IND applications are submitted, about 70 percent will successfully complete phase 1 and go on to phase 2; about 33 percent of the original 100 will complete phase 2 and go to phase 3; and 25 to 30 of the original 100 will clear phase 3 (and, on average, about 20 of the original 100 will ultimately be approved for marketing). This is consistent with 1990–2001 data on the number of commercial[5] INDs received and NDAs received and approved, shown in figure 7.3. The average annual number of NDAs approved (eighty-five) was 21 percent of the average annual number of INDs received (403).

As table 7.2 shows, there are seven different kinds of new drug applications. New molecular entities (NMEs) account for only about a third of all new drug approvals, but they probably account for the vast majority of pharmaceutical R&D expenditure,[6] and they are the NDAs that are most likely to increase longevity.[7]

DiMasi (2001) argues that mean drug development time has increased

5. "Commercial INDs" are applications that are submitted primarily by companies whose ultimate goal is to obtain marketing approval for a new product. There is another class of filings broadly known as "noncommercial" INDs. The vast majority of INDs are, in fact, filed for noncommercial research. These types of INDs include "Investigator INDs," "Emergency Use INDs," and "Treatment INDs."

6. Cross-sectional firm-level estimates support this hypothesis. When we compute a ("reverse") regression of a firm's average annual R&D expenditure on its average annual number of NDA approvals, by type, the number of NMEs is positive and highly significant, and the number of other NDAs is not significantly different from zero.

7. Forty-two percent of the NMEs approved during 1990–2001 were "priority-review approvals"—that is, considered by the FDA to represent "significant improvement compared to marketed products, in the treatment, diagnosis, or prevention of a disease." Only 14 percent of non-NME NDAs approved were priority-review approvals.
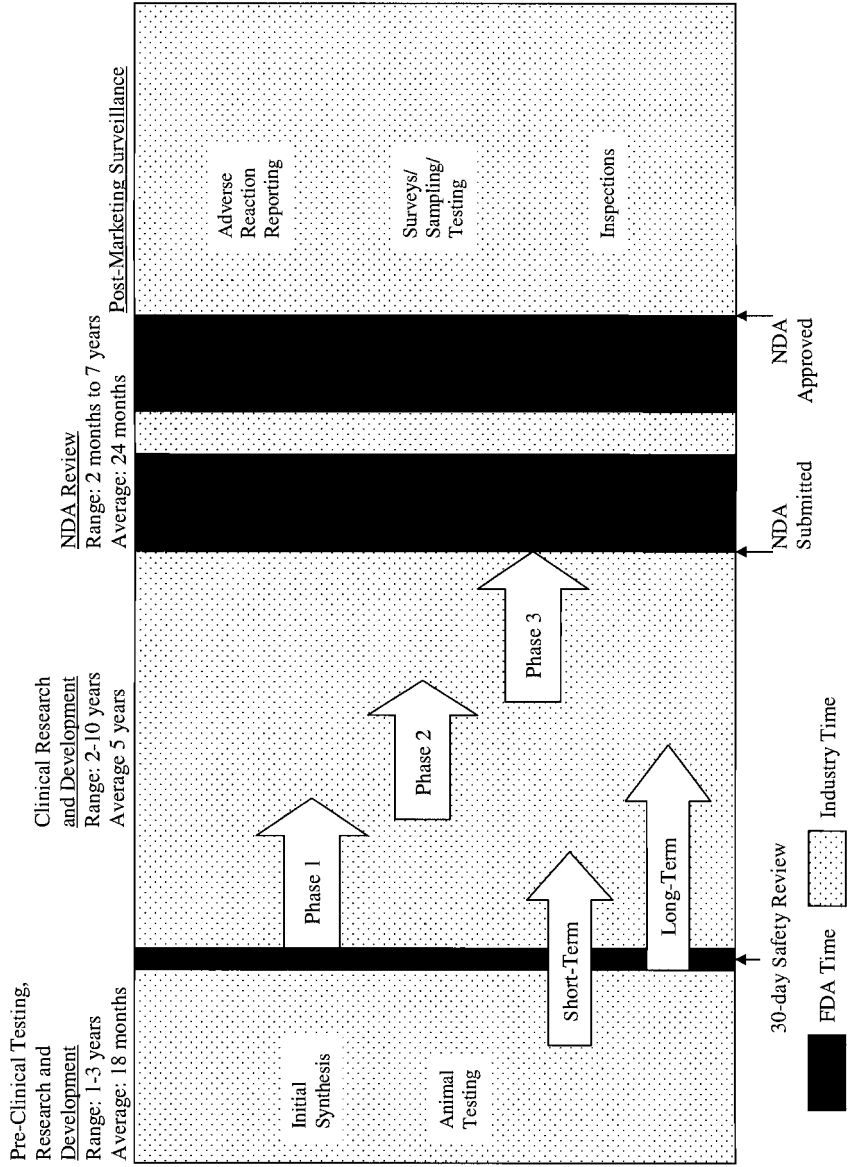
Pre-Clinical Testing,
Research and
Development
Range: 1-3 years
Average: 18 months

Clinical Research
and Development
Range: 2-10 years
Average: 5 years

NDA Review
Range: 2 months to 7 years
Average: 24 months

Post-Marketing Surveillance

Initial
Synthesis

Animal
Testing

Phase 1

Phase 2

Phase 3

Short-Term

Long-Term

30-day Safety Review

NDA
Submitted

NDA
Approved

Adverse
Reaction
Reporting

Surveys/
Sampling/
Testing

Inspections

FDA Time

Industry Time

Fig. 7.2   New drug development timeline

Table 7.1   **Phases of clinical research**

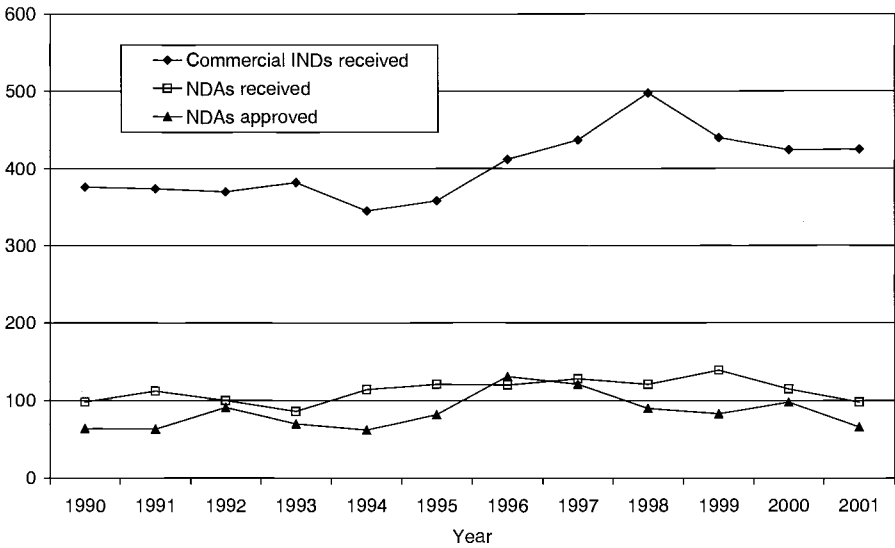| Phase	No. of patients	Length	Purpose	Percent of drugs successfully tested
1	20–100	Several months	Mainly safety	70
2	Up to several hundred	Several months to 2 years	Some short-term safety, but mainly effectiveness	33
3	Several hundred to several thousand	1–4 years	Safety, effectiveness, dosage	25–30

Fig. 7.3 Commercial INDs received and NDAs received and approved, 1990–2001

Table 7.2   **Types of new drug applications (NDAs), 1990–2001**

Percent of NDAs approved	NDA type
46	New formulation
35	New molecular entity
10	New manufacturer
6	New combination
2	New ester, new salt, or other noncovalent derivative
1	New indication[a]
1	Drug already marketed but without an approved NDA

[a]Beginning in 1994, new indications were tracked as efficacy supplements, not as NDAs.

sharply since the 1960s (see figure 7.4). His figures indicate that, for the last twenty years, mean drug development time has been 14.2 years, substantially longer than the FDA's estimate of 8.5 years.

Pharmaceutical Research and Manufacturers of America (PhRMA) provides statistics, based on its annual survey of pharmaceutical firms, on the distribution of 1999 pharmaceutical R&D expenditure by function. These are shown in table 7.3. These figures suggest that as much as 36 percent of R&D expenditure occurs during the preclinical phase of drug development, which is, on average (according to DiMasi's estimates), about eight years before NDA approval. Another 29 percent of R&D expendi-
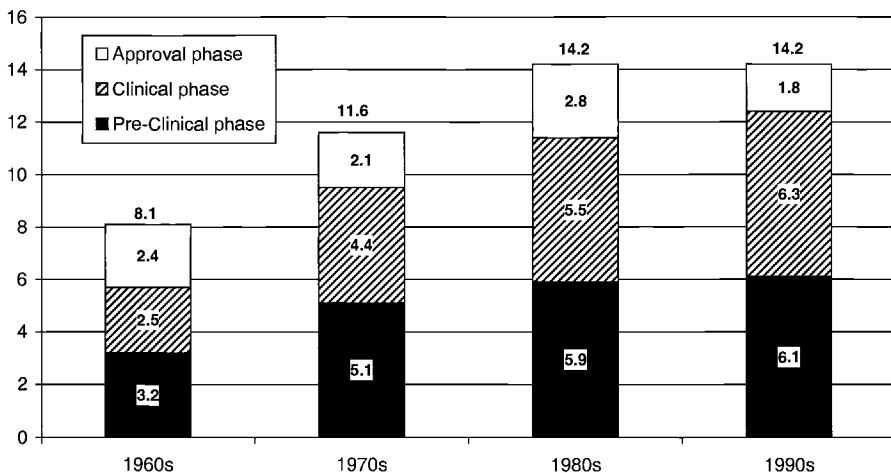

Fig. 7.4   Average number of years for drug development, 1960s to 1990s
Source: DiMasi (2001).

