

The Origins, History, and Design of the Resident Match

Alvin E. Roth, PhD

THE YEAR 2002 MARKED 50 YEARS of the resident match. Much has been learned, both about the kinds of problems the market for interns experienced in the first half of the 20th century and about the operation of clearinghouses like the resident match.

A BRIEF HISTORY OF THE US INTERNSHIP MARKET

The chief symptom that something was amiss in the early market for interns was that hospitals began to try to hire interns earlier than their competitors, so medical students often could only consider offers from one hospital at a time, without knowing their prospects at other hospitals. The situation in the 1920s is conveyed in a letter from the dean of the Columbia College of Physicians and Surgeons¹:

For a number of years attempts have been made to defer the appointment of hospital internes until towards the close of the fourth year. The Association of American Medical Colleges, the Council on Education of the American Medical Association, and the American Hospital Association have all passed resolutions favoring this idea. The difficulty has been in persuading someone to take the lead.

This is to inform you that it has been decided to defer the appointments of internes at the Presbyterian Hospital in the City of New York until some time in April.

It is earnestly hoped that other hospitals and schools will be able to act in a similar manner.

That hope was in vain. A decade later, a survey of hospitals by Reginald Fitz² at Boston University found appoint-

In the early 1900s, competition among hospitals for interns and among medical students for good internships led to increasingly early offers of internships to students. By the 1940s, appointments were often made as early as the beginning of the junior year of medical school. Hospitals thus had little information about students' performance, and students frequently had to make a final decision to accept or reject an offer without knowing which other offers might be forthcoming. From 1945 through 1951, efforts were made to enforce a uniform date for accepting offers. However, students were still faced with offers having very short deadlines, compelling them to accept or reject offers without knowing what other offers might be forthcoming. Hospitals often had to scramble for available students, since if an offer was rejected, it was often too late for them to reach their next preferred candidate. A centralized clearinghouse was thus developed as a way of alleviating this chaos and allowing a larger role to the preferences of both students and hospitals. This evolved into the current matching program, whose algorithm continues to be updated to take account of changing needs of applicants, such as growth in the number of couples who seek 2 positions in the same vicinity.

JAMA. 2003;289:909-912

www.jama.com

ments spread over the first semester of the students' senior year. He goes on to say:

Nearly a year ago the third year classes of the Harvard Medical School and of Tufts Medical College wrote letters to the Boston Committee suggesting, in effect, that it would be highly desirable from these students' viewpoint if some arrangement could be established by which intern appointments could be made in various hospitals at about the same time. . . . As one student put it, there are very few men who have the conceit to pass up a very good appointment in one locality offered early simply on the gamble of competing for a somewhat more desirable appointment made later in another locality.

The problem worsened until, in 1945, there was an attempt to establish and enforce a uniform time for intern appointments. In proposing the new plan, Joseph Turner summarized

the current state of internship appointments³:

Twenty-five and more years ago, the selection of internes by most hospitals took place in the last half and even the last quarter of the senior year. That selection has now been advanced on the school calendar to the beginning of the junior year and, indeed, inquiries now come to me even from sophomores. The dates of examinations and selection have been pushed farther and far-

Author Affiliation: Department of Economics, Harvard University, Cambridge, Mass.

Financial Disclosure: Dr Roth has served as a paid consultant to the National Resident Matching Program.

Corresponding Author and Reprints: Alvin E. Roth, PhD, Department of Economics, Harvard University, 183 Baker Library, Boston, MA 02163 (e-mail: al_roth@harvard.edu).

On Call Section Editors: Joseph K. Lim, MD, and Stephen J. Lurie, MD, PhD; Editors: Ethan M. Basch, MD, R. Sonia Batra, MD, MPH, Natalie Holt, MD, Alison J. Huang, MPhil, MD, Nina Kim, MD, Vincent Lo Re, MD, Dena E. Rifkin, MD, and Mrugeshkumar K. Shah, MD, MPH.

