THE PRODUCTIVITY POTENTIAL OF THE PHYSICIAN ASSISTANT*

RICHARD ZECKHAUSER
MICHAEL ELIASTAM, M.D.

ABSTRACT

The physician assistant is a title applied to individuals who assume at least a portion of physicians' traditional duties. This paper develops a production function methodology to estimate the potential contribution of physician assistants in the delivery of medical care. This methodology is applied to a paradigm delivery mode, an urban health center. Following upon a discussion of the possibilities for delegation and the efficient assignment of medical tasks to physician assistants, a numerical production function is estimated for physicians and physician assistants working together. When taking on his most productive assignments, it is found, a physician assistant can replace half of a full-time physician.

I. INTRODUCTION

President Nixon noted in his Health Message to Congress on February 18, 1971:

> One of the most promising ways to expand the supply of medical care and to reduce its costs is through a greater use of allied health personnel, especially those who work as physician's and dentist's assistants, nurse pediatric practitioners, and nurse midwives. Such persons are trained to perform tasks which must otherwise be performed by doctors themselves,

The authors are, respectively, Professor of Political Economy and graduate student, Kennedy School of Government, Harvard University.

* The authors are indebted to a referee who suggested an extremely helpful reorganization as part of his valuable critique. They would also like to thank Professors John Dunlop and Howard Raiffa for their helpful comments. This research was done under contract with the Health Services and Mental Health Administration, U.S. Department of Health, Education, and Welfare. A more extensive paper [10] discusses physician assistant roles, licensure, training, classification, etc., and is available from the authors on request. [Manuscript received February 1972; accepted January 1973.]

even though they do not require the skills of a doctor. Such assistance frees a physician to focus his skills where they are most needed and often allows him to treat many additional patients.

A great variety of personnel with diverse backgrounds and training might be called upon to assume these tasks. The inclusive title of physician assistant (hereafter called PA) suggests this range of possibilities.[1] The purpose here is to develop and apply a methodology to calculate the possibilities for substituting PAs for physicians.

Experience with PAs to date is very limited, and it is not possible to make firm estimates about the proportion of the physician's task that can be taken over by such personnel. A variety of approaches might be employed to give an indication of the appropriate magnitudes: (1) examine the experience of other countries; (2) make a comparative study that looks at the use of support personnel categories presently within the medical care industry; (3) look to the marketplace and make productivity calculations on the basis of salaries received.

The first two approaches might yield substantial, useful results, but only if significant resources were devoted to their pursuit. The structural characteristics of the health care sector rule out method (3). Noncompetitive elements such as barriers to entry and monopsonistic hiring practices have prevented many categories of personnel from reaching their free market price equilibria.[2] This is particularly true for PAs. No single major health care industry exists that is sufficiently free from explicit or implicit regulation and licensing to allow the unhindered play of market forces to bring about significant employment of PAs. Market prices for PA labor services can hardly be employed as firm guides for productivity estimates. Similarly, the supply of physicians, the factor of production for which they are supposed to substitute, is strongly influenced by nonmarket factors such as academic-performance-based restrictions on the purchase of medical school positions and stringent licensing requirements.

We rely on a fourth procedure which entails assessment of the productive potential of physicians and physician assistants working in an urban health center, a setting that serves as a paradigm for our analysis. The investigation proceeds in three steps. First, we define an abstract methodology designed to capture the substitution possibilities and productivity relationships to be examined. Next, we turn to our paradigm and, on an conjectural basis, derive

1 A referee suggested that it might be more appropriate to employ the term "physician-extended personnel" to apply to the diverse group of individuals who assume physicians' duties. He argued that the physician assistant was a specific type of extender. There is an urgent need for a widely accepted classification in this manpower area.

2 Were these noncompetitive elements to be removed, many current health manpower difficulties might be resolved. For example, in Massachusetts the number of nurses aides has been allowed to rise rapidly in recent years. This, together with a recession that has lured many nurses back into practice, has produced a surplus of available nursing services. *Boston Globe,* Aug. 15, 1971.

estimates of the ways in which physician assistants can take over some of the tasks of physicians. The descriptive task of the paper is then completed by incorporating these estimates into our methodological framework.

II. THE METHODOLOGY–A PRODUCTION FUNCTION FOR MEDICAL CARE

The language of economics provides a terse summary of what we are about: We wish to define and estimate a production function for medical output. An urban health center, a physician's office, a hospital emergency ward are all examples of production processes functioning in the medical sphere. The inputs to these processes are physicians, allied health personnel, drugs, bandages, stethoscopes, plaster, linoleum, etc. Different quantities of these inputs can be combined to produce different levels of output–the health care delivered. To keep track of each of these items individually would be to work as a cataloging clerk. What we seek is the analytic insight and power that derives from a simple and manageable form for the production function.

Traditionally, inputs are aggregated into categories that identify different factors of production: labor, capital, land. Our interest here is with physicians and physician assistants. They shall serve as the disaggregated argument of our production function. All other inputs are lumped together under the single heading, support.

