

Evaluación de la calidad de la atención médica*

Avedis Donabedian

Introducción

El propósito de este trabajo es describir y evaluar los métodos que se usan actualmente para estimar el grado de calidad de la atención médica, así como proponer algunas orientaciones destinadas a estudios futuros. El trabajo se ocupa de los métodos antes que de los resultados, y de la evaluación de la metodología en general antes que de una crítica detallada de los métodos empleados en determinados estudios.

No se ha querido hacer un examen exhaustivo de la bibliografía sobre el tema. Se han incluido, por cierto, los estudios más importantes; otros se han seleccionado tan sólo como ejemplos ilustrativos. Los omitidos no son por ello menos dignos de mención.

Este trabajo está dedicado casi exclusivamente a la evaluación del proceso de atención médica desde el punto de vista de la interacción entre el médico y el paciente. Por consiguiente, se han excluido los procesos relacionados fundamentalmente con la provisión efectiva de atención médica en el nivel de la comunidad. Tampoco se refiere este análisis a los aspectos administrativos del control de la calidad. Muchos de los estudios que se analizan aquí se originaron en la necesidad imperiosa de evaluar y controlar la calidad de la atención en los programas de atención médica ya establecidos. Sin embargo, estos estudios se examinarán sólo en términos de su contribución a los métodos de evaluación y no en función de sus objetivos sociales más amplios. El autor se ha mantenido, en general, dentro del territorio conocido de la atención que proporcionan los médicos y ha evitado incursionar en otros tipos de atención de la salud. Tampoco se ha considerado el difícil problema de la eficiencia económica como dimensión mensurable de la calidad.

Tres estudios de índole general sobre la evaluación de la calidad han sido sumamente útiles en la preparación de esta reseña. El primero es una obra clásica en la que el autor Mindel Sheps hace un excelente examen de los distintos métodos¹. Un ensayo más reciente de Peterson brinda una valiosa evaluación de este tema². En el informe de Lerner y Riedel se analiza un trabajo reciente sobre la calidad y se plantean varias cuestiones de importancia general³.

Definiciones de la calidad

La evaluación de la calidad debe apoyarse en una definición conceptual y operativa de lo que significa la "calidad de la

atención médica". En este aspecto fundamental se presentan muchos problemas, pues la calidad de la atención es una idea extraordinariamente difícil de definir. Tal vez la definición más conocida sea la de Lee y Jones⁴ en forma de ocho "artículos de fe", algunos enunciados como atributos o propiedades del proceso de la prestación de atención y otros, como metas u objetivos de ese proceso. Estos "artículos" transmiten claramente la impresión de que los criterios de calidad no son nada más que juicios de valor que se aplican a distintos aspectos, propiedades, componentes o alcances de un proceso denominado atención médica. En ese sentido, la definición de calidad puede ser casi cualquier cosa que se quiera que sea, si bien por lo común es un reflejo de los valores y metas vigentes en el sistema de atención médica y en la sociedad más amplia de la que ésta es una parte.

Pocos estudios empíricos analizan en profundidad cuáles son las dimensiones y los valores pertinentes en un momento y en un entorno determinados. Klein *et al.*⁵ encontraron que 24 "funcionarios administrativos" aplicaban, en conjunto, 80 criterios para evaluar la "atención del paciente". Los autores llegaron a la conclusión de que la atención del paciente, al igual que la moral, no se puede considerar como un concepto unitario y "...al parecer es probable que nunca se llegue a un único criterio general para medir la calidad de la atención del paciente".

Las dimensiones y los criterios que se elijan entre los muchísimos posibles para definir la calidad tendrán, por cierto, profunda influencia en los enfoques y métodos que se empleen para evaluar la atención médica.

Modos de enfocar la evaluación: qué es lo que se debe evaluar

El resultado de la atención médica, en términos de recuperación, restauración de las funciones y supervivencia, se ha utilizado frecuentemente como indicador de la calidad de esa atención. Como ejemplos se pueden citar los estudios de la mortalidad perinatal^{6,7}, las tasas de letalidad en la cirugía⁸ y la adaptación social de los pacientes dados de alta de los hospitales psiquiátricos⁹.

El uso de los resultados como criterio para estimar la calidad de la atención médica ofrece innumerables ventajas. Pocas veces se cuestiona la validez de utilizar los resultados

*Donabedian A. Evaluating the quality of medical care. *Milbank Mem Fund Q* 1966;44:166-203

Reproducción autorizada por los propietarios del copyright para la traducción española en homenaje a la figura del profesor Donabedian. Versión española supervisada por Fundación Avedis Donabedian y Revista de Calidad Asistencial

como dimensión de la calidad. Tampoco la estabilidad y validez de los valores de recuperación, restauración y supervivencia plantean ninguna duda en la mayoría de las situaciones y culturas, aunque tal vez no sea así en todas. Por otra parte, los resultados tienden a ser bastante concretos y, como tales, sujetos aparentemente a una medición más precisa.

Sin embargo, algunas consideraciones restringen el uso de los resultados para medir la calidad de la atención. En primer lugar, habría que determinar si el resultado de la atención es, en realidad, la medida que corresponde utilizar, puesto que los resultados reflejan no sólo el poder de la ciencia médica de alcanzar determinados efectos en ciertas condiciones dadas, sino también el grado en que se ha aplicado, en los casos objeto de estudio la, "medicina científica" como se la concibe actualmente. Pero, precisamente, el objetivo puede ser separar estos dos efectos. En algunas ocasiones, un determinado resultado puede no ser pertinente, por ejemplo, cuando se elige la supervivencia como criterio de éxito en una situación que no es mortal, pero que puede conducir a una condición de salud subóptima o invalidante¹⁰.

Aun en las situaciones en que los resultados son pertinentes y en que se ha elegido como criterio el resultado que corresponde, hay que tener en cuenta las limitaciones. En el resultado pueden influir muchos otros factores además de la atención médica, por lo que, si se quiere llegar a conclusiones válidas, habrá que tener cuidado para mantener constantes todos los demás factores distintos de la atención médica. En algunos casos, es preciso que transcurran largos períodos de tiempo, tal vez decenios, antes de que los resultados pertinentes sean evidentes. En esas ocasiones, los resultados no están disponibles cuando se necesitan para la evaluación y los problemas derivados de la necesidad de mantener la comparabilidad se acrecientan notablemente. Además, la tecnología médica no es totalmente efectiva y con frecuencia no se conoce con precisión el índice de éxitos que es dable esperar en una situación determinada. Por esta razón se deben hacer estudios comparativos de los resultados en situaciones controladas.

Si bien algunos resultados son por lo general fáciles de medir sin lugar a error (la muerte, por ejemplo), la medición de otros que no fueron definidos con tanta precisión puede resultar difícil. Entre estos se incluyen las actitudes y satisfacciones de los pacientes, la readaptación social y la discapacidad y rehabilitación física¹¹. Aun la validez aparente que se atribuye por lo común a los resultados, como criterios de éxito o fracaso, no es absoluta. Es una cuestión debatible, por ejemplo, si la prolongación de la vida en determinadas circunstancias es evidencia de buena atención médica. McDermott *et al.* han demostrado que, si bien la corrección de una luxación congénita de cadera en una posición determinada es considerada una prueba de buena medicina en el hombre de raza blanca, puede resultar invalidante para el indio navajo que pasa gran parte de su tiempo sentado en el suelo o sobre la montura¹². Por último, si bien los resultados considerados en conjunto podrían ser indicativos de buena o mala atención, no permiten conocer en qué consisten las deficiencias o ventajas a las que se podría atribuir un resultado determinado ni dónde se encuentran.