Table 7.3   Distribution of 1999 pharmaceutical R&D expenditure, by function

| Percent of R&D expenditure	Function
10.0	Synthesis and extraction
14.2	Biological screening and pharmacological testing
4.5	Toxicology and safety testing
7.3	Pharmaceutical dosage formulation and stability testing
29.1	Clinical evaluation: Phases I, II, and III
11.7	Clinical evaluation: Phase IV
8.3	Process development for manufacturing and quality control
4.1	Regulatory: IND and NDA
1.8	Bioavailability
9.0	Other
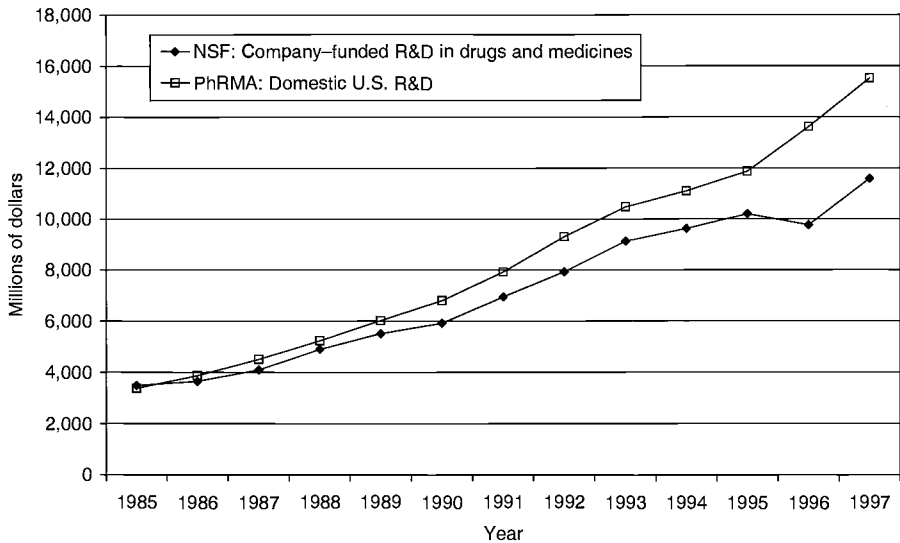
Fig. 7.5   Estimates of aggregate pharmaceutical industry R&D from NSF and PhRMA surveys

ture occurs during the clinical phase—which is, on average, about five years—before NDA approval. On average, then, the lag from R&D expenditure to new drug approval appears to be quite long and quite variable.

Long and variable lags are but one of the obstacles to a direct examination of the R&D-longevity relationship. There are several others: apparently inconsistent estimates of pharmaceutical R&D investment, smoothness of the aggregate time series R&D data, and lack of disaggregated data. We discuss these in turn.

7.1.1   Divergent Estimates

There are two distinct surveys that provide data on the amount of pharmaceutical industry R&D: the National Science Foundation (NSF) Survey of Industrial Research and Development,[8] and the PhRMA annual survey of research-based pharmaceutical companies. Figure 7.5 shows estimates of aggregate pharmaceutical industry R&D from the two surveys. Before 1990, the estimates differed by less than 10 percent, but in 1996 and 1997 the estimates differed by about 30 percent.

7.1.2   Smoothness

To identify the effect of pharmaceutical R&D investment on longevity, significant variability in R&D investment is required. As figure 7.5 suggests,

---

8. This survey is administered by the Census Bureau.

Table 7.4   Distribution of pharmaceutical R&D expenditure, by drug class,
                 1997–99 (%)

Drug class	1997	1998	1999	Average
Acting on the central nervous system and sense organs	29	30	23	27
Affecting neoplasms, endocrine system, and metabolic diseases	24	23	24	23
Acting on infective and parasitic diseases	22	22	14	19
Acting on the cardiovascular system	17	17	15	16
Acting on the respiratory system	6	4	4	5
Other human use	0	0	10	3
Biologicals	0	3	5	3
Acting on the digestive or genitourinary system	0	0	4	1
Diagnostic agents	2	1	0	1
Acting on the skin	0	0	1	0
Vitamins and nutrients	0	0	0	0

aggregate pharmaceutical R&D investment is very closely approximated by an exponential trend; that is, it exhibits very little variability.[9] This is not surprising: R&D in general is known to be very persistent, especially in comparison with ordinary investment.

7.1.3   Lack of Disaggregated Data

In principle, variability of R&D investment could be increased via disaggregation—for example, by class of drugs. Unfortunately, the NSF survey does not provide any disaggregated pharmaceutical R&D investment data. The PhRMA survey does, but as table 7.4 indicates, the drug classes are quite broad; 85 percent of investment during 1997–99 was in the largest four classes.

7.1.4   Are patent data likely to be useful?

R&D and patenting are known to be closely related (Griliches 1984). Perhaps patent data could supersede most of the limitations of the R&D data.

The U.S. Patent and Trademark Office (USPTO) publishes data on the number of patents granted for "drug, bio-affecting and body treating compositions" (patent class 514). Figure 7.6 presents annual data on the number of "drug patents" (patents in class 514) and total patents granted from 1980 to 2000. Drug patents do exhibit somewhat more variability than R&D expenditure.[10] However, drug patents track total patents quite closely.

Disaggregation of drug patents by therapeutic action appears to be in-

9. The R^2 of the regression of the PhRMA R&D series on an exponential trend is .9913.
10. The R^2 of the regression of the drug patent series on an exponential trend is .8897.
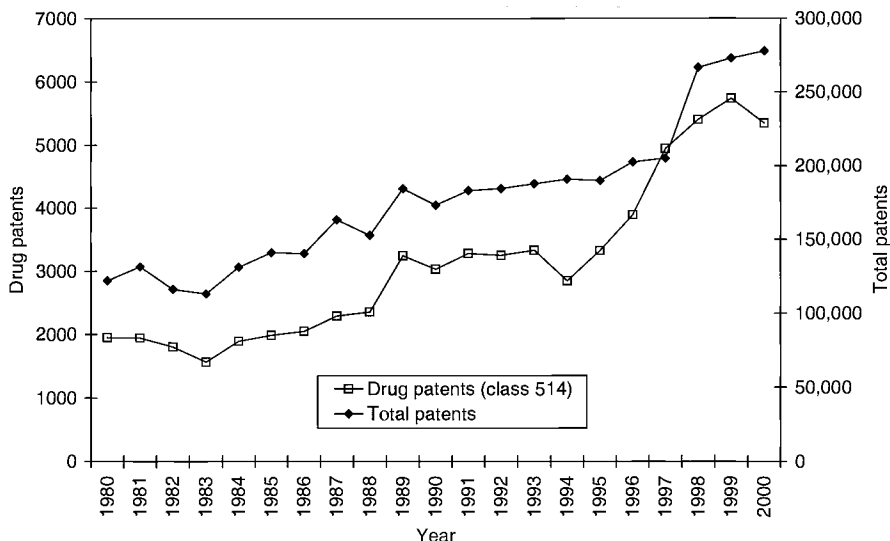
Fig. 7.6   U.S. drug patents granted and total patents granted, 1980 to 2000

feasible. Although certain subclasses of class 514 pertain to specific diseases (e.g., subclass 866 refers to diabetes, and subclass 883 refers to Hodgkin's disease),[11] these subclasses are "cross-reference art collections," and drug patents are not systematically classified by disease or therapeutic action.[12]

It appears that drug patents are often granted fairly early in the drug development cycle. According to PhRMA, "the average period of effective patent life (when a drug can be marketed) for new drugs introduced in the early to mid-1990s with patent-term restoration has been only 11–12 years. Innovators in other industries typically receive upwards of 18.5 years of effective patent life."[13] This suggests that, on average, patents are granted at least seven years prior to the market introduction of new drugs. Long and variable lags diminish the likelihood that drug patents can explain fluctuations in longevity.

Data on both pharmaceutical R&D expenditure and pharmaceutical patents are too aggregated, exhibit too little variability, and are too subject

11. See http://www.uspto.gov/go/classification/uspc514/sched514.htm.
12. The seventh (1999) edition of the International Patent Classification system appears to provide (in class A61P) a systematic classification of chemical compounds and medicinal preparations by therapeutic activity. For example, subclass 1/00 includes drugs for disorders of the alimentary tract or the digestive system, and subclass 1/18 covers drugs for pancreatic disorders (e.g., pancreative enzymes). (See http://www.wipo.int/classifications/fulltext/new_ipc/index.htm.)
13. See http://www.phrma.org/publications/publications/profile01/chapter8.phtml.

to excessively long lags to serve as a basis for testing our key hypothesis. But these limitations may be overcome by combining data from two different sources: First DataBank's National Drug Data File (NDDF)[14] and FDA data on NDA approvals.[15] These data sources enable us to compute the stock of drugs available (i.e., previously approved by the FDA) to treat a given condition in a given year.

The NDDF consists of a number of modules. One of these is the indications module, "the goal of [which] is to minimize the risks associated with drug use. The information in this module is intended to be used as a tool for assessing the appropriateness of drug therapy" (First DataBank 1998). We utilize just one part of the clinical module: the Drug Indications Master Table. This table links indications (diseases) to drugs (active ingredients): it lists all of the drugs appropriate for treatment of each indication. Indications are coded using the International Classification of Diseases, ninth revision (ICD9; http://www.cdc.gov/nchs/about/major/dvs/icd9des .htm),[16] the same classification system used in the mortality data we will analyze.[17]

Sample data from the NDDF Drug Indications Master Table, for two indications—tuberculosis and hypercholesterolemia—are shown in table 7.5.[18] The table lists eleven drugs appropriate for the treatment of tuberculosis and fourteen drugs appropriate for the treatment of hypercholesterolemia. All of the tuberculosis drugs are designated as "labeled," but

14. First DataBank, a wholly owned subsidiary of the Hearst Corporation, is a leading provider of electronic drug information. For more than two decades, it has delivered knowledge bases for various health care applications, including clinical decision support within the work flow. Its portfolio also includes comprehensive reference products, integrated content software, and specialty software for physicians and nutritionists. Many of these products help reduce the incidence of medication errors and adverse drug events, which can result in shorter hospital stays, lower medical costs, and improved patient care. The NDDF Plus knowledge base combines the drug information of the National Drug Data File with advanced clinical decision-support modules, to deliver complete descriptive, pricing, and clinical information for every drug approved by the FDA. Their staff includes clinicians, software engineers, and knowledge-base experts. It is found in installations ranging from retail pharmacies to hospital pharmacies and laboratories; physician and other health care professional practices; and e-health care companies, managed care organizations, and insurers.

15. Section 505 of the Federal Food, Drug, and Cosmetic Act states that "no person shall introduce or deliver for introduction into interstate commerce any new drug, unless an approval of an application . . . is effective with respect to such drug."

16. The International Classification of Diseases (ICD) is designed for the classification of Morbidity and Mortality information for statistical purposes, and for the indexing of hospital records by disease and operations, for data storage and retrieval. The ICD is developed collaboratively between the World Health Organization (WHO) and ten international centers, for purposes of ensuring that medical terms reported on death certificates are internationally comparable and lend themselves to statistical analysis. The ICD has been revised approximately every ten years since 1900 in order to reflect changes in understanding of disease mechanisms and in disease terminology.