See also p 913.

ther back, through the efforts of some hospitals to get ahead of others in the choice of candidates, for hospitals can exercise pressure on the selected candidates by requiring acceptance of offers of internship at once or within a short time. The student's dilemma is understandable; if the first offer of this kind comes from a hospital of his second or third choice, he loses out entirely if he declines and is not selected later by the hospital of his first choice.

In response to this situation, it was proposed that medical schools would not release information about students before an announced date.³ This "Cooperative Plan," adopted by the Association of American Medical Colleges, achieved some uniformity in appointment times: appointments for 1946 internships were largely made in the summer of 1945, and in subsequent years the dates at which information was released by medical schools was moved later into the senior year, and the date at which offers were made followed in step.

However, over the next few years, students had to make increasingly prompt decisions. In 1945, offers were to remain open for 10 days. By 1949, a deadline of 12 hours was rejected as too long. Hospitals were finding that if an offer was rejected after even a brief period of consideration, it was often too late to reach their next most preferred candidates before they had accepted other offers. Hospitals thus often pressured students to reply immediately; offers conveyed by telegram were often followed by telephone calls requesting an immediate reply.

The establishment of a clearinghouse, along the lines of what became the resident match, was proposed as a way of continuing to reap the benefits of uniform appointment dates while relieving pressure and congestion near the deadline. In his preface to the proposal for a clearinghouse, F. J. Mullin, dean of students at the University of Chicago School of Medicine, described the shortcomings of the Cooperative Plan⁴:

The most frequently voiced objections to the present Cooperative Plan are the following: Even when telegrams are filed early, the

offices cannot really release them all at once and the distribution gives much unfairness and inequality. . . . Many hospitals have resorted to phoning the students directly and putting pressure on students to make immediate decisions over the phone. . . . Students sometimes get panicky and accept poor internships way down on their lists because they have not heard from a higher position on their order of preference. . . . Students have resented pressure for immediate decisions put on them by phone communication from hospitals. Some hospitals have felt that other hospitals have violated the principles of the Cooperative Plan and have notified students early or have put undue pressure on students for immediate decisions.

Mullin outlined how a clearinghouse would work: rank-order lists would be solicited from students and hospitals, and used to produce a match.⁴ (Clearinghouses had earlier been tried on a regional level, eg, in a Philadelphia Pool Plan⁵ and a Boston Pool Plan [letter from William Castle to Reginald Fitz of Harvard Medical School, September 28, 1951, courtesy of N. C. Webb, MD].) Mullin further noted:

It should be made clear that under the proposed modification of the Cooperative Plan hospitals and students would still be completely free in making contacts and getting information about each other and in expressing their choice in selection of placement and applicants. The proposal calls for the establishment of a central clearing agency to act only as a mechanical facilitation in the final step in the final process of intern selection.

He concluded:

At the annual meeting of the Association of American Medical Colleges in October, 1950, this plan was discussed and it was voted to make a trial run for the present year without influencing the procedures already in effect.

Following the trial run, it was resolved that, for 1952 internships, a centralized match would be used to finalize internship appointments. Mullin and Stalnaker,⁶ announcing the centralized match, summarized its anticipated benefits as follows:

Under the plan the student will not be required to make a decision on the basis of a telephone call or within a very limited period of time. A last minute scramble, with its many uncertainties, is eliminated. No stu-

dent, under the plan, will receive telegraphic offers by a number of hospitals and wonder if he will receive other offers later. Hospitals will not send out telegraphic offers to many students only to receive no replies or negative ones, thus requiring them to send out additional offers at a later time to students who may, in the meantime, have taken another internship although they preferred the hospital involved.

This plan was implemented, with one crucial change. The algorithm outlined by Mullin and Stalnaker for turning rank-order lists into appointments met with objections from students. W. Hardy Hendren, then a student at Harvard Medical School, recounts how, after learning of the proposed algorithm, he organized the National Student Internship Committee, which proposed a different algorithm (oral communication). Hendren and his fellow students noted that under the originally proposed algorithm, a student could suffer by submitting a rank-order list that listed as first choice a position he or she was unlikely to obtain. This is worth describing in more detail, since the choice of matching algorithm had a large effect on the operation of the match.