Al presentar todas estas limitaciones al uso de los resultados como criterios para evaluar la atención médica no se ha

querido demostrar que los resultados son inadecuados como indicadores de la calidad, sino subrayar que se los debe utilizar con discernimiento. Los resultados siguen siendo en general los indicadores definitivos para dar validez a la eficacia y calidad de la atención médica.

Otra forma de enfocar la evaluación consiste en examinar el proceso mismo de prestación, en lugar de los resultados. Esta posición se justifica porque se supone que lo que interesa no es el poder de la tecnología médica de alcanzar los resultados previstos, sino determinar si se ha aplicado lo que ahora se conoce como "buena" atención médica. Las estimaciones se fundan en consideraciones como el grado de adecuación, integridad y minuciosidad de la información obtenida mediante la historia clínica, el examen físico y las pruebas diagnósticas; la justificación del diagnóstico y el tratamiento; la competencia técnica en la aplicación de los procedimientos diagnósticos y terapéuticos, incluso los quirúrgicos; las pruebas del manejo preventivo del paciente en condiciones de salud y de enfermedad; la coordinación y continuidad de la atención; el grado en que la atención resulta aceptable para el destinatario y muchas otras. Este enfoque requiere que se especifiquen con gran detenimiento las dimensiones, los valores y las normas que se usarán en la evaluación. Las estimaciones de la calidad obtenidas sobre esta base son menos estables y definitivas que las derivadas de la medición de los resultados. No obstante, pueden ser más pertinentes a la cuestión que nos ocupa: establecer si se practica correctamente la medicina.

Podría parecer que este análisis del proceso y los resultados significa una mera separación entre medios y fines. Tal vez sería más acertado pensar en una cadena ininterrumpida de medios antecedentes seguida de fines intermedios los que, a su vez, son medios para otros fines¹³. La salud misma podría ser un medio para alcanzar aun otro objetivo. Varios autores han señalado que esta formulación ofrece un enfoque valioso para la evaluación^{14,15}, la que podría considerarse como la medición de los puntos finales del procedimiento e incluirse bajo el encabezamiento general de "proceso" porque se funda en consideraciones similares con respecto a los valores, criterios y validación.

Una tercera manera de abordar la evaluación es estudiar no el proceso de atención en sí mismo, sino las situaciones en que tiene lugar y los medios y procedimientos de los que es el producto. Esto podría denominarse *grosso modo* evaluación de la estructura, aunque también puede abarcar procesos administrativos de distinta índole, que prestan apoyo y orientación a la provisión de atención. La evaluación así entendida se basa en aspectos tales como la adecuación de las instalaciones y los equipos; la idoneidad del personal médico y su organización; la estructura administrativa y el funcionamiento de programas e instituciones que prestan atención médica; la organización fiscal y varios más^{16,17}. Se da por sentado que si se cuenta con las condiciones y los medios adecuados, la buena atención médica estará asegurada. Este enfoque tiene la ventaja de que se refiere, al menos en parte, a información bastante concreta y accesible. Sin embargo, presenta el inconveniente importante de que con frecuencia la relación entre la estructura y el proceso o entre la estructura y el resultado no está bien establecida.

Fuentes de información y métodos para obtenerla

El enfoque que se adopte para evaluar la calidad determina en gran parte los métodos que se usarán para recoger la información requerida. Puesto que estos abarcan toda la gama de los métodos de las ciencias sociales, no se intentará hacer la descripción de cada uno de ellos. Hay cuatro, sin embargo, que merecen atención especial.

Las *historias clínicas* son los documentos de los que se parte para la mayoría de los estudios sobre el proceso de atención médica. No se debe olvidar, sin embargo, que adolecen de varias limitaciones. Dado que los consultorios de la mayoría de los médicos de la práctica privada no son fácilmente accesibles para el investigador y que los registros que estos llevan son por lo demás someros, las historias clínicas sólo se utilizan para evaluar la atención prestada en los hospitales, incluso en los departamentos de medicina ambulatoria, y en los planes de seguro médico. Tanto Peterson¹⁸ como Clute¹⁹ se han referido a las insuficiencias más comunes que se observan en los registros de la práctica médica general. Clute ha señalado, además, que en la práctica general "...la carencia de registros adecuados no es incompatible con el ejercicio de una medicina de buena y hasta excelente calidad..." Por otra parte, un estudio reciente sobre los consultorios privados de una muestra de médicos miembros de la Sociedad de Medicina Interna de Nueva York²⁰ revela que los resúmenes de los registros llevados en ellos se pueden utilizar para obtener estimaciones reproducibles de la calidad de la atención. Sin embargo, es difícil generalizar sobre este hallazgo, porque se refiere a un grupo especial de médicos cuyos registros son probablemente más adecuados que los del término medio. Además, por una u otra razón, la muestra original recogida para este estudio sufrió una pérdida de 61%.

Suponiendo que se tenga acceso al registro y que sea razonablemente adecuado, queda aún por establecer hasta qué punto es veraz y completo. Lembcke¹⁰ ha objetado que se puedan aceptar sin más las afirmaciones clave incluidas en el registro. No sólo ha cuestionado las declaraciones del médico sobre el paciente y su manejo, sino también la validez de los informes de los servicios de diagnóstico. Para verificar las afirmaciones del médico, se puede buscar en el registro, incluso en las anotaciones de las enfermeras, lo que al parecer sea la prueba más válida sobre el verdadero estado de cosas. La validez de los informes se confirma con el examen por personas competentes de las pruebas (películas, registros gráficos, diapositivas) en las que se basan los informes diagnósticos. Aun en las mejores circunstancias, el error del observador en la apreciación del material suele ser un problema²¹. Pero nada supera la incredulidad que produce el hallazgo de Lembcke en un hospital donde determinó que, sobre la base de las muestras de material de legrado, la verdadera incidencia de la hiperplasia uterina era entre el 5 y el 8%, en lugar de un 60 a 65%, según lo informado por el patólogo del hospital. En todo caso, los resultados de la verificación considerada como parte de la evaluación de la calidad deben ser objeto de detenido análisis. No cabe duda de que los errores en los informes diagnósticos se reflejan particularmente en la calidad de servicio de diagnóstico y en la atención que presta el hospital en general. Sin embargo, el desempeño del médico puede merecer un juicio favorable, prescindiendo de que los datos con que trabaja sean o no válidos.

Ello es así cuando el objeto de interés al emitir ese juicio son los principios que orientan las actividades del médico y no la validez absoluta de esas actividades.