Two technologies are considered. One employs physicians only; the other allows physician assistants to substitute for some of physicians' traditional tasks. Consider a simple abstract formulation: With traditional medical practice, the output of medical services, O, is a function of the inputs of physician time, P, and support, S. Support is treated as a flow; any stock requirements are converted to flows through application of the appropriate interest rates. Quite simply, support measured in dollars represents the difference between a physician's gross and net revenues; that is, it is the aggregate of his expenditures on all factors of production other than his own time.[3] It includes not only rent, supplies, etc., but nurses, technicians, and other traditional medical personnel. Output measured in a standard medical unit can be represented

$$O = G(P,S)$$

the production function offering constant returns to scale.

A production relationship that employs physician assistants to substitute in part for doctors represents a new technology. Identify output using this new technology as N. Its production function is

3 Financially computed figures may diverge from real inputs for a variety of reasons. For example, depreciation allowances that are used in determining net revenues are legislatively determined; they need not represent the actual diminution of capital.

$$N = F(P,A,S)$$

Both production functions are assumed to be homogeneous of degree one, with convex isoquants (allowing for substitution of support for labor inputs in both the old and the new technologies).[4]

Ideally, we would estimate the general form of $G(P,S)$ and $F(P,A,S)$ using observations on empirical situations where great varieties of factor mixes were employed. Unfortunately, we could find no hard data, in part because physician assistants are not in widespread use. Even had we found a great deal, we might have encountered estimation difficulties due to a paucity of independent observations. For example, if relative factor prices and production functions were everywhere the same, identical factor mixes would always be employed in each production process. Thus, our estimation procedure is forced to be modest in scope.

Empirical Foundation–The Substitution of Physician Assistants for Physicians

The primary purpose here is to determine how PAs can be used to substitute for physicians. Matters of exposition and computation will be simplest if we assume that for the set of extant prices, the amount of the productive input *support* that accompanies a given level of output is the same whatever the mix of labor inputs in the production process. This assumption need not rely on nonsubstitutability. For the sets of extant factor prices, the results of different optimization processes may yield the same optimal levels of support. If anything, directing attention to situations where support levels are not permitted to vary with mixes of labor factors underestimates the productivity potential of physician assistants' assuming the tasks of physicians.[5]

Represent support as the constant fraction k of output.[6] Employing the old production function, we have $O = G(P,kO)$; invoking the implicit function theorem, this yields $O = g(P)$. With the new technology, the production function

4 We are indebted to the referee for urging us to stress the substitutability of support for physician time using traditional delivery modes.

5 One indication of the gains from substitution of physician assistants into the production process is the dollar savings at a specified output level and a particular pair of prices for physicians and support. The curve indicating total savings when support is not allowed to vary can never rise above the curve which permits such variation; it may be below it. The optimum is reached where the slope of the total savings curve (i.e., the marginal savings) just equals the price of the physician assistant. Since the slopes, not the absolute heights, of the two curves are what matters, it may turn out to be the case that with no support variation permitted, more rather than less physician assistants may be employed to produce a particular level of output.

6 With support measured in dollars, and using the old technology, k would represent the differences between the physician's gross and net revenues, i.e., his expenditures on all other factor inputs. American Medical Association survey data for 1968 suggest that the typical practicing physician in the United States has an average gross income of $57,074 and a net income of $36,046. The implied support figure is $20,428. [2, pp. 73–74]

$N = F(P,A,kN)$ becomes $N = f(P,A)$. A major task of this paper will be to estimate the function $f(P,A)$.[7]

The substitution issue breaks into two convenient parts: What tasks can the physician assistant assume? What are the savings and expenditures of time involved when he does them? With this information in hand, we can return to the problem of estimating the production function.

III. POSSIBILITIES FOR DELEGATION

Legal, medical, and economic considerations play a role in determining the tasks that are most appropriately assumed by physician assistants. The lack of or expense of malpractice protection may preclude physician assistants from performing some dangerous though relatively simple tasks. Tasks that require highly developed medical judgment or advanced medical skills may be beyond the capabilities of physician assistants. (If they were to undertake the required training, they might as well become full-fledged physicians.) Finally, to train physician assistants for some specialized tasks of infrequent occurrence may be uneconomical. To employ a well-known economic analogy, physicians can perform all the tasks in this "pin factory," and the physician assistants can perform those for which they are specially trained. The remainder of this paper details procedures for determining which ones should be handled by which workers.