17. Information about drugs appropriate for treatment of specific indications can be obtained on a piecemeal basis from http://www.medscape.com.

18. The complete Drug Indications Master Table contains almost 7,000 links between indications and drugs.

Table 7.5   **Sample data for two indications from National Drug Data File (NDDF) Drug Indications Master Table, by International Classification of Diseases, Ninth Revision (ICD9) code**

ICD9 code	Indication	Drug	Labeled or unlabeled
0119	Tuberculosis	Capreomycin	L
0119	Tuberculosis	Isoniazid	L
0119	Tuberculosis	Cycloserine	L
0119	Tuberculosis	Ethambutal	L
0119	Tuberculosis	Ethionamide	L
0119	Tuberculosis	Aminosalicyate Sodium	L
0119	Tuberculosis	Acetylcysteine (INH)	L
0119	Tuberculosis	Pyrazinamide	L
0119	Tuberculosis	Rifampin	L
0119	Tuberculosis	Rifampin and Isoniazid	L
0119	Tuberculosis, Pulmonary	Rifapentine	L
272	Hypercholesterolemia	Lovastatin	L
272	Hypercholesterolemia	Pravastatin	L
272	Hypercholesterolemia	Simvastatin	L
272	Hypercholesterolemia	Cholestyramine	L
272	Hypercholesterolemia	Colestipol	L
272	Hypercholesterolemia	Probucol	L
272	Hypercholesterolemia	Fluvastatin	L
272	Hypercholesterolemia	Atorvastatin	L
272	Hypercholesterolemia	Niacin (SA-Lipotropic)	L
272	Hypercholesterolemia	Cerivastatin	L
272	Hypercholesterolemia	Garlic	L
272	Hypercholesterolemia	Psyllium, Bran	U
272	Hypercholesterolemia	Neomycin	U
272	Hypercholesterolemia	Conj. estrogen, M-Progesterone	U

three of the hypercholesterolemia drugs are designated as "unlabeled."[19] According to the American Medical Association's Council on Scientific Affairs,

> Unlabeled uses are defined as the use of a drug product for indications or in patient populations, doses, or routes of administration that are not included in FDA-approved labeling. The prevalence and clinical importance of prescribing drugs for unlabeled uses are substantial. Unlabeled indications are especially common in oncology, rare diseases, and pediatrics. Thus, the prescribing of drugs for unlabeled uses is often necessary for optimal patient care.[20]

19. About 25 percent of the almost 7,000 entries in the Drug Indications Master Table are designated "unlabeled."

20. See http://www.ama-assn.org/ama/pub/article/2036-2420.html. See also Cranston et al. (1998).

We will construct estimates of the stock of drugs available to treat specific conditions, both excluding and including unlabeled indications.

The NDDF Drug Indications Master Table lists all of the drugs appropriate for treating given conditions that were available *in the year 1999.*[21] We want to determine the number of drugs appropriate for treating given conditions that were available *in each of the years 1979–98* (the years for which we have mortality data). To determine this, we identified, from published and unpublished FDA data, the year in which each of the drugs listed in the NDDF Drug Indications Master Table was first approved as an NME by the FDA. The FDA provided us with a list of all 821 NMEs approved by the FDA during the period 1950–93. We extended this list through 1998 using another unpublished FDA data file and data posted on the FDA website. The FDA data on NME approvals are illustrated in table 7.6, which shows NMEs approved in calendar year 2000.

We aggregated the data in the NDDF Drug Indications Master Table up to the (approximately) two-digit ICD9 level, to be consistent with the Centers for Disease Control (CDC) Mortality Data (described below). There is considerable variation across diseases—even diseases in the same broad disease groups—in the extent and timing of increases in the stock of available drugs. This is illustrated by figure 7.7, which shows, for two conditions—diseases of the thyroid gland and diseases of other endocrine glands—the number of drugs available to treat the condition in year t, as a percent of the number of drugs available to treat the condition in 1979.[22] Between 1979 and 1984, the number of drugs available to treat diseases of the thyroid gland increased 29 percent, while the number of drugs available to treat diseases of other endocrine glands increased only 13 percent. However, between 1984 and 1998, the number of drugs available to treat diseases of the thyroid gland did not increase at all, while the number of drugs available to treat diseases of other endocrine glands increased 33 percent.

The algorithm we adopted is based on the assumption that a drug linked to a condition in the NDDF Drug Indications Master Table became available to treat the condition in the year that the drug was first approved as an NME by the FDA. We know that this assumption is incorrect in at least some cases, because some of a drug's indications may be added years after the drug was first approved as an NME. Table 7.7 provides examples of New Indication approvals, and the predecessor NME approvals. Amantadine hydrochloride was initially approved as an NME in 1966 and designated as an antiviral/anti-influenza/systemic drug. Seven years later, a new indication of the drug was approved by the FDA, and it was also classified as an anti-Parkinson drug. Unfortunately, although we have complete data on NME approvals, data on New Indication approvals are incomplete.

21. Unfortunately, earlier versions of the NDDF Drug Indications Master Table are not available.
22. In 1979, there were seven drugs for treating diseases of the thyroid gland and thirty-eight drugs for treating diseases of other endocrine glands.

Table 7.6   New Molecular Entities (NMEs) approved in calendar year 2000

| NDA number	Generic name	Trade name	Dosage form	Applicant	Classification	Approval date
20-989	Cevimeline HCl	Evoxac	Capsule	Snowbrand	1S	1/11/2000
21-014	Oxcarbazepine	Trileptal	Tablet	Novartis Pharms	1S	1/14/2000
20-987	Pantoprazole Sodium	Protonix	Tablet	Wyeth Ayerst	1S	2/2/2000
21-107	Alosetron HCl	Lotronex	Tablet	Glaxo Wellcome	1P	2/9/2000
21-084	Perfluoroalkylpolyether (PFPE) Polytetrafluoroethylene (PTFE)	Skin Exposure Reduction Paste	Paste	U.S. Army Med Res and Material Command	1,4P	2/17/2000
20-789	Zonisamide	Zonegran	Capsule	Elan Pharms	1S	3/27/2000
20-971	Articaine HCl 4%; Epinephrine	Septocaine	Injectable	Deproco	1,4S	4/3/2000
21-119	Verteporfin	Visudyne	Injectable	QLT PhotoTherapeutics	1P	4/12/2000
20-938	Meloxicam	Mobic	Tablet	Boehringer Pharms	1S	4/13/2000
21-130	Linezolid	Zyvox	Tablet	Pharmacia and Upjohn	1P	4/18/2000
21-081	Insulin Glargine	Lantus	Injectable	Aventis Pharms	1S	4/20/2000
20-823	Rivastigmine Tartrate	Exelon	Capsule	Novartis Pharms	1S	4/21/2000
21-174	Gemtuzumab Ozogamicin	Mylotarg	Injectable	Wyeth Ayerst	1PV	5/17/2000
21-176	Colesevelam HCl	Welchol	Tablet	GelTex	1S	5/26/2000
20-986	Insulin Aspart Recombinant	NovoLog	Injectable	Novo Nordisk	1S	6/7/2000
20-715	Triptorelin Pamoate	Trelstar Depot	Injectable	Debio Recherche	1S	6/15/2000
20-883	Argatroban	Acova	Injectable	Texas Biotech	1S	6/30/2000
20-484	Tinzaparin Sodium	Innohep	Injectable	Dupont Pharms	1S	7/14/2000
20-610	Balsalazide Disodium	Colazal	Capsule	Salix Pharms	1S	7/18/2000
20-941	Docosanol	Abreva	Cream	Avanir Pharm	1S	7/25/2000
21-214	Unoprostone Isopropyl	Rescula	Solution	Ciba Vision	1P	8/3/2000
21-197	Cetrorelix Acetate	Cetrotide	Injectable	Asta Medica	1S	8/11/2000
21-226	Lopinavir; Ritonavir	Kaletra	Capsule	Abbott Labs	1,4P	9/15/2000
21-248	Arsenic Trioxide	Trisenox	Injectable	Cell Therapeutics	1PV	9/25/2000
20-687	Mifepristone	Mifeprex	Tablet	Population Council	1P	9/28/2000
20-873	Bivalirudin	Angiomax	Injectable	The Medicines Co.	1S	12/15/2000
21-204	Nateglinide	Starlix	Tablet	Novartis	1S	12/22/2000
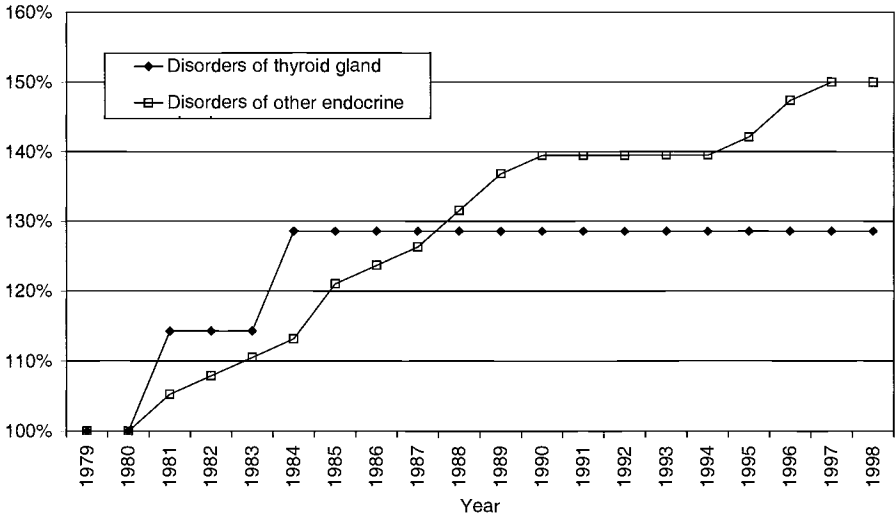
Fig. 7.7   Number of drugs available to treat condition in year *t*, as percentage of number of drugs available to treat condition in 1979

Table 7.7   Examples of NDA approvals of new indications for existing drugs

Ingredient name	Chemical type	Approval date	Therapeutic class
Amantadine	New Molecular Entity	18-Oct-66	Antiviral; anti-Influenza; systemic
Hydrochloride	New Indication	18-Apr-73	Anti-Parkinson drugs
Clotrimazole	New Molecular Entity	3-Feb-75	Fungicides (topical)
	New Indication	29-Jul-96	Antifungal (Candidiasis)
Cromolyn Sodium	New Molecular Entity	20-Jun-73	Bronchodilator
	New Indication	3-Jan-97	Respiratory
Cyclosporine	New Molecular Entity	14-Nov-83	Immunomodulators
	New Indication	22-May-97	NSAID
Cyproheptadine	New Molecular Entity	17-Oct-61	Antihistamine/oral
Hydrochloride	New Indication	18-Sep-69	Appetite stimulation
Fluoxetine	New Molecular Entity	29-Dec-87	Antidepressants
Hydrochloride	New Indication	28-Feb-94	Obsessive Compulsive Disorder
Fluticasone	New Molecular Entity	14-Dec-90	Steroids
Propionate	New Indication	7-Nov-97	Respiratory
Glycopyrrolate	New Molecular Entity	11-Aug-61	Miscellaneous Upper GI drugs
	New Indication	6-Feb-75	Anticholinergic agent
Goserelin Acetate	New Molecular Entity	29-Dec-89	GNRH agonists
	New Indication	18-Dec-95	Antineoplastic hormones
Lansoprazole	New Molecular Entity	10-May-95	Proton pump inhibitors
	New Indication	17-Jun-97	Systemic antibiotics— H. Pylori indication
Mebutamate	New Molecular Entity	11-Jul-61	Anti-hypertensive agents
	New Indication	31-Jan-75	Sedatives and hypnotics

Even if complete data on New Indication approvals by the FDA were available, in light of extensive unlabeled drug use, it is not clear how they should be used. A drug approved as an NME might be frequently prescribed for many years for a condition that is "off-label."