Mucho se ha debatido la cuestión de cuán completas son las historias clínicas, así como la posibilidad de establecer si, al evaluar la calidad de la atención sobre la base del contenido de esa historia, se está atribuyendo un valor a la historia clínica o a la atención prestada. La confusión se origina porque el registro de la historia clínica es de por sí una dimensión de la calidad de la práctica médica, así como el medio de información que permite evaluar la mayoría de las restantes dimensiones. Estos dos aspectos se pueden separar cuando se dispone de una fuente alternativa de información sobre el proceso de la atención como, por ejemplo, la observación directa de la práctica médica^{18,19}. Rosenfeld²² resolvió el problema de separar el registro de la historia clínica y la atención médica prestada examinando las razones por las cuales se le adjudicaba un valor menor a la calidad de la atención en la historia de cada paciente examinada. Demostró que el hecho de que la calidad de la atención se considerase de menor valor se debía en parte a lo que se podía considerar un registro inadecuado (pruebas "presuntas") y en parte a razones que no eran atribuibles al registro (pruebas "reales"). También encontró que la clasificación adjudicada a los hospitales solía ser la misma, alta o baja, con respecto a ambos tipos de errores, lo que indica la existencia de una correlación entre esos errores. Puesto que lo más probable es que los registros de rutina más completos correspondan a las salas, la comparación entre los servicios privados y los de sala de cada hospital en función de los tipos de razones por las que la calidad de estos servicios recibió una clasificación más baja podría haber brindado mayor información sobre esta importante cuestión. Otros investigadores han intentado subsanar las omisiones en las historias clínicas complementando esos informes mediante entrevistas con el médico a cargo del paciente y efectuando las enmiendas correspondientes^{23,25}. Lamentablemente, solo en uno de esos estudios (duración de la ingreso en los hospitales de Michigan) se incluye un informe sobre las diferencias que originó esta medida. En este estudio, "la información médica adicional recogida mediante entrevistas personales con los médicos de cabecera fue lo suficientemente importante en el 12,6% del total de casos estudiados para que se justificara reclasificar la evaluación realizada respecto a la necesidad de ingresar al paciente o a la adecuación de la duración del ingreso, o respecto a ambas"^{3,25}. Cuando la información obtenida en una entrevista se usa para modificar o completar la historia del paciente, es posible que haya que suponer que esta información adicional tiene la misma o mayor validez que la contenida en la historia. Morehead, cuya experiencia en la aplicación de este método es muy amplia, señaló que: "Muchos de los encuestadores que colaboran en el presente estudio habían aplicado el método de entrevistar al médico en trabajos anteriores sin obtener buenos resultados... El encuestador... se encontraba en la incómoda posición de verse obligado a elegir entre aceptar sin cuestionar las aseveraciones de que la atención médica era realmente óptima, o llegar a la conclusión de que esas afirmaciones eran inexactas"²⁶. En un estudio anterior, en el que se indicaba que se recurría al método de las entrevistas para complementar los datos disponibles²⁴, se desechó la

información verbal a menos que la corroborasen las medidas tomadas o pruebas concretas²⁷.

Otra cuestión metodológica que requiere consideración es establecer si corresponde usar como base para la evaluación la totalidad de la historia clínica o resúmenes de ella. La cuestión se plantea porque se supone que los extractos y resúmenes pueden ser preparados por personas menos especializadas a fin de que el experto, cuyos servicios son difíciles de obtener, pueda concentrarse en la verdadera tarea de la evaluación. Sin embargo, la preparación de resúmenes exige discernimiento para decidir qué es pertinente e importante. Por esta razón, sólo se los ha usado como un primer paso para evaluar la calidad en estudios en los que se utilizan criterios muy específicos y detallados¹⁰. Aun así, es poco lo que se sabe acerca del grado de confiabilidad del proceso de extraer información o qué validez tiene cuando se lo compara con una lectura más experimentada del material. El estudio sobre los especialistas en medicina interna de Nueva York al que ya se ha hecho referencia reveló una gran concordancia entre los resúmenes de historias clínicas de consultorios preparados por los médicos y por personas altamente capacitadas que no pertenecían a esa profesión²⁰.

Si bien continúa la controversia acerca de si la historia clínica debe ser utilizada como fuente de información, algunos autores han intentado disminuir su dependencia de los registros médicos eligiendo para la evaluación categorías diagnósticas que estén respaldadas por comunicaciones escritas además de las anotaciones del médico²⁸. Esto explica, en parte, que se recurra con frecuencia a los informes sobre operaciones quirúrgicas como material para los estudios de la calidad de la atención.

Las historias clínicas de los pacientes atendidos en la práctica general son demasiado imprecisas para que se las pueda utilizar como base para la evaluación. La otra alternativa es la *observación directa* de las actividades del médico por un colega calificado^{18,19}. La principal limitación de este método parecería ser los cambios que probablemente se producirían en la actividad habitual del médico que sabe que es objeto de observación. En contra de esto se ha aducido que a menudo el médico no está enterado del verdadero propósito del estudio, que se acostumbra rápidamente a la presencia del observador y que le resulta imposible modificar hábitos arraigados en el ejercicio de su profesión. Aun cuando ocurran cambios, es probable que resulten en una estimación mayor de la calidad y no en lo contrario. A pesar de estas afirmaciones, la medición de los efectos que la observación puede tener en la práctica médica sigue siendo un problema no resuelto.

A quienes han utilizado el método de la observación no se les ha escapado que con él no se obviaba el problema de establecer cuán completa es la información. El médico con frecuencia sabe bastante acerca del paciente por haber tenido contacto con él en ocasiones anteriores. De ahí que sea necesario elegir para la observación casos "nuevos" y situaciones que exijan hacer un reconocimiento a fondo, sea cual fuere la experiencia anterior que se tenga con el paciente. Por otra parte, no todas las actividades del médico para el manejo del paciente son explícitas. Algunos aspectos de la atención, no sujetos a la observación directa, deben excluirse del esquema de evaluación.

Otro problema que se puede presentar es la percepción selectiva del observador. No es lo habitual que el observador primero registre imparcialmente los hechos y sea después el juez de esos mismos hechos. Es muy posible que sus conocimientos y criterios influyan en su manera de percibir los hechos, lo que puede ocasionar una cierta distorsión de lo percibido.

Uno de los métodos indirectos de obtener información consiste en estudiar *conductas y opiniones* que permitan hacer inferencias sobre la calidad. Maloney *et al.* se han referido a un enfoque *sociométrico*, según el cual, cuando los médicos buscan atención para ellos y sus familias expresan con sentido crítico opiniones válidas sobre la capacidad de sus colegas de brindar atención de alta calidad²⁹. Se comprobó que los médicos que eligieron tenían reputación de estar más calificados que otros. Pero tanto la sensibilidad como la especificidad, cuando se usan como criterio estimaciones más estrictas de la calidad de la atención, carecen de validez. Georgopoulos y Mann³¹ emplearon un método, que se podría denominar *autorreferencial*³¹, para evaluar la calidad de la atención en algunos hospitales comunitarios. Este método surgió de estudios anteriores en los que se comprobó que la gente suele ser bastante perspicaz cuando juzga la "eficacia" de las organizaciones en las que trabaja³². Los hospitales fueron clasificados por grado sobre la base de opiniones acerca de la calidad de la atención médica y otras características emitidas por miembros del personal directivo, profesional y técnico de cada hospital, por personas relacionadas con esas instituciones y por gente bien informada de la comunidad. La perspicacia y coherencia de las opiniones permitió que se pudiesen clasificar los hospitales con un grado de confiabilidad aparentemente satisfactorio, a pesar del carácter elogioso de las respuestas que en el 89% de los casos calificaron la calidad de la atención médica en los hospitales de "muy buena", "excelente" o "sobresaliente" y de "regular" en casi ninguno. Los autores suministran abundantes pruebas de un alto grado de interrelación entre distintas opiniones emitidas en forma separada. Pero en pocos casos se respalda la validez de las opiniones recurriendo a criterios verdaderamente independientes para evaluar la calidad de la atención.