ALGORITHMS FOR MATCHING INTERNS AND RESIDENTS

Mullin and Stalnaker⁶ had described a clearinghouse in which students ranked individual hospitals and hospitals ranked students in groups, with "1" being reserved for the most preferred students up to the number of available positions, "2" for the next most preferred group, and so forth. The proposed algorithm first matched all hospitals and students that were each others' first choices (1-1 rankings). Then, hospitals would be matched with students in their second group if those students had ranked the hospital first (2-1 rankings), followed by matches of hospitals' first choices with students' second choices (1-2 rankings), and so on (2-2, 3-1, 3-2, 1-3, 2-3 . . .). The intention appears to have been to give an advantage to students, since when preferences conflicted, students' first choices were considered earlier than hospitals' first choices.

However, consider a student who listed as his first choice a hospital to which he did not match, but whose second-choice hospital included him among its first choices. That hospital might fill all its positions in the 1-1 and 2-1 steps of the algorithm and have no position available for the student, a 1-2 match. Thus, it was possible for a student to suffer by ranking first a hospital to which he or she could not match; the student could end up at a hospital he liked very little, even though his second-choice hospital had ranked him first. Mullin and Stalnaker⁷ discussed this possibility:

... There was dissatisfaction caused by the student fear of being penalized for taking a "flyer." Following a meeting in New York, an ad hoc student committee made proposals involving a change in the procedure of matching which was supported by other student groups. The National Interassociation Committee, after consideration of the suggested changes known as the Boston Pool Modification, adopted them as the official method to be used in the matching.

The Boston Pool algorithm updated rank-order lists as it went along, tentatively matching students to hospitals that presently ranked them in the first group and deleting a student from a hospital's list only when that student was tentatively matched to a hospital the student preferred (at which point initially lower-ranked students could move into the first group on the hospital's list). Thus, a student who ranked first a hospital to which he or she did not match could nevertheless be assured that if the second-choice hospital did not fill with students it preferred, he or she would get a position there. This Boston Pool algorithm is equivalent to a "deferred acceptance" algorithm, which can be interpreted as one in which hospitals make offers to applicants, starting at the top of each hospital's rank-order list, and each applicant holds on to the best offer he or she has received so far but can later reject it if a better offer is forthcoming.⁸

The change in algorithms was fortunate for the longevity of the match, which became the National Resident

Matching Program (NRMP), because the Boston Pool algorithm had another property the Mullin and Stalnaker⁶ algorithm lacked. It produced outcomes that were *stable*, in the sense that no applicant and hospital who were not matched with one another preferred each other to their assigned matches.

The importance of stability has since become clear. If an algorithm produces unstable outcomes, then there are applicants and hospitals who would both prefer to be matched to one another than to accept the results of the match (as in the example described earlier). This creates mutual incentives for the unhappy pairs to circumvent the match.

For example, when the British market for interns experienced increasingly early appointments in the 1960s, each region of the British National Health Service devised its own centralized clearinghouse. Several used algorithms very similar to that of Mullin and Stalnaker. These all failed and were abandoned after interested applicants and hospitals learned to circumvent them. In contrast, clearinghouses that produced stable outcomes succeeded and remained in use.^{9,10} This and related evidence strongly suggest that, had the originally proposed match algorithm not been replaced, we would not now be looking back on 50 years of operation of the NRMP.

In the 50 years since the inception of the match, changes in medicine have been reflected in the demands on the match and in the design of the algorithm. Fifty years ago, the vast majority of US medical graduates were men; today, the match accommodates couples who may submit rank-order lists of pairs of positions, to obtain jobs together. Fifty years ago, most internships were nonspecialized rotations; today, many specialties require more than 1 residency, and the match accommodates applicants who need to combine 2 positions by allowing them to submit a primary rank-order list for (typically) second-year positions and supplemental lists for first-year positions to

be applied if a second-year position is obtained.