Dr. William Schwartz, chief of medicine at Tufts New England Medical Center and a strong advocate of the physician assistant concept, suggests that:

> . . . many of the physician's tasks now considered sacrosanct might well be taken over by non-physicians. . . . [They] could be trained in a few months to perform therapeutic abortions. . . . Orthopedic health care specialists, trained to distinguish between uncomplicated and complicated fractures, might provide service to a number of small communities or might work in large hospital emergency rooms, treating the great majority of bone injuries. . . . [T]echnicians might well carry out most diagnostic examinations of the bowel or the stomach. . . . Needle biopsies of kidney or liver could be performed by 'health care specialists' limited to this particular role.[8]

Table 1 presents a suggested breakdown of delegable and nondelegable duties according to categories of complaint at the urban health center that provides the observations and data for our calculations in the remainder of this

7 Another interpretation of $f(P,A)$ is that for each value of N it gives the output N isoquant lying on the planes $S = kN$.

8 Testimony of William B. Schwartz, M.D., before the U.S. Senate Committee on Labor and Public Welfare, Subcommittee of Health, May 10, 1971.

paper. This conjectural task delineation was carried out by Michael Eliastam, M.D., on the basis of his experience at the center and his study of the physician assistant concept. In any particular context, the assignment of tasks will be an evolutionary process with forces of law and economics as well as medicine influencing the course of development.

IV. THE POTENTIAL USE OF PHYSICIAN ASSISTANTS IN AN URBAN HEALTH CENTER

Once we understand the types of tasks that can be delegated, the next step is to estimate the savings in physician time that can be achieved. The magnitude of the savings may depend on the setting as well as on the case mix. Some physicians—diagnostic consultants, for example—may encounter very few patients whose needs could be wholly or partly met through the use of physician assistants. Others, perhaps physicians manning an outpatient pediatric clinic, may be able to delegate a significant proportion of their tasks. The data estimates below are derived directly from the experience of the center under study.[9] Their strongest implications are for other medical delivery institutions experiencing equivalent case mixes.[10]

The Services of the Center

The center studied serves a target population of roughly 25,000; it is generally thought that this is an efficient scale of operations for a facility of this sort. The center receives approximately 60,000 visits to physicians per year. The major practitioner categories are internist, with 25,000 patient visits, and pediatrician, with 30,000 visits.

Center services can be separated into adult and pediatric medicine. The staff of internists serving the adults has eight full-time equivalent (f.t.e.) physicians working a 48- to 50-hour week. Here the internist is functioning as a primary care physician (a role traditionally taken on by the general practitioner), not as a consultant in internal medicine. Approximately 50 percent of the internist's time is devoted to direct patient contact. Each f.t.e. internist sees one-eighth of the total number of patients—roughly 60 per week. At the rate of

9 The center, which wishes to remain anonymous, provided this information as part of a private communication. This type of center is one of an expanding class of new modes of health care delivery that are assigned such labels as health centers, group practices, and HMOs. The selection of this example was designed to reinforce our feeling that the success of the physician assistant concept must come with the simultaneous development of new production methods in medical care.

10 By returning to the delegation possibilities for particular tasks, they can be extrapolated (with some caution) to institutions where case mixes differ.

TABLE 1

POTENTIAL TASKS FOR DELEGATION TO PHYSICIANS' ASSISTANTS: EXAMPLES OF GENERAL PROCEDURES IN SICK AND WELL PATIENTS

Delegable Duties	Nondelegable Duties
A. Record history	A. Evaluate subtle and unusual signs
B. Conduct physical exams	B. Indicate need for more specific diagnostic tests and procedures than in initial examination
C. Perform pelvic exam and PAP smear	C. Formulate final diagnosis
D. Obtain blood, urine, throat culture, sputum, etc., specimen	D. Perform uncommon diagnostic or therapeutic tasks such as spinal tap, thoracentesis (insertion of tube into chest cavity)
E. Carry out routine preventive measures: EKG, chest X-ray, eye testing	E. Provide in-depth psychological evaluation
F. Provide general information on drugs, diet, activity	F. Initiate a drug regimen
G. Treat common minor illnesses	

24 contact hours per week, the internist averages 24 minutes of direct contact with each visiting patient.[11]

Patient Contact Units

For convenience, we arbitrarily defined a modular unit for patient contact time: one patient contact unit = 12 minutes. Of course, few visits require exactly some multiple of 12 minutes. To provide summary measures, visits for different categories of complaint were grouped into clusters according to average number of patient contact units per visit. The numbers of visits within each category provide the data for our later calculations.

Time Breakdown by Category of Complaint

Table 2 presents a breakdown by category of complaint. The right-hand column of percentage breakdowns was taken directly from the center's case records. The secondary breakdowns were made by Dr. Eliastam on the basis of his experience during a two-year association with such a center.

11 The patient contact percentage agrees well with the observation studies of Wood [9] and Chase and Craig [5]. The report by Theodore and Sutter [8] gave a higher percentage, perhaps because it suffered from the bias one might expect from a questionnaire. The length of visit estimate corresponds with the findings of Altman et al. [1].