Toma de muestras y selección

Cuando se va a efectuar un muestreo, lo primero que hay que hacer es especificar con precisión el universo que será objeto del muestreo, el cual, a su vez, depende de la clase de generalizaciones que se quiera hacer. Por lo común, los estudios sobre la calidad se refieren a uno de estos tres objetivos: 1. la atención que realmente presta una categoría predeterminada de proveedores de atención; 2. la atención que realmente recibe un grupo determinado de personas, y 3. la capacidad de un grupo determinado de proveedores de proporcionar atención. En los dos primeros casos se necesitan muestras representativas tanto de los posibles proveedores o beneficiarios, como de la atención proporcionada o recibida. En el tercero se requiere una muestra representativa de los proveedores, pero no necesariamente de la atención recibida. Otro aspecto aun más importante es que se seleccionen de manera uniforme aspectos significativos de la atención. Tal vez sea necesario estudiar cómo se cumplen las tareas en ciertas situaciones clínicas que originan

mucha tensión y en las que, por consiguiente, se ponen más de manifiesto la competencia o las debilidades latentes para desempeñarlas. Hasta se pueden crear situaciones hipotéticas para evaluar la capacidad de desempeño en ciertos aspectos de la atención³³⁻³⁵. Las distinciones que se han hecho antes, especialmente entre la evaluación de la atención realmente provista y la capacidad de proporcionar atención, son convenientes para evaluar los procedimientos de muestreo utilizados en los principales estudios sobre la calidad. Según estos criterios, ciertos estudios pertenecen a una determinada categoría, mientras que otros reúnen las características de varias categorías de modo tal que la generalización resulta difícil. Por ejemplo, en el primer estudio sobre la calidad de la atención proporcionada a las familias de los miembros del sindicato de conductores de camión, los resultados se refieren sólo al manejo de determinadas categorías de enfermedad que requieren internación con respecto a un grupo de población determinado²⁸. En el segundo trabajo de esta serie se alcanza un grado de generalización algo mayor obteniendo una muestra representativa (con exclusión de las variaciones estacionales) de todas las enfermedades que requieren internación en el mismo grupo de población²⁶. En ninguno de los dos estudios se prevé suministrar información sobre toda la atención prestada por una muestra representativa de médicos.

El grado de homogeneidad del universo que se va a muestrear tiene, por cierto, gran importancia en todo proyecto de muestreo o selección. La pregunta que corresponde formular aquí es en qué medida es uniforme la atención que presta un médico. ¿Existen determinadas categorías diagnósticas, niveles de dificultad o aspectos de la atención en los que un médico se desempeña mejor que en otros? ¿Es posible encontrar realmente en ese médico una "capacidad general de desempeñarse bien al prestar la atención médica"¹⁸ o, por el contrario, la calidad de su desempeño es bastante dispar? También cabría preguntarse si la atención que prestan las distintas subdivisiones de una misma institución es más o menos del mismo nivel ya sea en términos absolutos o en relación con el desempeño de instituciones comparables. Makover, por ejemplo, da por sentada explícitamente la homogeneidad cuando sostiene: "No se hizo ningún intento de relacionar el número de historias clínicas que se iban a estudiar con el número de asociados a los planes de seguro médico. La atención médica prestada a uno u otro individuo es una prueba válida de la calidad y no debe existir ninguna, o sólo una mínima, variación aleatoria que sea afectada por el ajuste según el tamaño de la muestra"²³. Rosenfeld inició su estudio enunciando la hipótesis de que en las instituciones existe una correspondencia en los niveles de atención de las distintas especialidades y para varias categorías de enfermedades²².

No abundan las pruebas empíricas relacionadas con la homogeneidad. Los estudios de Peterson y Clute sobre la práctica general de la medicina^{18,19} mostraron que había un alto grado de correlación entre el desempeño de los médicos en distintos componentes o aspectos de la atención (historia clínica, reconocimiento físico, tratamiento, etc.). Por su parte, Rosenfeld también demostró que no había grandes diferencias en las clasificaciones de la calidad de varios diagnósticos seleccionados en distintos campos de la práctica médica (medicina

general, cirugía y obstetricia-ginecología). Si bien las diferencias entre los hospitales parecieron ser mayores cuando se consideraron los distintos campos de la práctica médica, no fueron suficientes para modificar las posiciones de los tres hospitales estudiados.

En los dos estudios de la atención recibida por las familias de los miembros del sindicato de conductores de camión^{26,28} se citó un porcentaje de atención óptima y subóptima casi idéntico en todas las poblaciones estudiadas. Esa debe haber sido una coincidencia, ya que el porcentaje de atención óptima, en el segundo estudio, varió mucho según la clase de diagnóstico, de 31% en el caso de la medicina a 100% en el de la oftalmología (nueve casos solamente). Si existe esa variabilidad, el "conjunto de diagnósticos" de la muestra de pacientes atendidos debe ser un asunto de gran importancia en la evaluación. En los dos estudios de las familias de los miembros del sindicato citado, se consideró que las diferencias de ese conjunto habían llevado a asignar una menor clasificación a la medicina y una mayor a la obstetricia y ginecología en el segundo estudio en comparación con el primero. El hecho de que el mismo factor pueda producir efectos en dos sentidos contrarios indica la complejidad de las interacciones que debe considerar el investigador. "La explicación más probable de la menor clasificación asignada a la medicina en el estudio actual (que es el segundo) es la naturaleza de los casos analizados". El factor al que puede atribuirse eso es la menor capacidad para manejar la enfermedad "que no encuadró dentro de un patrón bien reconocido". En el caso de la obstetricia y la ginecología, los resultados del segundo estudio "...difieron en un aspecto importante de los del anterior en que se hicieron serias preguntas sobre el tratamiento de un número de pacientes mucho mayor. El estudio anterior se centró sobre todo en cirugía abdominal mayor, en tanto que el grupo seleccionado al azar comprendió pocos de esos casos y tuvo más pacientes con afecciones menores"²⁶. En estudios como éstos, en que la atención recibida por la totalidad o parte de la población es objeto de estudio, las variaciones observadas emanan, en parte, de diferencias de la materia incluida en el diagnóstico y, en parte, de patrones institucionalizados de práctica relacionados con dicha materia. Por ejemplo, los nueve casos de oftalmopatía recibieron óptima atención porque "ese es un campo muy especializado en que los médicos que carecen de la formación pertinente raras veces se arriesgan a practicar procedimientos"²⁶.