When the FDA receives an NDA, it assesses the drug's "therapeutic potential" and classifies it as either a "Priority Review" drug—one that represents a "significant improvement compared to marketed products, in the treatment, diagnosis, or prevention of a disease"—or a "Standard Review" drug—one that "appears to have therapeutic qualities similar to those of one or more already marketed drugs."[23] Two diseases that have similar increases in the total number of NMEs approved may have quite different increases in the number of priority review NMEs approved. For example, sixteen drugs for treating syphilis were approved during 1979–98, but only five (31 percent) of these were priority review drugs. The total number of drugs approved for treating lymph cancer was lower—fourteen—but ten (71 percent) of these drugs were priority review drugs. In our empirical analysis, we will distinguish between the stock of priority review drugs available to treat a condition and the stock of standard review drugs available.

7.2   Model

The basic model we will estimate is

(6) $\mathrm{MORT}_{it} = \beta\, \mathrm{DRUG\_STOCK}_{it} + \alpha_i + \delta_t + \varepsilon_{it},$

where MORT_{it} is an indicator of mortality (e.g., mean age at death) from ICD9 disease i ($i = 00, 01, \ldots, 99$)[24] in year t ($t = 1979, 1980, \ldots, 1998$) and $\mathrm{DRUG\_STOCK}_{it}$ is the stock of drugs available to treat disease i in year t.

The fixed disease effects (α_is) control for any determinants of mortality that vary across diseases but do not vary over time. The year effects (δ_is) control for any determinants of mortality that vary over time but do not vary across diseases. If the estimate of β is positive and significant, that indicates that diseases with *above-average* increases in the stock of drugs had *above-average* changes in the mortality indicator.

We can allow for different effects of priority-review and standard-review drug approvals by estimating the more general model

(7) $\mathrm{MORT}_{it} = \beta_P\, \mathrm{PRI\_STOCK}_{it} + \beta_S\, \mathrm{STD\_STOCK}_{it} + \alpha_i + \delta_t + \varepsilon_{it},$

---

23. Applications for new indications are also classified by therapeutic potential, and the therapeutic potential of a new indication may differ from the therapeutic potential of the NME.

24. The set of 100 diseases is exhaustive.

where PRI_STOCK$_{it}$ is the stock of priority-review drugs available to treat disease i in year t and STD_STOCK$_{it}$ is the stock of standard-review drugs available to treat disease i in year t.

We think it is worthwhile to briefly discuss how this model relates to the literature on endogenous technical change and on embodiment.

7.2.1   Endogenous Technical Change Model

In Romer's (1990) model of endogenous technological change, labor is used to produce either output or ideas:

$$L = L_Y + L_A,$$

where L is total labor, L_Y is labor used to produce output, and L_A is labor used to produce ideas.

The production function for output is

$$Y = K^\alpha (AL_Y)^{1-\alpha} \quad (0 < \alpha < 1),$$

where Y is output, K is capital, and A is stock of ideas.

The production function for ideas is

$$\Delta A = A_{+1} - A = \delta L_A^\lambda A^\phi,$$

where $0 < \lambda \leq 1$ and ϕ may be either positive or negative. The DRUG_STOCK variable corresponds to Romer's "stock of ideas" variable (A). In the empirical analysis, we count only the "ideas" (NMEs) that have been approved by the FDA.

While the model we will estimate is consistent with Romer's embodied technical change model, there are other ways in which one might specify the drugs-mortality relationship. New products and ideas do not diffuse instantaneously throughout the economy or health care system. After new drugs are introduced, some people continue to use old drugs. Hence, measures of the *vintage distribution of drugs used* to treat a given disease in a given year might be preferable to a simple count of the *number of drugs available* to treat the disease. In practice, however, measurement of the vintage distribution of drugs used by disease and year is far more difficult than measurement of the number of drugs available by disease and year. Vintage data can be constructed from the National Ambulatory Medical Care Survey (NAMCS), a survey of physician office visits that collects data on patient diagnoses and drugs prescribed. Unfortunately, prior to 1989, it was conducted in only three years: 1980, 1981, and 1985.[25] The NAMCS is an approximately 1-in-10,000 survey of office visits, so it is subject to considerable sampling error. More than one diagnosis is recorded in a significant number of visits; in these cases it is difficult to allocate or assign drugs to diseases. Drug vintages are subject to left-censoring: the vintages of drugs

---

25. It has been conducted annually since 1989.

that existed prior to the creation of the FDA in 1939 can't be determined. Finally, interpretation of the coefficient on the number of drugs available is perhaps more straightforward than interpretation of the coefficient on the mean age of drugs used.

7.2.2   Embodiment

Implicit in this specification is the hypothesis that the technical progress generated by pharmaceutical R&D is *embodied* in new drugs. Solow and other economists have recognized since the late 1950s that there are two kinds of technical progress: *disembodied* and *embodied.* Suppose that agent i in the economy (e.g., a firm or government agency) engages in R&D. If technical progress is disembodied, another agent (j) can benefit from agent i's R&D whether or not he purchases agent i's products. But if technical progress is embodied, agent j benefits from agent i's R&D only if he purchases agent i's products. Solow conjectured that most technical progress was embodied. In one paper (Solow 1962, p. 76), he assumed that "*all* technological progress needs to be 'embodied' in newly produced capital goods before there can be any effect on output."

A number of econometric studies have investigated the hypothesis that capital equipment employed by U.S. manufacturing firms embodies technological change—that is, that "each successive vintage of investment is more productive than the last" (Bahk and Gort 1993, p. 566). Equipment is expected to embody significant technical progress due to the relatively high R&D intensity of equipment manufacturers. According to the NSF, the R&D intensity of machinery and equipment manufacturing is about 50 percent higher than the R&D intensity of manufacturing in general, and 78 percent higher than the R&D intensity of all industries.

One method that has been used to test the equipment-embodied technical change hypothesis is to estimate manufacturing production functions, including (mean) vintage of equipment as well as quantities of capital and labor. Bahk and Gort (1993) argued that "we can take due account of the effect of vintage by measuring the average vintage of the stock" (p. 565). Similarly, Sakellaris and Wilson (2001) stated that "a standard production function estimation (in logs) provides an estimate of embodied technical change by dividing the coefficient on average age [of equipment] by the coefficient on capital stock" (capital's share in total cost; p. 11).

These studies have concluded that technical progress embodied in equipment is a major source of manufacturing productivity growth. Hulten (1992) found that as much as 20 percent (and perhaps more) of the Bureau of Labor Statistics (BLS) total factor productivity change (in manufacturing) can be directly associated with embodiment—the higher productivity of new capital than old capital. For equipment used in U.S. manufacturing, best-practice technology may be as much as 23 percent above the average level of technical efficiency.

Bahk and Gort (1993) concluded that "Industrywide learning appears to be uniquely related to embodied technical change of physical capital. Once due account is taken of the latter variable, residual industrywide learning [disembodied technical change] disappears as a significant explanatory variable" (p. 579). And Sakellaris and Wilson (2001) estimate that "each vintage is about 12 percent more productive than the previous year's vintage (in the preferred specification)" (p. 2) and that equipment-embodied technical change accounted for about two-thirds of U.S. manufacturing productivity growth between 1972 and 1996.

Estimation of equations (6) and (7) enables us to test the pharmaceutical-embodied technical progress hypothesis—the hypothesis that newer drugs increase longevity—and to estimate the contribution of new drugs to longevity increase.

One might be concerned that estimation of these equations could result in overestimation of the average longevity impact of pharmaceutical innovation. Suppose that the expected effect of a new drug on mean age at death is higher for some diseases than for others: instead of a single β in equation (6), there is a distribution of β_is. One might hypothesize that pharmaceutical companies would devote most of their research budgets to diseases where the expected effect of a new drug on mean age at death is highest, and therefore that most new drugs would be developed for such diseases. However a more rational investment strategy would be to invest heavily in diseases where the *total* (as opposed to *average*) expected increase in life-years is greatest. Suppose that the expected effect of a new drug on mean age at death from disease A is six months and that the expected effect of a new drug on mean age at death from disease B is one month. If ten times as many people suffer from disease B as from disease A, then the social (and presumably private) benefits to investment in disease B is higher, even though the benefit per patient is lower. Since firms will not necessarily invest more in diseases with high benefits per patient,[26] it is not obvious that the estimate of β will be an overestimate of the (weighted) mean of the β_is (weighted by number of new drugs).