TABLE 2
AN URBAN HEALTH CENTER DISTRIBUTION OF COMPLAINTS BY CATEGORY AND PATIENT CONTACT UNITS REQUIRED FOR TREATMENT

Category of Complaint	Patient Contact Units Required Per Visit	Percent of Total Patient Visits to Center	
	Adult Medicine		
Chronic illness	2	41%	
	3	10	51%
Acute illness	1	8	
	2	4	
	3	4	16
Subacute illness	1	7	
	2	3	
	3	3	13
Miscellaneous	1	4	
	2	3	
	3	3	10
Checkup	2	7	7
Trauma	1	1	
	2	1	2
Psychological	2	1	1
Total	Weighted average = 2,	100%	100%
	Pediatric Service		
Chronic illness	2	5%	5%
Acute illness	1	16	
	2	18	
	3	5	39
Subacute illness	1	6	
	2	5	11
Miscellaneous	1	9	9
Checkup	1	26	
	2	3	29
Trauma	1	6	
	2	1	7
Psychological	–	0	0
Total	Weighted average = 1.42	100%	100%

To illustrate some entries in the table, consider the classification "chronic illness" care. Patients with stable diabetes might need only two patient contact units, or roughly 24 minutes per visit. A severely hypertensive patient, unresponsive to therapy and requiring, in addition, significant psychological support, might require 50 percent more time—three units.

Breakdown of Pediatrician Time

The pediatric service employs seven f.t.e. pediatricians for its 30,000 annual visits, which averages out to 85 visits per doctor per week. Assuming that the pediatricians spend 50 percent of their time in direct patient contact, their 48-hour week would enable them to spend 17 minutes, on average, with each visiting patient. The average number of patient contact units per visit is 1.42.[12] The lower portion of Table 2 presents the pediatric breakdown.

Production Using Physicians and Physician Assistants

Table 2 presents the patterns of production when the old technology is employed. Unfortunately, there is no empirical data available that would enable us to prepare an equivalent table for production employing physician assistants as well as physicians. The best that could be done was to have Dr. Eliastam provide some conjectural estimates. The primary significance of the numbers presented here rests in their rough order of magnitude, and in the ways in which they are employed. Other sources, other physicians, other contexts might yield particular estimates that differ substantially from those given here. But there is likely to be less variance among gross estimates which are our primary interest, for they indicate the gains achieved when physician assistants join the medical delivery team.

V. THE EFFICIENT ASSIGNMENT OF PHYSICIAN ASSISTANTS TO TASKS

Merely assigning physician assistants to tasks will not be sufficient, as some patterns of delegation will yield greater savings than others. The principle to be followed is that of comparative advantage—what might be called *comparative efficiency* in this context.

The Comparative Efficiency Assignment Procedure

In order to determine the maximum comparative efficiency, the following procedure is used:

12 Bergman's observation study [4] yielded a 48 percent contact time figure and a somewhat shorter (11 minute) average patient visit. Deisher et al. [6] found an average visit time in accordance with ours.

1. Score all tasks on the amount of physician time required to perform them.

2. Score these tasks on (a) the amount of PA time required to perform them, and (b) the required amount of accompanying physician supervision time.

3. Compute the net savings in physician time created when the task is assigned to a PA (the score from 1, above, minus the score from 2b).

4. Rank all tasks in terms of a PA's comparative efficiency. Comparative efficiency is indicated by the ratio of net savings in physician time if a PA performs the task/PA time required to perform the task (the score from 3 divided by the score from 2a).

5. Assign tasks to the PAs, starting with those that offer the highest PA comparative efficiency ratio.[13]

A variety of complications may enter the procedure. There will be increasing returns to scale over certain ranges for certain tasks if (a) PAs perform tasks more efficiently as they perform them more often; (b) there is learning by doing; (c) physician supervisory time per PA-performed task declines as more tasks are taken on by the PA. With increasing returns, it may be necessary to consider the task-assignment procedure on a more aggregative basis. All, or at least a substantial portion, of the time devoted to a particular task would be treated as a single, indivisible task.

A related set of difficulties arises because PA training may be task-related, in which case the same PA may not be able to perform all possible delegable duties. For example, in a one-year training period, a "specialist" physician assistant might be taught to deliver chronic care to diabetic, hypertensive, and cardiac patients or to deal with routine procedures in emergency room care, but not both. If PAs were readily divisible commodities, it would be optimal to hire fractions of differently trained PAs to handle different sorts of tasks. With fractional PAs not on the market, it may become essential to look to groups of tasks most efficiently performed by single PAs.[14] The procedure for efficient assignment of PAs to tasks presents other problems as well that might enable a talented operations researcher to earn his keep.

Different Types of Physician Assistants

Just as not all physician assistants will be trained for the same tasks, not all assistants will have the same basic capabilities. The task assignments and time

13 To illustrate: Task A takes 24 minutes for a physician. It takes 30 minutes for a PA, with a supervisory requirement of 5 minutes. It offers net physician savings of 24 – 5 = 19. Its ratio of savings to PA time is 19/30 = .633. Task B takes 12 minutes for a physician; a PA does it in 14 minutes, with 1 minute of supervision. Its net savings/PA time ratio is 11/14 = .786. Task B should be undertaken before Task A.