El muestreo y la selección influyen en varios factores, además de la generalización y la homogeneidad, y reciben la influencia de los mismos. Las dimensiones particulares de la atención que despiertan interés (tratamiento preventivo o técnica quirúrgica, para citar dos ejemplos bastante distintos) pueden determinar la selección de situaciones de atención médica para fines de evaluación. Las escogidas guardan relación también con la naturaleza de los criterios y patrones usados y el sistema de clasificación y calificación adoptado. Cualquier intento por tomar muestras de situaciones problemáticas, más bien que de diagnósticos u operaciones tradicionales, se puede dificultar mucho por la forma en que se archivan y clasifican las fichas clínicas. Eso es lamentablemente porque un análisis de las operaciones o de los tipos de diagnóstico establecidos da una idea de en qué se fundamentó el diagnóstico o

se hizo la operación. Eso deja inexplorado un segmento complementario de la práctica, a saber, las situaciones en que puede haberse indicado un diagnóstico o tratamiento similar, pero no se emitió ni se administró.

Patrones de medición

La medición depende de la formulación de patrones. Para fines de evaluación de la calidad, los patrones se derivan de dos fuentes.

Los patrones *empíricos* se derivan de la práctica real y se emplean en general para comparar la atención médica prestada en un ámbito con la de otro o con promedios y escalas de valores estadísticos obtenidos a partir de un mayor número de ámbitos similares. El Estudio de las Actividades Profesionales se basa, en parte, en ese método³⁶.

Los patrones empíricos se basan en niveles de atención cuyo logro puede demostrarse y, por esa razón, gozan de un cierto grado de credibilidad y aceptabilidad. Además, sin patrones normativos claros, deben formularse observaciones empíricas en ámbitos seleccionados con esa finalidad. Un interesante ejemplo proviene del trabajo de Furstenberg *et al.*, quienes emplearon patrones de prescripción en clínicos de atención médica y hospitales de atención ambulatoria como norma para juzgar el ejercicio particular³⁷.

Al emplear patrones empíricos hay que tener cierta seguridad de que el material clínico en los distintos ámbitos comparados sea similar. En el Estudio de las Actividades Profesionales esto se tiene en cuenta al notificar los patrones de atención en los hospitales agrupados por tamaño. Sin embargo, la mayor deficiencia está en que la atención puede parecer adecuada en comparación con la que se presta en otros casos y ser inferior a la que puede lograrse con la cabal aplicación de los actuales conocimientos de medicina.

Los patrones *normativos* se derivan, en principio, de las fuentes que fijan legítimamente las normas del conocimiento y la práctica en el sistema de atención médica predominante. En la práctica, son fijados por libros de texto comunes o publicaciones¹⁰, grupos de médicos²⁵, profesionales de máxima idoneidad que sirven de jueces²⁶ o un grupo de investigadores en consulta con profesionales calificados²². Los patrones normativos pueden tener niveles muy altos y representar la "mejor" atención médica que se puede prestar o tener un nivel más modesto que represente una atención "aceptable" o "adecuada". En todo caso, su característica distintiva se centra en que emanan de un conjunto de conocimientos y valores legítimos más bien que de ejemplos específicos de la práctica real. Como tales, su validez depende del grado de concordancia sobre hechos y valores dentro de la profesión o, al menos, entre su personal directivo. Donde las fuentes igualmente legítimas difieren en sus puntos de vista, el juicio relativo a la calidad es también ambiguo.

Se ha puesto en tela de juicio la pertinencia de ciertos patrones normativos, formulados por un grupo, para el campo de práctica de otro. Por ejemplo, Peterson y Barsamian indican que, si bien el examen del líquido espermático del esposo debe preceder a la intervención quirúrgica para tratar el síndrome de Stein-Leventhal, no se practicó ni una sola vez y se suprimió ese requisito de los criterios de evaluación³⁸. Tam-

bién se ha expresado insatisfacción en relación con la aplicación al ejercicio general de patrones y criterios formulados por especialistas que ejercen en el ámbito académico. En los principales estudios sobre la práctica general se ha tenido en cuenta ese asunto. Sin embargo, poco se sabe respecto de las estrategias de la "buena" práctica general y de sus similitudes y diferencias en relación con las de la práctica especializada en el ámbito académico.

Algunos investigadores han empleado ambas clases de patrones, normativos y empíricos, para evaluar la atención. Rosenfeld empleó patrones normativos, pero incluyó en su diseño una comparación entre los hospitales universitarios y los comunitarios. "El empleo del hospital docente como testigo da el elemento de flexibilidad necesario para ajustarse a la siempre cambiante base científica de la práctica de la medicina. Ningún patrón escrito, por bien redactado que esté, sería adecuado en cinco años"²². Lembcke se basó en la experiencia adquirida en los mejores hospitales para formular un factor correctivo que suavice la excesiva rigidez de sus patrones normativos. Ese factor, expresado en términos de un porcentaje aceptable de cumplimiento con el patrón, se formuló para tener en cuenta las situaciones imprevistas en los patrones propiamente dichos. Sin embargo, tiene el efecto de ser más permisivo también en su aspecto real. Eso se debe a que el factor de corrección puede constar, en parte, de desviaciones aceptables con respecto al patrón y, en parte, de otras que podrían ser inaceptables.

Los patrones también se pueden diferenciar por su grado de especificidad y de indicaciones detalladas. En un extremo, el médico evaluador puede recibir instrucciones muy sencillas, por ejemplo, "como una vara que sirva para medir la calidad de la atención prestada, tendrá que determinar si habría tratado a ese paciente de ese forma durante ese período de hospitalización en particular"²⁶. En el otro extremo, se puede formular un "sistema lógico" casi impenetrable en el que se especifiquen todas las reglas sobre decisiones que son aceptables para justificar el diagnóstico y el tratamiento^{38,39}. La mayoría de los casos se sitúan en un punto intermedio.

Los patrones muy precisos y estrictamente detallistas guardan relación con la selección de determinadas clases de diagnóstico para fines de evaluación. Cuando se pretende evaluar una muestra representativa de toda la atención prestada, no se puede dar más que una guía general al evaluador. Por ejemplo, Lembcke, que ha hecho hincapié en la necesidad de formular criterios específicos, ha tenido que establecer una clasificación de diagnóstico de la cirugía de la pelvis igualmente detallada¹⁰. Además de la especificidad del diagnóstico, los patrones estrictamente detallados guardan relación con la selección previa de dimensiones específicas de la atención para efectos de evaluación. Ciertas clases de diagnóstico, como el de intervenciones quirúrgicas, se prestan más a inclusión dentro de este método. Eso es evidente en el intento hecho por Lembcke de ampliar su sistema de verificación al diagnóstico distinto del quirúrgico⁴⁰. Los juicios claros, casi empíricos, de idoneidad pierden su nitidez. Los datos resumidos en cada clasificación de diagnóstico se asemejan más a una descripción de los patrones de manejo, con un número insuficiente de criterios normativos para hacer una evaluación definitiva. En

este caso, se adopta la alternativa de la comparación con una institución con criterio establecido.

Obviamente, cuanto más generales y con menos indicaciones detalladas sean los patrones, más habrá que depender de las interpretaciones y normas de la persona encargada de la evaluación real de la atención. Con mayor especificidad, el grupo investigador puede ejercer, en conjunto, mucho más control de las dimensiones de la atención que se deben subrayar y de lo que constituyen patrones aceptables. Los patrones empleados en situaciones estructuradas y no estructuradas parecen tener mucho en común, según lo indica el grado de concordancia entre la clasificación "intuitiva" y la dirigida en el estudio de Rosenfeld²² y entre la "cualitativa" y "cuantitativa" en el de Peterson y colaboradores¹⁸. En realidad, estos dos últimos grupos fueron tan similares que podrían usarse indistintamente.