The object of the most recent redesign of the algorithm, used since 1998, was that it should yield a match as favorable as possible to applicants while producing a stable outcome that accommodated these contemporary requirements.^{11,12} The current (Roth-Peranson) algorithm is a deferred acceptance algorithm that can be viewed as a process in which applicants offer to come to residency programs, starting with the residency program at the top of the applicant's rank-order list, rather than one in which the programs make offers to applicants, starting at the top of the program's rank-order list.

Today, there are also matches for fellowship positions, many initiated in the 1980s and 1990s, after fellowship markets exhibited the unraveling of appointment dates that characterized the intern market before the resident match. So, 50 years after the inception of the resident match—years that have included significant changes and some controversy—it continues to serve the function for which it was designed.

This article has focused on the match process, which has occurred in the context of larger changes in the organization of US medical education. Historical background information has been provided by Starr¹³ and Ludmerer.¹⁴ Gale and Shapley¹⁵ and Roth and Sotomayor¹⁶ provide further reading on stability and matching. Roth and Xing¹⁷ describe other markets that have experienced unraveling of transaction times; see also my more formal contemporary overview.¹⁸

Funding/Support: Parts of this work were supported by a grant from the National Science Foundation.

Acknowledgment: I received helpful comments on an earlier draft of the manuscript from Robert Beran, Muriel Niederle, Elliott Peranson, and N. C. Webb, MD.

REFERENCES

1. Darrach W. Letter. *Bull Assoc Am Med Coll*. 1927; 2:68.
2. Fitz R. Concerning interns and internships. *J Assoc Am Med Coll*. 1939;14:101-104.

3. Turner J. Intern selection: wanted, an orderly plan. *J Assoc Am Med Coll.* 1945;20:26-32.
4. Mullin FJ. A proposal for supplementing the Co-operative Plan for appointment of interns. *J Assoc Am Med Coll.* 1950;25:437-442.
5. Hatfield JN. How Philadelphia hospitals appoint internes. *Natl Hosp Forum.* February 1935.
6. Mullin FJ, Stalnaker JM. The matching plan of internship appointment. *J Med Educ.* 1951;26:341-345.
7. Mullin FJ, Stalnaker JM. The matching plan for internship placement: a report of the first year's experience. *J Med Educ.* 1952;27:193-200.
8. Roth AE. The evolution of the labor market for medical interns and residents: a case study in game theory. *J Political Economy.* 1984;92:991-1016.
9. Roth AE. New physicians: a natural experiment in market organization. *Science.* 1990;250:1524-1528.
10. Roth AE. A natural experiment in the organization of entry level labor markets: regional markets for new physicians and surgeons in the UK. *Am Econ Rev.* 1991;81:415-440.
11. Roth AE, Peranson E. The effects of the change in the NRMP matching algorithm. *JAMA.* 1997;278:729-732.
12. Roth AE, Peranson E. The redesign of the matching market for American physicians: some engineering aspects of economic design. *Am Econ Rev.* 1999;89:748-780.
13. Starr P. *The Social Transformation of American Medicine.* New York, NY: Basic Books Inc; 1982.
14. Ludmerer KM. *Time to Heal: American Medical Education from the Turn of the Century to the Era of Managed Care.* New York, NY: Oxford University Press; 1999.
15. Gale D, Shapley L. College admissions and the stability of marriage. *Am Math Monthly.* 1962;69:9-15.
16. Roth AE, Sotomayor M. *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis.* New York, NY: Cambridge University Press; 1990.
17. Roth A, Xing X. Jumping the gun: imperfections and institutions related to the timing of market transactions. *Am Econ Rev.* 1994;84:992-1044.
18. Roth AE. The economist as engineer: game theory, experimental economics and computation as tools of design economics. *Econometrica.* 2002;70:1341-1378.

It is art that *makes* life, makes interest, makes importance . . . and I know of no substitute whatever for the force and beauty of its process.
—Henry James (1843-1916)