14 Part-time PAs could conceivably handle the indivisibility problem. If tasks can be scheduled in advance, it will be possible to deal with, say, eight hours worth of tasks

assessments for this paper were carried out for what the National Academy of Science has labeled Type A assistants. Type A assistants have highly developed skills and assume general responsibilities. Type B assistants, in this classification scheme, have highly developed skills in a specific area. They might, for example, perform traditional physician tasks in connection with a renal dialysis unit. Types C and D assistants are the limited-skill counterparts of A and B, performing general and special tasks, respectively. Though the analysis here is directed to Type A assistants, the methodology could be applied equally well to the less skilled and autonomous and the most specialized categories of assistants.[15]

The Assignment of Duties to the Physician Assistant

We are now equipped to perform the operation that was the initial goal of this analysis. On paper, at least, we can allow physician assistants to take over some of the physician's burden of examination, diagnosis, and therapy. But before turning to these substitution procedures, the reader may wish to refer back to Table 1 which laid out the health delivery duties that could and could not be delegated to the physician assistant.

To assign physician assistants to tasks, we followed the five-step comparative efficiency procedure outlined above. Tables 3 and 4 show the results of these assignments. They detail how physician assistant time would most efficiently substitute for physician time in the urban health center.

What we wish to find out is what combinations of the two inputs, physician time and PA time, are required to produce the present output of the center. Thus, as the PA input is increased, the physician input is reduced just enough to keep the output level constant. For example, in Table 3 the first PA saves 120 physician patient contact units; that is, he saves one-half of an internist. The substitution of the second PA is computed for a situation starting with one PA and seven and one-half full-time physicians. It is assumed that the PA works a 40-hour week which, allowing for lunch breaks, waits for patients, etc., translates into 32 hours of patient contact, or 160 patient contact units.

What might be called the major rows of the tables detail the three categories of numbers needed to carry out the comparative efficiency procedure just cited. To illustrate this somewhat complex format, look at the chronic illness row under the 2nd physician assistant column; note that all time amounts are in patient contact units. The numbers 55 and 25 represent the time per week (in patient contact units) the PA is devoting, respectively, to 2- and 3-unit visits.

---

with a single, quarter-time PA. With some medical ailments and some patient populations, scheduling may offer the opportunity to make effective use of part-time medical personnel. In particular, this might be a good place to employ nurses who have been upgraded to PA status.

15 See [7; 10, pp. 6–12] for further discussion of the responsibilities that could be assumed by each category of assistant.

TABLE 3
AN URBAN HEALTH CENTER WITH EIGHT FULL-TIME EQUIVALENT INTERNISTS
REPLACEMENT SCHEDULE OF INTERNISTS BY PHYSICIAN ASSISTANTS[a]

Adult Medicine

Category of Complaint	Patient Contact Units Required Per Visit	Weekly Total Patient Contact Units for Physicians Only Delivery Mode	Physician Assistants			
			1st	2nd	3rd	4th
Chronic illness	2	394	160 $\boxed{20}$ 140	55 $\boxed{5}$ 46	73 $\boxed{10}$ 49	
	3	144		25 19	87 56	
Acute illness	1	38				26 $\boxed{7}$ 18
	2	38				24 14
	3	58				25 14
Subacute illness	1	33				22 $\boxed{5}$ 14

	2	29				13 8
	3	43				14 8
Miscellaneous	1	19				*3* 7 5
	2	29				7 5
	3	43				8 5
Checkup	2	67		*10* 80 67		
Trauma	1	5				*2* 6 4
	2	10				8 5
Psychological	2	10				
Total		960				
Net savings in physician time units			120	117	94	83

[a]120 patient contact units take half-time for one internist.

TABLE 4
AN URBAN HEALTH CENTER WITH SEVEN FULL-TIME EQUIVALENT PEDIATRICIANS
SCHEDULE OF PEDIATRICIANS BY PHYSICIAN ASSISTANTS[a]
Pediatric Service

Category of Complaint	Patient Contact Units Required Per Visit	Weekly Total Patient Contact Units for Physicians Only Delivery Mode	Physician Assistants			
			1st	2nd	3rd	4th
Chronic illness	2	59				*6* 45 29
Acute illness	1	95		*12* 36 28	*10* 20 15	*3* 8 5
	2	213		55 43	53 36	11 7
	3	89		13 9	32 20	8 5
Subacute illness	1	35				*5* 21 12

	2	59				37 20
Miscellaneous	1	53				30 $\boxed{4}$ 19
Checkup	1	154	130 $\boxed{20}$ 115	45 $\boxed{5}$ 39		
	2	35	30 26	11 9		
Trauma	1	35			40 $\boxed{6}$ 27	
	2	12			15 9	
Total[b]		839				
Net saving in physician time units			121	111	91	79

[a]120 patient contact units take half-time for one pediatrician.

[b]Total of 839 differs from 120 × 7 = 840 due to rounding.