7.2.3   Other Medical Innovations

New drugs are not the only type of medical innovation that might be hypothesized to contribute to longevity increase. Another important kind of innovation, and one that is also regulated by the FDA, is medical devices. If a company seeks to market a medical device, FDA approval of a pre-market approval application (PMA) is required. Premarket approval by the FDA is the required process of scientific review to ensure the safety and

---

26. In 1983, Congress passed the Orphan Drug Act, in an attempt to encourage firms to develop drugs for the treatment of rare diseases (diseases borne by fewer than 200,000 Americans). See Lichtenberg and Waldfogel (2003).
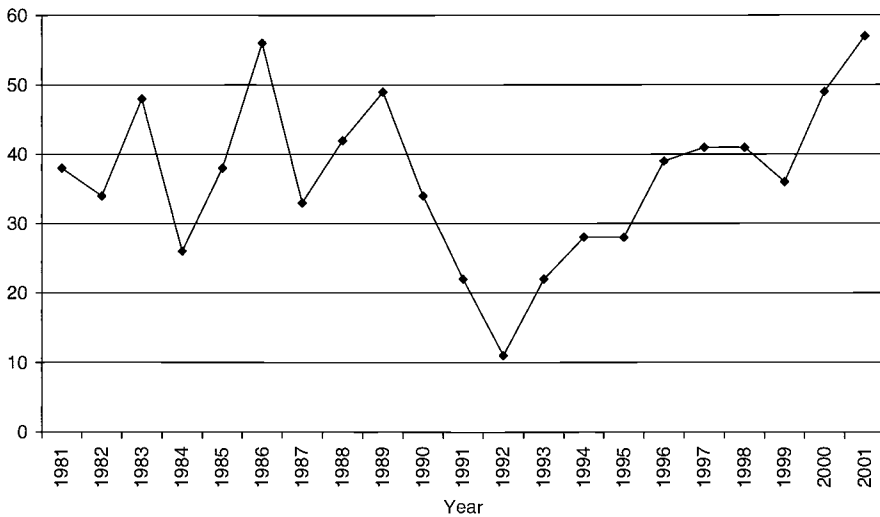
Fig. 7.8   Number of original medical device PMAs reviewed by the FDA, 1981–2001

effectiveness of all devices classified as class III devices. The FDA maintains a PMA database (http://www.fda.gov/cdrh/pmapage.html). From this database, one can construct estimates of the number of PMA approvals, by various characteristics. Figure 7.8 shows the number of original PMAs reviewed by the FDA during the period 1981–2001. One characteristic is the identity of the advisory committee that has jurisdiction over the device. As table 7.8 indicates, there are nineteen advisory committees, but two committees account for more than half of all original PMAs.[27] Moreover, it would be difficult to construct a "mapping" from PMAs classified by advisory committee to mortality data classified by ICD9 code. PMAs are also classified by "product code," but the number of distinct product codes is extremely large (almost as large as the number of PMAs—over 5,000), they do not appear to be hierarchically organized, and it would be difficult to construct a mapping from product codes to ICD9 codes. (Unfortunately, neither First DataBank nor anyone else produces a *Device* Indications Master Table.)[28]

27. Over 90 percent of PMAs are "supplemental" PMAs: applications to modify the design, manufacturing, or other aspects of original PMAs.

28. The PMA database includes "device description/indications" information, but disease coding of this information would be challenging and costly. Here is sample information about PMA Number P010054, Approval for the Elecsys Anti-HBs Immunoassay and Elecsys PreciControl Anti-HBs: "The Elecsys Anti-HBs Immunoassay is indicated for: The qualitative determination of total antibodies to the hepatitis B surface antigen (HBsAg) in human serum and plasma (EDTA). The electrochemilumin-escence immunoassay 'ECLIA' is intended for use on the Roche Elecsys 2010 immunoassay analyzer. Assay results may be used as an aid in

Table 7.8   Distribution of medical device PMAs reviewed by the FDA, 1981–2001

No. of original PMAs	Cumulative percentage of all original PMAs	Advisory committee
338	38	Ophthalmic
168	57	Cardiovascular
74	65	Microbiology
56	71	General and plastic surgery
45	76	Immunology
32	80	Gastroenterology and urology
28	83	Orthopedic
23	86	Radiology
22	88	Obstetrics/Gynecology
18	90	General hospital
17	92	Clinical chemistry
15	94	Anesthesiology
12	95	Ear, nose, and throat
12	97	Dental
10	98	Physical medicine
8	99	Neurology
6	99	Hematology
6	100	Clinical toxicology
1	100	Pathology

Hence, due to data limitations, the models we will estimate will include measures of only one type of medical innovation: new drugs. Suppose that, in reality, mean age at death depends on both new drugs and other medical innovations, but the model we estimate includes only the first variable. If the rates of pharmaceutical and other medical innovation are positively correlated across diseases—diseases with high rates of new drug introduction also have high rates of other medical innovations—then the drug coefficient will overstate the effect of new drugs per se. It is not clear that there is a positive correlation: new drugs and other medical innovations may be substitutes, rather than complements.[29] But, for the sake of argument, suppose that the correlation is positive. Indeed, suppose that there is a *perfect* positive correlation and that the estimated regression coefficients

the determination of susceptibility to hepatitis B virus (HBV) infection for individuals prior to or following HBV vaccination, or where vaccination status is unknown. Assay results may be used with other HBV serological markers for the laboratory diagnosis of HBV disease associated with HBV infection. A reactive assay result will allow a differential diagnosis in individuals displaying signs and symptoms of hepatitis in whom etiology is unknown. The detection of anti-HBs is indicative of laboratory diagnosis of seroconversion from hepatitis B virus (HBV) infection. The Elecsys PreciControl Anti-HBs is indicated for: The preciControl Anti-HBs is used for quality control of the Elecsys Anti-HBs immunoassay on the Elecsys 2010 immunoassay analyzer. The performance of the PreciControl Anti-HBs has not been established with any other Anti-HBs assay."

29. If the correlation is negative, we would *underestimate* the benefit of new drugs per se.

capture the effect of medical innovation in general, not just new drugs. Under certain reasonable assumptions, we can still identify the contribution of new drugs per se to ability to work.

Suppose that the marginal health benefit of $1 million of pharmaceutical R&D is equal to the marginal health benefit of $1 million of other biomedical R&D. Then the fraction of the overall health benefits of biomedical research attributable to new drugs is equal to the ratio of pharmaceutical R&D expenditure to total biomedical R&D expenditure. According to the Global Forum for Health Research (2002), the pharmaceutical industry accounted for 42 percent of global health R&D funded by advanced and transition countries in 1998. If only 42 percent of the estimated effect of new drug approvals on mean age at death is attributable to new drugs (and the remaining 58 percent is attributable to other medical innovations), then we should multiply the benefit estimates we obtain by 42 percent. We will therefore calculate the social (internal) rate of return to new drug development under two alternative assumptions: (a) new drugs account for 100 percent of the estimated impact on mortality, and (b) new drugs account for 42 percent of the estimated impact on mortality.

7.3   Measurement of Changes in the Age Distribution of Deaths, by Cause of Death

We used the Compressed Mortality File (CMF) to measure changes in the age distribution of deaths, by cause of death. The CMF is a county-level national mortality and population data base spanning the years 1968–99, produced by the Office of Analysis and Epidemiology, National Center for Health Statistics, Centers for Disease Control and Prevention. The mortality database of the CMF is derived from the U.S. records of deaths that occurred in 1979–99.

Deaths are classified by (a single) *underlying cause* in the CMF, as well as in other mortality databases, such as the World Health Organization's Global Mortality Database.[30] The person completing the death certificate is instructed to report, according to his or her best medical opinion, "the chain of events leading directly to death, proceeding from the *immediate cause* of death (the final disease, injury, or complication directly causing death) to the *underlying cause* of death (the disease or injury that initiated the chain of morbid events which led directly to death)."[31] For example,

---

30. See http://www3.who.int/whosis/menu.cfm?path=whosis.mort&language=english. Misallocation of deaths by cause would result in measurement error in the dependent variable. This would increase the standard error of β but would not necessarily result in any bias. Since multiple cause-of-death data files are available for the United States (see http://www.cdc.gov/nchs/products/elec_prods/subject/mortmcd.htm#description1), this issue may be explored in future research.

31. See the CDC's "Instructions for completing the cause-of-death section of the death certificate," http://www.cdc.gov/nchs/data/dvs/cod.pdf.

Table 7.9   **Sample cause-of-death section of death certificate**

		Approximate interval between onset and death
Rupture of myocardium (immediate cause)		minutes
Due to (or as a consequence of):	Acute myocardial infarction	6 days
Due to (or as a consequence of):	Chronic ischemic heart disease (underlying cause)	5 years

part I of the cause-of-death section of the certificate might be completed as in table 7.9.

The system used to classify deaths changed in 1979 and again in 1999. The three classification schemes are different enough to make direct comparisons of cause of death difficult, so our analysis is confined to the period 1979–98.[32]

Counts and rates of death can be obtained by place of residence (U.S. state and county), age, race (white, black, and other), gender, year, and underlying cause of death (four-digit ICD code or group of codes). There are seventeen age groups: under 1 day, 1–6 days, 7–27 days, 28–364 days, 1–4 years, 5–9 years, 10–14 years, 15–19 years, 20–24 years, 25–34 years, 35–44 years, 45–54 years, 55–64 years, 65–74 years, 75–84 years, over 85 years, and unknown. We excluded infant deaths (age less than one year) and deaths at unknown ages. For each approximately two-digit ICD9 code and year, we calculated two statistics: mean age at death,[33] and the fraction of deaths that occurred before age sixty-five.

Data on the number of deaths, population, and the crude death rate, by age group for 1979 and 1998 are presented in table 7.10. The crude death rate declined in every age group except the highest (over eighty-five years). Despite this, the crude death rate for the entire population increased, due to aging of the population. The fraction of deaths occurring before age sixty-five decreased from 32 percent in 1979 to 24 percent in 1998.

Figure 7.9 shows a comparison of mean age at death (from all causes) to life expectancy at birth over the period 1979–97 (Anderson 1999).[34] Life ex-

32. Data for 1979–99 are available at the website http://wonder.cdc.gov/mortsql.html.
33. We assumed that deaths in a given age interval occurred at the mean of the lower and upper ages of the interval; for example, deaths at ages 1–4 occurred at age 2.5. We assumed that deaths at ages greater than 85 occurred at age 89.5.
34. Life expectancy (e_x)—the average number of years of life remaining for persons who have attained a given age (x)—is the most frequently used life table statistic. Calculation of the complete life table is derived from the probability of death (q_x), which depends on the number of deaths (D_x) and the midyear population (P_x) for each single year of age (x) observed during the calendar year of interest. There are two types of life tables: the generation or cohort life table and the current life table. The current life table (upon which these life expectancy figures are based) does not represent the mortality experience of an actual cohort. Rather, the current life table considers a hypothetical cohort and assumes that it is subject to

Table 7.10   **Number of deaths, population, and crude death rate, by age group, 1979 and 1998**

	1979			1998		
Age group	Death count	Population	No. of deaths per 100,000 population	Death count	Population	No. of deaths per 100,000 population
1–4 years	8,108	12,637,000	64	5,251	15,189,749	35
5–9 years	5,278	16,947,000	31	3,530	19,920,862	18
10–14 years	5,868	18,445,000	32	4,261	19,241,808	22
15–19 years	21,085	21,348,000	99	13,788	19,539,327	71
20–24 years	27,634	21,096,000	131	16,839	17,674,134	95
25–34 years	47,941	36,038,000	133	42,516	38,774,410	110
35–44 years	57,723	25,114,000	230	88,866	44,519,859	200
45–54 years	135,265	22,935,000	590	146,479	34,584,884	424
55–64 years	286,966	21,448,000	1,338	233,724	22,675,970	1,031
65–74 years	449,255	15,338,000	2,929	458,982	18,395,293	2,495
75–84 years	493,676	7,598,000	6,497	681,663	11,952,189	5,703
Over 85 years	328,725	2,197,000	14,962	612,575	4,053,650	15,112
Total	1,867,524	221,141,000	844	2,308,474	266,522,135	866

Source: CDC Compressed Mortality File.

pectancy at birth is higher than mean age at death. For example, in 1997 life expectancy at birth was 76.5 years and mean age at death was 71.9 years. However, the 1979–97 increase in mean age at death (4.0 years) was greater than the increase in life expectancy at birth (2.6 years).