Cuando los patrones no son muy específicos y el evaluador debe emitir su propio juicio para poder hacer la evaluación, hay que emplear los servicios de jueces muy expertos y cuidadosos. Lembcke alega que un sistema mucho más preciso y detallado como el suyo no necesita jueces expertos. "Se dice que con un libro de cocina, cualquiera que sepa leer puede cocinar. Lo mismo sucede con la revisión de la atención médica con criterios objetivos, y casi en la misma proporción; cualquiera que sepa suficiente terminología médica para entender las definiciones y criterios puede preparar resúmenes de casos y cuadros para verificación médica. Sin embargo, la aceptación, interpretación y aplicación final de los resultados debe ser la responsabilidad de un médico o un grupo de médicos"⁴¹. El "sistema lógico" formulado por Peterson y Barsamian parece adaptarse bien a la clasificación sistematizada en computadora una vez que los hechos básicos se han recopilado, presuntamente por un especialista en resúmenes de fichas^{38,39}.

Por supuesto, las dimensiones de la atención y los valores empleados para juzgarlas están incorporados en los criterios y patrones usados para su evaluación⁴². Por tanto, esos patrones pueden diferenciarse por selectividad e inclusividad en la elección de las dimensiones que es preciso evaluar. Las dimensiones seleccionadas y el juicio emitido respecto de sus valores constituyen la definición funcional de la calidad en cada estudio.

Como ya se señaló, la selección previa de las dimensiones permite establecer procedimientos, patrones y criterios precisos. Lembcke¹⁰ ha hecho gran hincapié en la necesidad de seleccionar algunas dimensiones específicas de atención dentro de determinadas clases de diagnóstico, en lugar de tratar de hacer evaluaciones generales de dimensiones inespecíficas que, en su opinión, carecen de exactitud. Emplea las siguientes dimensiones: confirmación del diagnóstico clínico, justificación del tratamiento (incluso cirugía) e integridad del procedimiento quirúrgico. Dentro de cada dimensión y en cada clase de diagnóstico, se usan a menudo una o más actividades previamente definidas para clasificar los resultados correspondientes a esa dimensión en su conjunto. Son ejemplos de ello la compatibilidad del diagnóstico de pancreatitis con las concentraciones de amilasa sérica o la del diagnóstico de cirrosis del hígado con los resultados de una biopsia, la práctica de pruebas de sensibilidad antes de la antibioterapia en casos de bronquitis aguda y el control del nivel de glucemia en la sangre en casos de diabetes.

Además de la medida en que se realiza la selección previa de las dimensiones, la evaluación de la calidad difiere con respecto al número de dimensiones usadas y la minuciosidad con que se exploran los resultados en cada dimensión. Por ejemplo, Peterson y colaboradores¹⁸ y Rosenfeld²² emplean un gran número de dimensiones. Por otra parte, Peterson y Barsamian^{38,39} se concentran en dos dimensiones básicas, a saber, justificación del diagnóstico y de la terapia, pero exigen prueba completa de la justificación. Huntley *et al.*⁴³ dan un ejemplo de un método más simplificado al evaluar la atención ambulatoria con dos criterios solamente: el porcentaje de exámenes que excluyen ciertos procedimientos de rutina y el porcentaje de anomalías encontradas cuya trayectoria no se siguió.

Cualquier juicio emitido respecto de la calidad es incompleto cuando se usan sólo algunas dimensiones y cuando las decisiones sobre cada una se adoptan basándose en pruebas parciales. Algunas dimensiones, como los cuidados preventivos o la atención psicológica y social de la salud y la enfermedad, se excluyen a menudo de la definición de calidad y de los patrones y criterios que le permiten funcionar. Son ejemplos de ello la exclusión intencional de la atención psiquiátrica del estudio de Peterson¹⁸ y la de la relación entre el paciente y el médico y la actitud de los médicos en los estudios sobre la calidad de la atención en el Plan de Seguro Médico de la Zona Metropolitana de Nueva York²⁷. Rosenfeld²² insistió especialmente en incluir los resultados de determinadas medidas exploratorias entre los criterios de atención superior; pero la atención se consideró buena sin esas medidas. A falta de instrucciones específicas a los jueces, en el estudio de Morehead *et al.*²⁶ se incluyen historias de casos que, en su opinión, recibieron óptima atención, en los que cualquier falla en materia de atención preventiva podría haber tenido graves consecuencias para el paciente.

Otra característica de medición es el nivel en que se fija el patrón. Los patrones pueden ser tan estrictos que nadie cumpla con ellos o tan tolerantes que todos reciban una "buena" clasificación. Por ejemplo, en el estudio de la práctica general notificado por Clute¹⁹, el examen de la tensión arterial, la medición de la temperatura corporal, la otoscopia y la aplicación de inmunizaciones no sirvieron para clasificar a todos los médicos porque todos ellos tuvieron un buen desempeño.

Escalas de medición

La capacidad de hacer una distinción entre diferentes grados de desempeño depende de la escala de medición usada. En muchos estudios de calidad se emplea un número mínimo de divisiones para clasificar la atención, considerada como un conjunto, en categorías como "excelente", "buena", "regular" o "mala". La posición relativa de una persona en cada conjunto se puede especificar más computando el porcentaje de casos en cada categoría de la escala. En otros estudios se asigna cierta calificación a los resultados de determinados elementos de la atención y se acumulan los valores respectivos para obtener un índice numérico que, por lo general, va de 0 a 100. Esas prácticas dan lugar a interrogantes relativos a las escalas de medición y al funcionamiento legítimo de las mismas. A continuación se describen algunas de ellas.

Quienes defienden la primera práctica señalan que no es posible lograr un mayor grado de precisión con los métodos

actuales. Algunos han llegado a reducir las clases a sólo dos: óptima y subóptima. Clute¹⁹ emplea tres y reconoce que la del medio es dudosa o indeterminada. Además, la atención médica tiene un aspecto de totalidad o nulidad que no es posible captar en la calificación numérica normal. La atención puede ser buena en muchos aspectos y desastrosamente inadecuada en su conjunto por un error fundamental de uno de sus elementos. Por supuesto, este no es un problema común, si se demuestra que los resultados obtenidos con diversos elementos de atención guardan una estrecha correlación.

Quienes han empleado el sistema de calificación numérica han señalado que se pierden mucha información al emitir un juicio general³⁸ y que la calificación numérica, acumulada a partir de determinada calificación secundaria, da una idea no sólo del conjunto, sino de la evaluación de cada parte. Rosenfeld²² ha resuelto este problema valiéndose de un sistema de asignación de una calificación cualitativa particular a los elementos de la atención y de una global basándose en reglas arbitrarias de combinación que dan cabida al atributo de totalidad o nulidad de la calidad de la atención médica. Como ya se señaló, hubo mucha concordancia entre la clasificación intuitiva y estructurada en el estudio de Rosenfeld²² y entre la cualitativa y cuantitativa en el de Peterson *et al.*¹⁸.