Length of visit is indexed in the far left numerical column of the table. To deliver equivalent care, the physician would require 46 and 19 units, as indicated by the lower right-hand numbers in the physician assistant columns. In delivering these 2- and 3-unit chronic care visits, the PA imposes an additional 5-unit supervisory load on the internist, as indicated by the upper right-hand numbers in italics. Thus, the 55 patient contact units of the PA, together with a fraction of the five additional supervisory units imposed upon physicians, replace the 46 patient contact units using the physicians-only mode of delivery. After the second PA is brought into the center, there will still be 208 = 394 - (140 + 46) units of 2-unit chronic care still left to be delivered directly by the internist.

Note that these numbers are marginal representations; they compute the net changes from the situation when one less PA is working in the center. Formulating in a marginal framework avoids the complications that ensue if the first-assigned PA duties must be reassigned when later PAs join the delivery team. In the not unlikely situation where changes in PA numbers would require extensive reassignments, our marginal tables are not an appropriate representation of the specific tasks undertaken by particular PAs.

First, observe that in all cases the PA takes more time than the physician. Only infrequently will it be best for the PA to attempt to duplicate the exact approach the physician would take. To compensate for the physician's superior skill and judgment, the PA will have to substitute a more time-consuming approach to medical care delivery.

For the most part, the more straightforward the task, the smaller the relative time payoff to the greater knowledge and capability of the physician. Height, weight, pulse, blood pressure—these are things that can be measured as quickly by the PA as by the MD. Any semblance of parity vanishes with complex tasks, even though they may be delegable. Cardiac assistants have been trained to auscultate the heart (listen to and define heart sounds and murmurs). On average, they may require more than twice as much time as an internist to perform this sophisticated procedure, part of any comprehensive cardiac evaluation.

The grouping of PA assignments within one or, at most, a few related categories reflects three economic considerations:[16] (1) There may be learning by doing within a category of tasks. (2) Other things equal, supervisory time requirements per unit of care delivered decline as more of a particular type of care is delivered. (3) It may be overly time-consuming, and thus relatively inefficient, to train a PA to handle a great variety of tasks. If a single PA is working in a category that offers more than one PA's-worth of delegable duties, then, subject to the limitations of easy triage, he will take on the most elementary tasks—the ones where his relative efficiency is highest.

16 Maintenance of PA morale provides an argument against grouping tasks, and for sharing of assignments.

A number of factors underlie the details of the assignments made in the tables. As mentioned earlier, we confined our attention to the most highly trained group of assistants. We allowed them to assume solely clinical tasks.[17] Substitution does not extend to paperwork, administration, etc. The constraints that have established the physician as the unsubstitutable deliverer of care do not apply to his administrative duties; they might already have been delegated where feasible.

Though the PAs in the table are assumed to be in the highest skill category, their backgrounds, training, and types of task assignments may differ. With adult medicine, for example, the fourth assistant is a PA with a difference. He is an ex-Army corpsman with extensive experience in triage. He is assigned to the admissions area of the clinic where he effectively screens patients and takes those he can handle. In the acute illness area, he deals with upper respiratory infections, tonsillitis, and the common cold. He covers the more routine cases in the miscellaneous and subacute categories. Minor trauma is a specialty for him. He sutures uncomplicated wounds, sets simple fractures, and administers tetanus prophylaxis.

Similarly, within the pediatric service, the different PAs will have different preparations and roles. There will be a shifting around of early PAs as more are assigned to the center. The most general key observation to be made is that additional PAs offer declining marginal productivity in taking over physicians' tasks.

VI. ESTIMATING THE PRODUCTION FUNCTION

The first major difficulty in estimating the production function is measuring medical output. We use a surrogate measure: the number of patient visits treated. Not all complaints will require that the same amounts of resources be devoted to their treatment. To standardize, we employ the mix of cases observed at the urban health center; the weighted average is thus derived from the percentages derived from Tables 3 and 4.

Table 5 presents the relevant numbers derived from Tables 3 and 4. The two sets of rows in the table are divided into four pairs of numbers. The first two rows contain input pairs that produce the same level of output for adult medicine that was achieved with the old technology when eight physicians worked alone. To state this concept in parallel words, these rows contain input pairs P,A that lie on the isoquant defined by $f(P,A) = g(8)$. Consider the leftmost

17 Some of these proposed assignments must be contingent on consumer acceptance. Experience may show that intimate tasks, though simple to perform, cannot be delegated to PAs in all circumstances. This might also be true for otherwise delegable tasks intimately connected with vital functions. For example, it may be inappropriate for a PA to do a spinal tap.

TABLE 5

	Old Delivery Mode	New Delivery Mode					
Adult medicine	P = 8	P	=	7.5	7.01	6.62	6.27
		A	=	1.0	2.00	3.00	4.00
Pediatric service	P = 7	P	=	6.5	6.03	5.65	5.32
		A	=	1.0	2.00	3.00	4.00

pair. It shows that the introduction of the first PA saves one-half of a physician. This figure is derived from the bottom of Table 3. Reference there shows that the first physician assistant saves 120 physician time-units, half of the 240 delivered by a full-time physician. The bottom two rows show PA substitution possibilities in pediatric medicine.