7.4   Empirical Results

Estimates of equations (6) and (7) are presented in table 7.11. All equations are estimated via weighted least squares, where the weight is equal to the number of deaths.[35] In the first column, the dependent variable is mean age at death, drugs not labeled for a given indication are excluded, and we do not distinguish between priority-review and standard-review drugs.

the age-specific death rates observed for an actual population during a particular period. Thus, for example, a current life table for 1997 assumes a hypothetical cohort subject throughout its lifetime to the age-specific death rates prevailing for the actual population in 1997. The current life table may thus be characterized as rendering a "snapshot" of current mortality experience and shows the long-range implications of a set of age-specific death rates that prevailed in a given year.

35. Let N_{it} represent the number of deaths from disease i in year t and N_i represent the average annual number of deaths from disease i during 1979–98. The estimates obtained when the weight is N_i are very similar to the estimates obtained when the weight is N_{it}. This reflects stability over time in the relative number of deaths caused by different diseases. About 81 percent of the increase in mean age at death during the period was within-disease increase; only 19 percent was between-disease increase (attributable to a changing mix of diseases).
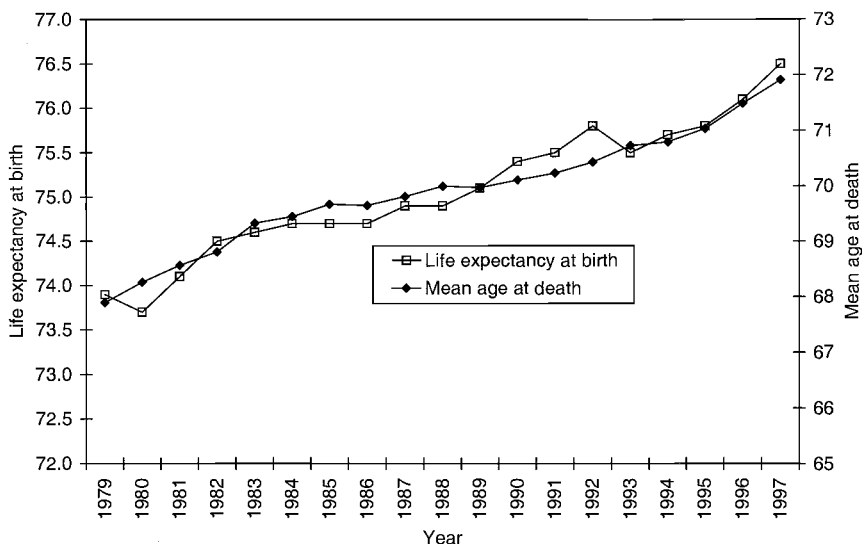
Fig. 7.9   Mean age at death versus life expectancy at birth, 1979–97

The coefficient on the total stock of drugs is positive but only marginally significant (*p*-value = .11). The second column is the same, except that unlabeled drugs listed in the NDDF Drug Indications Master Table are included. This has a modest positive effect on the point estimate of β but reduces its standard error, so that the estimate is now highly significant (*p*-value = .02). This is consistent with the American Medical Association Council on Scientific Affairs' observation that "the prevalence and clinical importance of prescribing drugs for unlabeled uses are substantial" and with the hypothesis that increases in the stock of (labeled and unlabeled) drugs to treat a condition increase the mean age at which people die from that condition.

   In the equations reported in columns (3) and (4), the dependent variable is an alternative statistic of the age distribution of deaths: the fraction of deaths that occur before the age of sixty-five. These estimates seem to confirm the estimates in the first two columns: when unlabeled indications are excluded, β is insignificantly different from zero, but when they are included, β is negative and highly significant, indicating that increases in the stock of drugs reduce the probability of dying before the age of sixty-five.

   In columns (5) and (6), the dependent variable is again mean age at death, but the stock of drugs is classified by therapeutic potential (i.e. disaggregated into priority-review and standard-review drugs). Whether or not unlabeled indications are included, β_P (the coefficient on the stock of priority-review drugs) is positive and highly significant, and β_S (the coefficient on the stock of standard-review drugs) is insignificantly different

Table 7.11 **Estimates of equations (6) and (7)**

Dependent variable	Mean age at death (1)	Mean age at death (2)	Fraction dying before age 65 (3)	Fraction dying before age 65 (4)	Mean age at death (5)	Mean age at death (6)	Fraction dying before age 65 (7)	Fraction dying before age 65 (8)	Mean age at death (9)	Mean age at death (10)	Fraction dying before age 65 (11)	Fraction dying before age 65 (12)
Unlabeled indications	Excluded	Included	Excluded	Included	Excluded	Included	Excluded	Included	Excluded	Included	Excluded	Included
DRUG_STOCK	0.01022	0.01268	−0.0001	−0.00029								
t-statistic	1.58	2.27	0.81	2.66								
p-value	0.1147	0.0232	0.417	0.0079								
PRI_STOCK					0.05324	0.07485	−0.00084	−0.00161	0.04501	0.06493	−0.0006	−0.00142
t-statistic					2.43	4.01	1.98	4.47	2.51	4.03	1.76	4.56
p-value					0.0152	<.0001	0.0484	<.0001	0.0122	<.0001	0.0794	<.0001
STD_STOCK					−0.00691	−0.00868	0.00019	0.00017				
t-statistic					0.65	1.05	0.94	1.05				
p-value					0.5129	0.2946	0.349	0.2923				

Notes: All equations are estimated via weighted least squares, where the weight is equal to the number of deaths. All equations include disease and year fixed effects.

from zero. This is not surprising, since, as discussed earlier, priority-review drugs are those that represent a "significant improvement compared to marketed products, in the treatment, diagnosis, or prevention of a disease," while standard-review drugs are those that "appear to have therapeutic qualities similar to those of one or more already marketed drugs." Once again, the estimate of β_P is larger and more significant when unlabeled indications are included than it is when they are excluded.

Columns (7) and (8) report analogous regressions, in which the dependent variable is the fraction of deaths that occur before the age of sixty-five. Once again, we cannot reject the null hypothesis that the stock of standard-review drugs has no effect on mortality, but the hypothesis that the stock of priority-review drugs has no effect on mortality can be rejected, especially when unlabeled indications are included.

Since β_S is not significant in any equation in columns (5)–(8), in columns (9)–(12) we estimate models that include *only* the stock of priority-review drugs. The estimates of β_P in columns (9)–(12) are fairly similar to their counterparts in columns (5)–(8). We will use the estimate of β_P in column (10) to evaluate the contribution of pharmaceutical knowledge-capital accumulation to the increase in the mean age at death during the period 1979–98.

Sample mean[36] values of the dependent and independent variable in 1979 and 1998 are shown in table 7.12. Mean age at death increased by 3.8 years from 1979 to 1998: ΔAGE_DEATH = 3.8 years. The mean stock of priority-review drugs increased by 6.0 drugs: ΔPRI_STOCK = 6.0 drugs. The estimated contribution of the increase in the stock of priority-review drugs to the increase in mean age at death is $\beta_P \cdot \Delta$PRI_STOCK = .065 · 6.0 = 0.39 years. The increase in the stock of priority-review drugs is estimated to have increased mean age at death by 0.39 years (4.7 months) during this period. Hence, about 10 percent of the total increase in mean age at death is due to the increase in the stock of priority-review drugs.[37]

Now we will attempt to compare the value of the longevity benefit of pharmaceutical knowledge-capital accumulation to its cost. During the period 1979–98, 508 NMEs (about 25 per year) were approved by the FDA. The Office of Technology Assessment estimated that the average cost of an NME approval is $359 million.[38] More recently, DiMasi, Hansen, and Grabowski (2003) examined the research and development cost of sixty-eight randomly selected new drugs by conducting a survey of ten pharmaceutical firms. These data were used to estimate the average pre-tax cost of new drug development. The costs of compounds abandoned

36. These are weighted means, weighted by the number of deaths.
37. This estimate may be conservative, because it includes only the *within-disease* increase in mean age at death.
38. This is the cost of *all* NMEs approved—both priority-review and standard-review. About 40 percent of NMEs are priority-review NMEs.

Table 7.12   **Sample mean values of the dependent and independent variables in 1979 and 1998**

Year	Mean age at death	Stock of priority-review drugs
1979	69.6	4.6
1998	73.4	10.6
Change	3.8	6.0

Note: Means are weighted by the number of deaths.

during testing were linked to the costs of compounds that obtained marketing approval. The estimated average out-of-pocket cost per new drug was $403 million (in 2000 dollars). They estimated that, on average, $121 million was spent during the preclinical phase and that $282 million was spent during the clinical phase. As figure 7.4 indicates, during the 1980s and 1990s the average duration of each phase was about six years, and the clinical phase was followed by an approval phase that lasted about two years. We will assume that annual expenditures were constant within each of the three phases: the industry spent about $20 million per drug in each of the first six years of drug development, $47 million per drug in each of the next six years (years 7–12) of drug development, and nothing during the approval phase (years 13–14). Hence, total expenditure on all 508 NMEs was $10.2 billion during each of the first six years of drug development, and $23.9 billion during each of the next six years (years 7–12) of drug development.