Un grave problema, todavía sin resolver, en la formulación de un sistema de calificación numérica, es la forma de ponderación de los distintos elementos en el proceso de cálculo del total. En la actualidad, ese es un asunto arbitrario. Por ejemplo, Peterson *et al.*¹⁸ formulan la siguiente escala: historia clínica 30, examen físico 34, uso de ayudas de laboratorio 26, terapia 9, medicina preventiva 6, fichas clínicas 2, total 107. Daily y Morehead²⁴ asignan diferentes grados de importancia, de la manera siguiente: fichas 30, exámenes de diagnóstico 40, tratamiento y seguimiento 30, total 100. Peterson *et al.* dicen que "se asigna la mayor importancia al proceso de emisión de un diagnóstico, ya que sin éste la terapia no puede ser racional. Además, la terapia se encuentra en proceso de cambio constante, mientras que la forma de tomar la historia y practicar el examen físico se ha modificado muy poco en el correr del tiempo"¹⁸. Daily y Morehead no ofrecen ninguna justificación de sus índices de ponderación, pero tal vez se podrían presentar en su nombre argumentos igualmente persuasivos. Sigue existiendo el problema de la búsqueda de confirmación externa⁴⁴.

El problema de los índices de ponderación está relacionado con el de naturaleza más general del valor de los puntos de información o de procedimientos en el proceso de atención médica. Rimoldi *et al.*³⁴ emplearon la frecuencia con que se usaron determinados puntos de información para solucionar un problema de ensayo como medida del valor de dicho punto. Williamson hizo que los expertos clasificaran determinados procedimientos en un medio específico para la práctica de pruebas de diagnóstico, en una escala que osciló entre "muy útil" y "muy nocivo". Los resultados obtenidos en cada prueba se clasificaron luego empleando índices cuantitativos de "eficiencia", "pericia" y "competencia" general, según la frecuencia y naturaleza de los procedimientos empleados³⁵.

Un problema encontrado en la interpretación de la calificación numérica es el significado del intervalo numérico de los

puntos de la escala. La calificación numérica asignada con fines de evaluación de la calidad generalmente carece de intervalos iguales. No se debe usar como si tuviera esa propiedad.

Confiabilidad

La confiabilidad de las evaluaciones es un punto de gran importancia en los estudios de calidad, en los que tantas cosas dependen del juicio emitido, aun cuando se usan patrones detallados. En varios estudios se ha prestado cierta atención al acuerdo entre los jueces. Se tiene la impresión de que se considera aceptable. Basándose en otras 14 visitas hechas por observadores, Peterson *et al.*¹⁸ consideraron que había suficiente acuerdo para poder agrupar todas las observaciones en una de las seis clases principales de atención, después de hacer las modificaciones necesarias par tener en cuenta el sesgo de parte de cada observador. En el estudio de Daily y Morehead "se hicieron varias verificaciones del trabajo de los dos médicos internistas entrevistadores, pidiéndoles que entrevistarán a los mismos médicos. Las diferencias en la calificación dada a los médicos de familia basándose en la clasificación asignada por separado no fueron superiores a 7%"²⁴. Rosenfeld²² prestó gran atención a los análisis de fiabilidad de las pruebas y formuló índices matemáticos de "acuerdo" y "dispersión" para medirla. Esos índices muestran que existe bastante acuerdo, pero es difícil hacer una evaluación precisa puesto que no se ha comprobado que ningún otro investigador haya usado esas mismas medidas. Morehead *et al.*²⁶, en el segundo estudio de la atención médica que recibieron los familiares de los miembros del sindicato de conductores de camión, indican que hubo acuerdo inicial entre dos jueces al asignar la atención a una de dos clases en 78% de los casos. Esa cifra aumentó a 92% después de reevaluar los casos de desacuerdo de los dos jueces.

En contraste con la confiabilidad de la evaluación entre jueces, poco se ha dicho sobre la de los juicios emitidos respecto de la calidad en repetidas ocasiones por la misma persona. Para someter a ensayo la variación del juicio de un mismo observador, Peterson *et al.*¹⁸ les pidieron a dos de ellos que volvieran a visitar a cuatro de los médicos que habían visitado antes. El grado de concordancia del juicio de un mismo observador fue menor que el existente entre uno y otro, en parte porque las nuevas visitas fueron más breves y, por tanto, guardaron relación con una muestra de menor tamaño.

El principal mecanismo para lograr un mayor grado de fiabilidad es la especificación detallada de criterios, patrones y procedimientos empleados para evaluar la atención. De hecho, el esfuerzo por duplicar el proceso representó un gran impulso para el establecimiento de los sistemas de clasificación más rigurosos de Lembcke y de Peterson y Barsamian. Lamentablemente, no existen estudios comparativos de confiabilidad entre los métodos de evaluación muy específicos en sus indicaciones y los que no lo son. Los datos sin elaborar de Rosenfeld podrían permitir una comparación de la fiabilidad de los juicios "intuitivos" y la de los juicios estructurados de los mismos dos evaluadores. Se podrían analizar los datos inéditos de Morehead *et al.*²⁶ de la misma manera que los de Rosenfeld²² para dar información útil sobre la relación existente entre el grado de fiabilidad y el método de evaluación. Los datos parciales que se han publicado sugieren que la fiabilidad lograda por Morehead *et al.* después

del examen, al emplear el método menos indicativo, es similar a la lograda por Rosenfeld, que empleó una técnica mucho más indicativa.

Morehead y colaboradores preguntaron si la fiabilidad proveniente de la especificación detallada de patrones y criterios no se logra a costa de una menor validez, lo que reviste importancia. "A menudo, esos criterios obligan a introducir en un rígido marco acciones o factores similares que quizá no sean apropiados en una situación dada, debido a las infinitas variaciones de la reacción del cuerpo humano a la enfermedad...". El grupo de estudio rechaza la hipótesis de que esos criterios sean necesarios para evaluar la calidad de la atención médica. Opinan unánimemente que la flexibilidad es tan importante para los evaluadores al juzgar cada caso como para el médico competente al confrontar el problema clínico de una paciente determinado²⁶.

Las razones del desacuerdo entre los jueces aclaran los problemas de evaluación y las posibilidades de lograr mayor confiabilidad. Rosenfeld descubrió que "casi la mitad de las diferencias eran atribuibles a situaciones no cubiertas debidamente por los patrones o en las que estos eran ambiguos. En otro ámbito surgieron diferencias sobre preguntas reales, porque un consultor pasó por alto un punto de información importante en el registro. Por tanto, al parecer, se podría lograr un grado de concordancia mucho mayor con patrones revisados y mejores métodos de orientación de los consultores²². Menos del 25% de los casos de desacuerdo se refieren a diferencias de opinión en lo relativo a los requisitos para el manejo de los pacientes. Esta es una función de ambigüedad del sistema de atención médica y fija un límite máximo de duplicación. Morehead *et al.* informan que en cerca de la mitad de los casos de desacuerdo inicial "hubo concordancia en cuanto al aspecto más serio de la atención del paciente, pero un evaluador afirmó más tarde que no había tenido en cuenta los corolarios de esa atención²⁶. Otras razones del desacuerdo fueron la dificultad para adherirse a diversas escalas de clasificación o el no haber tenido en cuenta todos los hechos. Del reducido número de casos de desacuerdo sin resolver (8% de todos los pacientes internados y 36% de los casos de desacuerdo inicial) más de la mitad se debieron a diferencias de opinión bien intencionadas respecto del manejo clínico del problema. El resto surgió de diferencias de interpretación de registros inadecuados o de los problemas técnicos relativos al punto en que se debe evaluar la atención insatisfactoria en una serie de hospitalizaciones²⁷.