The data provided by Table 5 put us in a position to estimate the production function for medical output. Admittedly, we have data on only two isoquants, each in a particular setting. Caution should be employed when extrapolating these results to significantly different production modes.

The data in Table 5 exactly fit the production function

$$c[eP^d + (1 - e)A^d]^{1/d}$$

where e takes the value .78 and d is equal to .66. The multiplier c is assigned the value 1.3937. This converts output levels, using the new production mode, to appropriate values vis-à-vis the old; that is, to eight physicians for adult medicine, seven for pediatrics.

This is a constant elasticity of substitution production function. Its form was originally derived to allow for "the properties of (i) homogeneity, (ii) constant elasticity of substitution between capital and labor, (iii) the possibility of different elasticities for different industries" [3, pp. 225–26]. The well-known fixed-coefficient and Cobb-Douglas production functions are special cases with elasticities of substitution of zero and one, respectively. Straight line isoquants imply an elasticity of infinity. The production function here has an elasticity of substitution of 2.94.

Note that this large value for the elasticity implies that the isoquants cut through both the physician and physician assistant axes.[18] We already knew about the first intersection, for in the old delivery mode physicians produce care

18 For example, with this production function, a single physician will require 37.09 physician assistants before he can produce the same output as eight physicians working alone. Solving for zero physicians, we find that 56.92 physician assistants would be required. These numbers are so far out of range of possible physician assistant/physician ratios that it does not seem important to worry about whether physician assistants alone could deliver medical care as the production function would imply. To avoid any possible difficulties, we can specify that our production function only applies for ratios of PAs to physicians less than, say, one to one. Beyond that, it may be inappropriate to

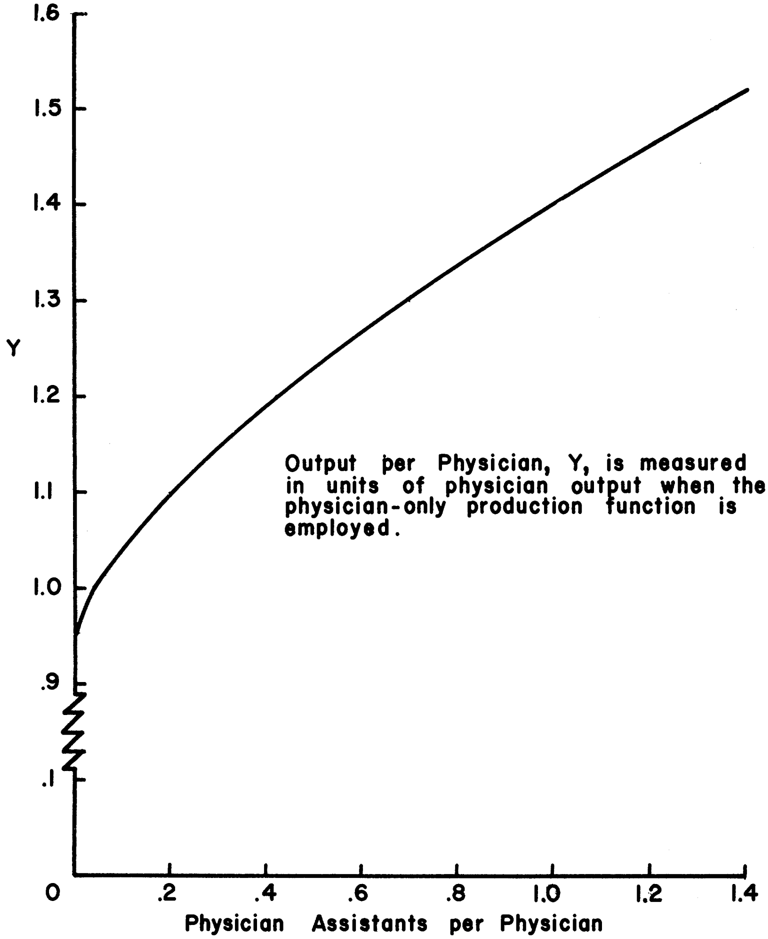

FIGURE 1
PRODUCTION USING PHYSICIANS AND PHYSICIAN ASSISTANTS

by themselves. Physician assistants would not have this solo capability. This would imply that as the number of physician assistants increased beyond the range of relevant values, the elasticity of substitution would necessarily diminish.

VII. APPLYING THE PRODUCTION FUNCTION

Figure 1 displays the values of $f(P,A)$ as a function of A. The figure has been normalized so that PAs are combined with a single physician (which may be below the efficient scale of operation). Observe that with very small fractions of PAs, the output levels with the new delivery mode are below what they would

apply a function with a lower elasticity of substitution. We are indebted to a referee for urging us to pay attention to the axis-cutting problem.

be with the old, despite the fact that the latter has no PAs. This loss is explained by substantial startup supervisory costs that are incurred once any PAs are introduced.