The increase in the stock of priority-review drugs is estimated to have increased mean age at death by 0.39 years during this period. There are about 2 million deaths per year, so the total number of life-years gained per year is $0.39 \cdot 2$ million $= 800,000$ life-years per year. To compute the social rate of return to new drug development, one must assign a value to a statistical life year (VSLY). Cost-effectiveness guidelines issued by the National Institute of Health (NIH) are based on a VSLY in the $50,000–$100,000 range.[39] Murphy and Topel (2003) argue that $150,000 is a more appropriate figure for the VSLY. I will calculate the social (internal) rate of return using alternative assumed values of a statistical life-year, ranging from $25,000 to $150,000. However, if we suppose, for concreteness, that the VSLY is $50,000, then the value of the annual gain in life-years is $800,000 \cdot \$50,000 = \$40$ billion. Presumably, knowledge capital does not depreciate (although it can be rendered obsolete by future knowledge capital accumulation), so even if no new drugs were approved after 1998, people

39. According to these guidelines, a medical procedure is cost effective if the associated cost per life-year gained is below $50,000 and is not cost effective if the associated cost per life-year gained is above $100,000. If the associated cost per life-year gained is in the range $50,000–$100,000, cost-effectiveness is indeterminate.
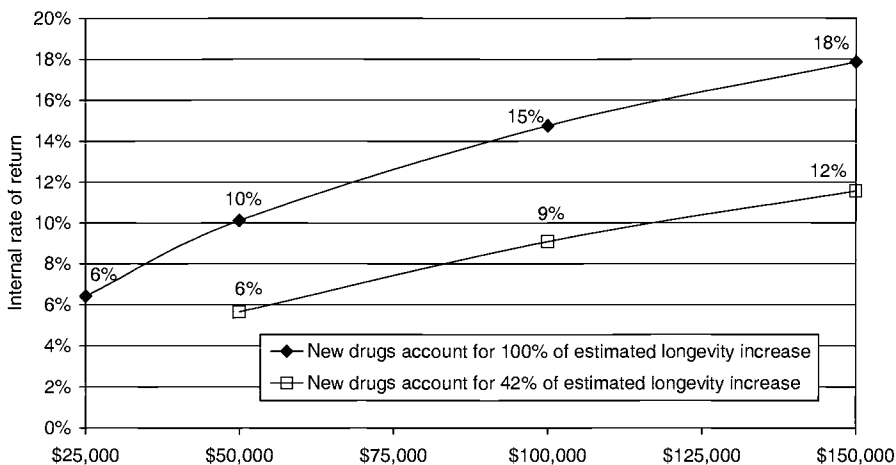
Fig. 7.10   Social rate of return to new drug development under alternative assumptions

Note: When new drugs are assumed to account for 42 percent of the estimated longevity increase, and the value of a statistical life-year is assumed to be $25,000, the internal rate of return is not defined (i.e., it is negative).

would continue to experience the 0.39-year higher life expectancy in all future years. In other words, the $40 billion may be viewed as an annuity, which the American public would start to receive fifteen years after the beginning of drug development.

Social (internal) rates of return to new drug development are shown in figure 7.10. The upper line shows the internal rate of return (IRR) for different assumed VSLY values, also assuming that 100 percent of the estimated longevity increase is due to new drugs per se (not to other medical innovations). For the range of VSLY values underlying the NIH cost-effectiveness guidelines, the IRR is in the 10–15 percent range. A VSLY of $150,000 implies an 18 percent IRR.

The lower line shows the IRR for different assumed VSLY values, assuming that only 42 percent of the estimated longevity increase is due to new drugs per se. As noted above, R&D investment by pharmaceutical firms is estimated to account for 42 percent of all (public and private) biomedical R&D investment. The IRR for a VSLY of $25,000 is not defined (i.e., it is negative). However, $25,000 is far below the range of VSLY values underlying the NIH cost-effectiveness guidelines. In this range ($50,000–$100,000), the IRR is in the 6–9 percent range. A VSLY of $150,000 implies a 12 percent IRR.

For several reasons, these IRR estimates are likely to be conservative. The estimates along the lower line assume that pharmaceutical innovation is *perfectly correlated* across diseases with other medical innovation. In re-

ality, new drugs and other medical innovations are less than perfectly complementary; they may sometimes even be substitutes. None of the estimates account for the existence of benefits other than increased longevity, including reduced hospital expenditure and reduced limitations on work and other activities (see Lichtenberg 2001, 2005). Moreover, these rates of return reflect only the value of increased longevity among Americans, who consume only about 5 percent of the world's drugs, by volume.

7.5   Summary and Conclusions

People value leisure time as well as goods, so longevity increase is an important part of economic growth, broadly defined. R&D is the principal source of economic growth, and the pharmaceutical industry is the most R&D-intensive sector of the economy. In this paper we have assessed the contribution of pharmaceutical R&D to longevity increase (hence to economic growth) by analyzing the relationship between FDA approvals of NMEs and changes in the age distribution of deaths, using longitudinal disease-level data on all diseases.

We computed the stock of drugs available (i.e., previously approved by the FDA) to treat a given condition in a given year by combining FDA data with data from First DataBank's National Drug Data File. We used the CDC's Compressed Mortality File to measure changes in the age distribution of deaths, by cause of death.

The estimates indicated that approval of standard-review drugs—drugs whose therapeutic qualities the FDA considers to be similar to those of already marketed drugs—has no effect on longevity, but that approval of priority-review drugs—those considered by the FDA to offer significant improvements in the treatment, diagnosis, or prevention of a disease—has a significant positive impact on longevity. Increases in the stock of (labeled and unlabeled) drugs to treat a condition increase the mean age at which people die from that condition, and reduce the probability of dying before the age of sixty-five.

The increase in the stock of priority-review drugs is estimated to have increased mean age at death by 0.39 years (4.7 months) during the period 1979–98. Ten percent of the total increase in mean age at death was due to the increase in the stock of priority-review drugs. The social rate of return to new drug development appears to have been quite high during the period 1979–98.

But the average annual number of priority-review drugs approved by the FDA during the period 2000–2003 (8) is only 60 percent as great as the average annual number of priority-review drugs approved by the FDA during 1990–99 (13.4).[40] Hence, there has recently been a 40 percent decline in

---

40. See http://www.fda.gov/cder/rdmt/pstable.htm.

the average annual contribution of pharmaceutical innovation to longevity increase. Hopefully this decline will soon be reversed.

References

Anderson, Robert. 1999. United States life tables. 1997. *National vital statistics reports* 47 (28). Hyattsville, MD: National Center for Health Statistics.

Bahk, Byong-Hyong, and Michael Gort. 1993. Decomposing learning by doing in new plants. *Journal of Political Economy* 101:561–83.

Christensen, L. R., and D. W. Jorgenson. 1973. Measuring the performance of the private sector of the U.S. economy, 1929–1969. In *Measuring economic and social performance,* ed. M. Moss, 233–351. New York: Columbia University Press.

Cranston, J. W., M. A. Williams, N. H. Nielsen, and R. J. Bezman, for the American Medical Association Council on Scientific Affairs. 1998. Unlabeled indications of Food and Drug Administration–approved drugs. *Drug Information Journal* 32:1049–61. http://www.ama-assn.org/ama/pub/article/print/2036-2420.html.

DiMasi, Joseph A. 2001. New drug development in the United States from 1963 to 1999. *Clinical Pharmacology and Therapeutics* 69 (5): 286–96.

DiMasi, Joseph A., Ronald W. Hansen, and Henry G. Grabowski. 2003. The price of innovation: New estimates of drug development costs. *Journal of Health Economics* 22 (2): 151–85.

First DataBank. 1998. *National drug data file documentation manual.* San Bruno, CA: First DataBank.

Food and Drug Administration (FDA) Center for Drug Evaluation and Research. 1999. From test tube to patient: Improving health through human drugs. September. http://www.fda.gov/cder/about/whatwedo/testtube-full.pdf

Global Forum for Health Research. 2002. *The 10/90 report on health research 2001–2002.* Geneva: Global Forum for Health Research, May. http://www.globalforumhealth.org/pages/index.asp

Griliches, Zvi. 1979. Issues in assessing the contribution of R&D to productivity growth. *Bell Journal of Economics* 10 (1): 92–116.

———, ed. 1984. *R&D, patents, and productivity.* Chicago: University of Chicago Press.

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,* ed. Zvi Griliches, 465–96. Chicago: University of Chicago Press.

Hulten, Charles R. 1992. Growth accounting when technical change is embodied in capital. *American Economic Review* 82 (4): 964–80.

Jones, Charles I. 1998. *Introduction to economic growth.* New York: Norton.

Jorgenson, Dale, and Zvi Griliches. 1967. Explanation of productivity change. *Review of Economic Studies* 34 (3): 249–83.

Lichtenberg, Frank. 2001. Are the benefits of newer drugs worth their cost? Evidence from the 1996 MEPS. *Health Affairs* 20 (5): 241–51.

———. 2005. Availability of new drugs and Americans' ability to work. *Journal of Occupational and Environmental Medicine* 47 (4): 373–80.

Lichtenberg, Frank, and Donald Siegel. 1991. The impact of R&D investment on productivity: New evidence using linked R&D-LRD data. *Economic Inquiry* 29 (April): 203–28.

Lichtenberg, Frank, and Joel Waldfogel. 2003. Does misery love company? Evidence from pharmaceutical markets before and after the Orphan Drug Act. NBER Working Paper no. 9750. Cambridge, MA: National Bureau of Economic Research, June. http://www.nber.org/papers/w9750.

Murphy, Kevin M., and Robert H. Topel. 2003. The economic value of medical research. In *Measuring the gains from medical research: An economic approach,* ed. Kevin M. Murphy and Robert H. Topel, 74–109. Chicago: University of Chicago Press.

National Science Foundation (NSF). Survey of Industrial Research and Development, http://www.nsf.gov/sbe/srs/sird/start.htm.

Nordhaus, William. 2003. The health of nations: The contribution of improved health to living standards. In *Measuring the gains from medical research: An economic approach,* ed. Kevin M. Murphy and Robert H. Topel, 74–109. Chicago: University of Chicago Press.

Pharmaceutical Research and Manufacturers of America (PhRMA). 2004. *Pharmaceutical industry profile 2004.* Washington, DC: PhRMA. http://www.phrma.org/publications/publications//2004-03-31.937.pdf.

Romer, Paul. 1990. Endogenous technological change. *Journal of Political Economy* 98 (October): S71–S102.

Sakellaris, Plutarchos, and Dan Wilson. 2001. The production-side approach to estimating embodied technological change. Federal Reserve Board Finance and Economics Discussion Series no. 2001-20. Washington, DC: Federal Reserve Board, April.

Solow, Robert M. 1956. A contribution to the theory of economic growth. *Quarterly Journal of Economics* 70 (February): 65–94.