The objective is to allocate physicians between the new and old delivery modes to maximize total output, $O + N$. If the dictates of efficiency are followed in making this allocation, total output can be indicated as a function of the total levels of physicians and assistants—of P_T and A_T, respectively. This relationship is expressed as

$$h(P_T, A_T) = \max_{P_N} f(P_N, A_T) + g(P_T - P_N)$$

For the estimated production function, the ratio of P_N to A_T that maximizes total output is 5.6 to 1.[19] If the number of physician assistants were to grow so that this ratio could not be maintained wherever PAs were in use, maximum output would be achieved by assigning all physicians to the new delivery mode and spreading physician assistants evenly among them.

Financial Implications of the Production Function

In the present situation where PAs are in very short supply, they will be able to function in conjunction with a sufficient number of physicians to enable them to produce their maximal marginal product. Each PA will be replacing roughly 50 percent of a physician. Even though there will be significant restrictions on the areas in which PAs can perform, and even though the tradition of low salaries of allied health manpower may exert pressures to prevent them from reaping their full economic value, we should still expect them to earn substantial salaries.

If the market were allowed to function in an unrestricted fashion, as long as the number of physicians was more than 5.6 times the number of PAs, the salary of the PA would come to half that of the physician. There is very little evidence to confirm or refute this prediction, but the following information is at least suggestive. In the physician assistant programs at Duke University, "this year's average starting salary has risen to $13,500 (with one promise of $18,000 after one year). . . ."[20]

19 This optimal ratio can be computed graphically in Figure 1. Starting at the point 0,1 (output using the old delivery mode), draw a straight line to the total product curve. The slope of this line gives the average and marginal products of the PA. This slope is maximized if the line is drawn tangent to the curve. With the optimal ratio, PA productivity is 50.74 percent that of the physician. This optimal ratio is sensitive to minor changes in the parameters of the production function. Fortunately, however, levels of output and PA productivity are comparatively insensitive to variation in the ratios employed. For efficiency purposes, it is not too important that an exact optimal ratio be computed and achieved.

20 Louis R. Pondy, Duke University, private communication (Aug. 1971).

These are quite substantial salaries to be received by individuals with the initial skills and extent of training of PAs—salaries well in excess of their opportunity costs. However, these salaries are in rough accord with what the production function would have suggested: roughly one-half of the net incomes of physicians working in similar delivery modes. If, as seems likely, physician assistants can reap their productive proportion of the scarcity rents of physicians, the gains to society from their use will come in real resources but not in financial savings. In the long run, the use of physician assistants can help shift outward the supply curve of medical capability. Such a shift will offer financial savings through a reduction in scarcity rents.

VIII. SUMMARY AND POLICY CONCLUSIONS

In the coming years, policy innovations in health care are expected to place tremendous new pressures on the demand side of the health care delivery market. The physician assistant may have a potentially useful role to play in helping to meet this demand. This paper developed a methodology to estimate the contribution of physician assistants in delivering medical care. With limited and conjectural data, it showed that this contribution could be significant.

The analysis here was carried out at a low level of aggregation. What can a PA do to increase medical delivery in a single type of delivery institution? What specific duties could he assume? And what would be the net savings in physician time? From these data we developed a production function for medical output using both physicians and physician assistants. With the optimal mix of factor inputs and the efficient distribution of task assignments, a PA achieves the productivity of half a physician.

REFERENCES

1. I. Altman et al. "The Office Practice of Internists, Part II, 'Patient Load.' " *Journal of the American Medical Association* 193 (1965).
2. American Medical Association. *Reference Data on the Profile of Medical Practice.* Chicago: The Association, 1971.
3. K. Arrow et al. "Capital-Labor Substitution and Economic Efficiency." *Review of Economics and Statistics* 43 (Aug. 1961).
4. A. B. Bergman. "Time-Motion Study of Practicing Pediatricians." *Pediatrics* 38 (1966).
5. J. S. Chase and W. R. Craig. "Time and Motion Study of General Practice." Unpublished paper, Harvard Medical School, 1968.
6. R. W. Deisher et al. "Changing Trends in Pediatric Practice." *Pediatrics* 25 (April 1960).

7. National Academy of Sciences. "New Members of the Physician's Health Team: Physician's Assistants." Washington: The Academy, 1970.
8. C. Theodore and G. Sutter. "A Report on the First Periodic Survey of Physicians." *Journal of the American Medical Association* 202 (Nov. 6, 1972).
9. L. A. Wood. "A Time and Motion Study." *Journal of College of General Practice* 5 (Aug. 1962): 379–81.
10. Richard Zeckhauser and Michael Eliastam. "The Physician Assistant." Teaching and Research Materials No. 8, Kennedy School of Government, Harvard University, April 1972.