———. 1962. Technical progress, capital formation, and economic growth. *American Economic Review, Papers and Proceedings* 52 (2): 76–86.

Comment   Eric J. Bartelsman

The main contribution of this paper is to provide estimates of the longevity impact of pharmaceutical R&D and estimates of the rate of return to this R&D. The bottom line of the paper is that pharmaceutical R&D investment provides a (monetized) return to U.S. residents of 18 percent through the effect of new and improved drugs on longevity. Further, 10 percent of the increase in longevity between 1979 and 1999 can be attributed to drug use. The study is important for a couple of reasons. The estimates may help guide policy to reduce the wedge between social and private returns to R&D, such as policy regarding intellectual property rights, patents, subsidies to R&D, and funding to the National Science Foundation, to name a few. Next, the paper contributes directly to the topic of this volume—namely, national accounting for intangible investment—by addressing questions such as these: what is the service flow generated by the unobserved stock of knowledge

Eric J. Bartelsman is professor of economics and director of the Economic and Social Institute at the Free University of Amsterdam.

built up by past research expenditures? How does R&D add to our wealth? How does the accumulation over time work its way into GDP figures?

In Lichtenberg's research, detailed information is drawn from a wide variety of sources, linked together, and aggregated to construct a data set with variation in the dimensions needed to identify the effect to be estimated. The choice of estimation strategy and the data-handling procedures were to a large extent predicated on the characteristics of the available raw data, and to a lesser extent on a theory describing the effect of pharmaceutical R&D on welfare of the nation.

The theory says that the stock of pharmaceutical knowledge is an input into the production function for "extended income," a welfare measure that includes leisure time and traditional GDP. The theory further states that net additions to the stock of knowledge are the output of an innovation function that has R&D performance as an input, alongside spillovers from the existing stock, as well as depreciation and obsolescence. The theory is the canonical endogenous growth model, with an innovation function and an output production function.

The empirical work requires molding the theory to fit the possibilities allowed by the available data. The rate of return on R&D calculation done in the paper is effectively simple. To start, a measure of real output is needed. To avoid the difficult issues of deflation and quality adjustment, a real indicator is defined—namely, the increase in longevity. Some possible refinements could include different weights for increases in longevity for different age groups or for increases in longevity for individuals in other drug-using countries (these are set to zero in the output measure considered here).

On the other side of the production function, there is a real input measure. As proxy for the productive flow from a stock pharmaceutical knowledge, the paper uses a count of the number of approved drugs that are available on the market each year to deal with each disease. The mean age of death or the percentage of people under age sixty-five is regressed on the number of approved drugs, with variation by disease over time, and fixed effects.

While the method is straightforward, the data collection and organization turn out to be of crucial importance. Figure 7C.1 shows how the sources of data were used. The output indicator derives from a CDC data file on deaths per year by disease. The data used in the paper are aggregated by discreet age groups, while supposedly the collection of the data was by actual age. Because the lengthening of age through drugs may take place at the high tail of the distribution, the concern is that the top-coding of the discrete age categories may create bias in the estimates.

The other place where categorization and aggregation may cause difficulties is in the definition of disease. At the highest level of the disease classification is the case where you get sick and then die. Lower down in the
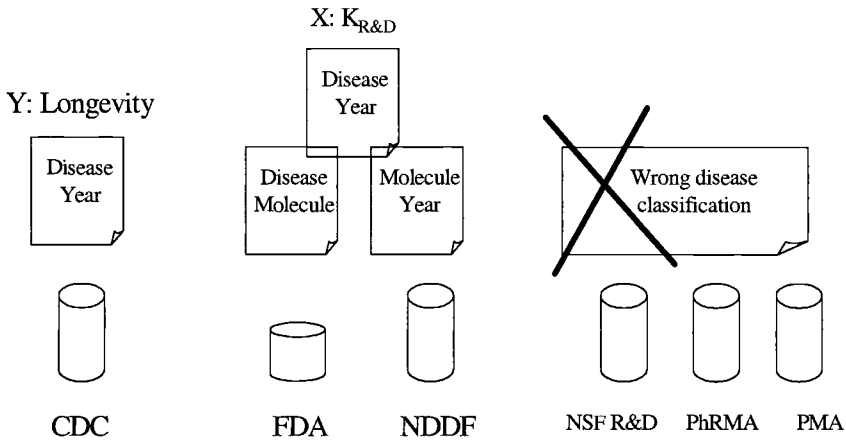
Fig. 7C.1   Research design Lichtenberg

hierarchy, you get a sniffle from some specific virus and then you die. The question is this: at what level of aggregation do you have the proper variation in your X and Y variables to get the best result? And how can you decide upon the proper level in the disease hierarchy to match up the information on drugs and death? Unfortunately, the manner in which the data are collected and published limits the ability of the researcher to define the appropriate match.

Other issues involved in calculating the rate of return are methods for valuing life-years. A considerable literature exists on this topic, and the paper uses standard estimates. More interesting for the calculation is the choice to use a stock of knowledge as the source for flow benefits to longevity. The direct approach would be to acknowledge that the R&D is embodied in the drugs and that the increase in person years actually comes from getting the prescription for the drugs. Without merged data on individual drug use and cause of death, the paper needs to work with the knowledge stock.

The knowledge stock is proxied by a very simple measure: number of approved drugs available each year for each disease. However, it is difficult to assess whether it's really the appropriate measure. A nice check on the validity, giving some more confidence in the measure, is to separately consider priority-review drugs.

The stock measure was compiled by merging a file on drugs by disease with a file on drug approvals by year, as shown in figure 7C.1. Now, the only file available of drugs by disease is from 1999. That may be a bit of a problem. If certain drugs were on a similar file in earlier years but got dropped subsequently, there may be an upward bias in the stock increase.

As an alternative, since the pharmaceutical knowledge is actually

embodied in the drugs, data on actual prescriptions by drug, by vintage of the approval date, provide a better measure of flows from the stock. The regression coefficient on impact could then be allowed to vary by vintage and would allow some estimate of the rate of obsolescence. That seems like a difficult data collection problem, but it is precisely the type of economic data that the Census Bureau is mandated to collect: sales of goods by product class. Scanner data from pharmacies would do the trick. If the nomenclature of the products sales data from the census matched those from the drug approval process, the proposed method would be a snap.

Besides data issues, real life-and-death problems make the estimation difficult. The interactions between diseases, cures, and causes of death require attention. The cause of death noted in the data file is the underlying cause precipitating the need for care, not the immediate cause of death. If certain drugs delay the onset of the immediate cause, this will not be measured. For this, a framework is needed with disease histories by person and transition matrixes by state. The empirics would gauge the effect of drugs on transitions from wellness to disease to another disease to the absorbing state of death.

As a suggestion for more feasible future work, I would like to see an explicit statement of the knowledge production function. This would link the R&D expenditures to additions to the stock, and the available flow from the stock would relate to the real output. In this framework longevity is a function of drug prescriptions, and the knowledge production function shows the arrival rate of the new drugs as a function of the R&D inputs. Some recent work on simultaneous models of production and innovation may guide further research.

This additional work would bring out the need to look at obsolescence of complementary inputs. Which wealth stocks related to health are written off owing to new drugs? Lichtenberg mentioned medical devices as being a possible reason why longevity could increase. I worry about the interaction between the various stocks on adding to the production of real output. For measuring asset stocks, it thus may not suffice for statistical agencies to do a study on depreciation rates once every few decades.

Finally, I think the paper misses closure. The paper never provides an estimate for private returns to R&D, so I can't evaluate the policy related to the wedge between social and private returns. Nor does the paper explicitly make a statement about the valuation of the current stock of intangible wealth from pharmaceutical knowledge. Having a ballpark guess of how large the intangible wealth of the sector is relative to its tangible wealth would be useful.

My closing remarks are not directly related to the paper but are suggestions for intangible investments that statistical agencies could make to improve the ability of researchers to undertake projects of the type we have
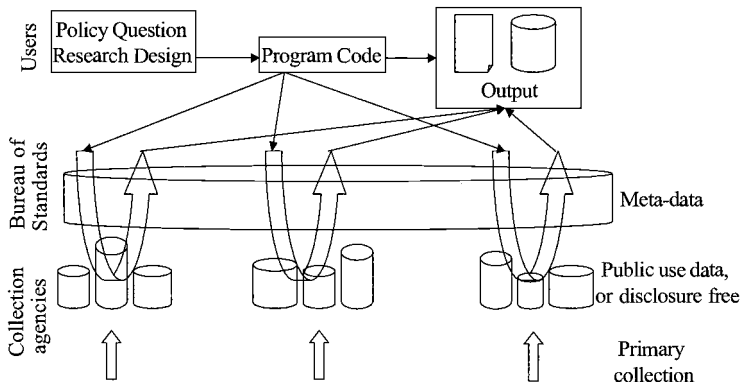
Fig. 7C.2   Statistical system for the new economy

just discussed. Figure 7C.1 shows a schematic of the research approach that was used in the paper. The research was feasible because data from different sources were made conformable and put into a database where the endogenous and exogenous variables were defined over the same units, in this case disease and year. The difficulty and time involved in this study lie in lining up the classifications to match. If the underlying data are collected using classifications that are orthogonal, no useful matching could have been done. For example, data on drug sales, or R&D expenditures by field of endeavor, could not easily be matched to the disease categories used in the study.

The basic strategy of statistical agencies is to line up data from various sources into so-called satellite accounts, at fairly high levels of aggregations. For this paper, satellite health accounts with dimensional detail that allowed linking of the various data sources would have helped.

A statistical system for the new economy would allow researchers access to data in a manner that allows complicated studies at relatively low cost in time. The statistical agencies would then not only invest in primary data collection and tabulation of results into static tables. They would collaborate on definitions of classifications and hierarchies needed to related data collected from disparate sources. Figure 7C.2 presents a view of such a statistical system. The intangible investments needed for this are similar to the expenditures being made at present in the business sector, as other papers in this conference have discussed. Firms harness information technology to change structure within their organization and the way they transact with other organizations. This is done by investing in data dictionaries that describe in as few dimensions as possible, but in a very detailed manner, everything about the products that they are buying and selling.

The collection agencies should do their best to listen to an integrating

organization—a bureau of standards—to decide on the classification schemes to use. The agencies should review all their separate little databases and build the metadata necessary so that a researcher or policymaker with a question can remotely design a research protocol and write the program code. The program code would get passed through the metadata and be translated to generate research outcomes from the underlying data sources. I think that the investments in this direction by statistical agencies would build a stock of data that would greatly enhance the ability to undertake the type of projects as presented by Lichtenberg.