Un último aspecto de la confiabilidad es el fallo ocasional en el desempeño de un evaluador, como se ha demostrado en forma tan impresionante en el estudio de Reosenfeld²². Se planteará aquí la pregunta, sin intención de contestarla, de qué es lo que hace el investigador cuando un segmento bien definido de sus resultados aparece totalmente aberrante.

Sesgo

Cuando varios observadores o jueces describen y evalúan el proceso de atención médica, uno de ellos puede emplear constantemente patrones más rígidos que otro o interpretar patrones previamente determinados en forma más estricta. Peterson *et al.*¹⁸ descubrieron que uno de sus observadores daba generalmente una clasificación más alta que el otro al evaluar

la práctica del examen físico, pero no a la prestación de otras clases de atención. Rosenfeld²² mostró que uno de los dos evaluadores asignaba regularmente una clasificación más baja a los mismos casos estudiados por ambos. El examen de cada caso de desacuerdo en el estudio hecho por Morehead y colaboradores²⁶ revela que, en la clase correspondiente a atención médica, el mismo evaluador dio a la atención una clasificación más baja en 11 de 12 casos de desacuerdo. En lo que respecta a casos quirúrgicos, un evaluador asignó una clasificación más baja a la atención que el otro en los ocho casos de desacuerdo. Al examinar las razones de desacuerdo en los casos médicos se tiene la impresión de que uno de los jueces tenía especial interés en la cardiología y exigía más claridad y seguridad en el tratamiento de los casos de cardiopatía.

Estos resultados indican claramente que hay que aceptar el sesgo como la regla más que la excepción y que los estudios de calidad se deben formular teniendo eso en cuenta. Por ejemplo, en el estudio de Rosenfeld²² cualquiera de los dos evaluadores cuyos servicios se emplearon en cada campo de ejercicio profesional habrían calificado a varios hospitales en el mismo orden, aunque uno fue siempre más generoso que el otro. Por otra parte, el estudio del ejercicio general hecho por Clute¹⁹, en Canadá, se ha criticado por comparar la calidad de la atención en dos regiones geográficas, aunque diferentes observadores examinaron la atención en las dos zonas en cuestión⁴⁵. El autor conocía ese problema y encontró métodos de comparación del desempeño de los observadores en las dos zonas geográficas, pero la deficiencia básica permanece igual.

El orden o la regularidad previamente determinados en el proceso de estudio puede guardar relación con el sesgo. Por tanto, es posible que haya que introducir procedimientos cuidadosamente planeados al formular investigaciones con fines de aleatorización. El estudio realizado por Peterson *et al.*¹⁸ parece ser uno de los pocos en los que se ha prestado atención a ese factor. Otra importante fuente de sesgo está en los conocimientos que tenga el evaluador respecto de la identidad del médico que prestó la atención o del hospital en que se dispensó. Se ha preguntado si es posible eliminar las características de identificación de los cuadros objeto de análisis³, pero poco se sabe sobre la factibilidad de ese procedimiento y sus efectos en la clasificación asignada. Puede producirse todavía otra clase de sesgo como resultado de los restringidos patrones y criterios de práctica que pueden surgir en ciertas instituciones o "facultades" de ejercicio de la medicina y a su alrededor. Mientras eso sea verdad o se sospeche de que sea, habrá que tomar las debidas precauciones para la consecución y asignación de jueces.

Validez

La eficacia de la atención, como se dijo, para lograr o producir salud y satisfacción, según la definición de una sociedad o subcultura particular para cada uno de sus miembros, es el elemento fundamental de comprobación de la validez de la calidad de la misma. La validez de los demás fenómenos como indicadores de la calidad depende, en último análisis, de la relación existente entre esos fenómenos y el logro de salud y satisfacción. Sin embargo, la conformidad de la práctica con los patrones aceptados tiene una clase de validez condicional o

parcial que puede ser más pertinente para fines de evaluación en determinados casos.

La confirmación de la validez de los detalles de la práctica de la medicina por sus efectos para la salud es de particular interés para las ciencias clínicas. En las publicaciones sobre medicina clínica se buscan datos para determinar si la penicilina fomenta la recuperación en ciertas clases de neumonía, los anticoagulantes en la trombosis coronaria o los corticosteroides en la carditis reumática; lo que indican ciertos análisis sobre la función hepática; y si la mastectomía simple o radical es el procedimiento que más prolonga la vida en determinadas clases de cáncer de mama. De los conocimientos generales sobre esas relaciones surgen los patrones de la práctica, cuya validez se ha comprobado plenamente en mayor o menor grado y mediante los cuales se juzga de ordinario el proceso de atención médica.

Los extremos de naturaleza intermedia o relativos a procedimiento representan a menudo mayores conjuntos de atención. Su relación con el resultado ha atraído la atención del investigador clínico y del estudiante de la organización de atención médica. Algunos ejemplos de esto último son los estudios de las relaciones existentes entre la atención prenatal y la salud maternoinfantil^{46,47} y entre los exámenes exploratorios múltiples y el estado de salud subsiguiente⁴⁸. Un interesante ejemplo del estudio de la relación existente entre un extremo relativo a procedimiento y otro es el intento de demostrar que existe una relación favorable entre la práctica de los exámenes rectales y vaginales por el médico y la confirmación patológica de apendicitis en apendectomía primarias, según se informa en el Estudio de las Actividades Profesionales⁴⁹.

En muchos estudios reseñados^{18,19,23,26,28} se trata de analizar la relación existente entre las propiedades estructurales y la evaluación del proceso de atención. Por ejemplo, en varios de estos estudios se ha demostrado que existe una relación entre la formación e idoneidad de los médicos y la calidad de la atención que prestan. Sin embargo, la relación es compleja y recibe la influencia de la clase de formación, su duración y el tipo de hospital en que se recibió. Los dos estudios de la práctica general^{18,19} han mostrado otras relaciones favorables entre la calidad y la disponibilidad de mejores consultorios para la práctica, la dotación de equipo de laboratorio y la institución de un sistema de registro de citas. No se demostró que hubiera ninguna relación entre la calidad y el pertenecer a asociaciones profesionales, el ingreso del médico o la dotación de equipo de radiografía del consultorio. Los dos estudios no concuerdan completamente respecto de la naturaleza de la relación existente entre la calidad de la práctica y la formación del médico en un hospital docente o en otro, el número de horas trabajadas o la naturaleza de la afiliación hospitalaria del médico. La acreditación de los hospitales, presuntamente una marca de calidad conferida principalmente por la observación de una amplia gama de patrones de organización, en sí y por su propia naturaleza, no parece guardar relación con la calidad de la atención, al menos en la ciudad de Nueva York²⁶.

Aunque, sin duda alguna, la estructura y el proceso están relacionados, los pocos ejemplos citados indican claramente la complejidad y ambigüedad de esas relaciones. Este es el resultado, en parte, de los muchos factores que intervienen y, en

parte, de interacciones mal entendidas de los mismos. Por ejemplo, basándose en varios resultados^{26,38}, se podría proponer razonablemente que los factores relativos a los hospitales y a los médicos influyen en la calidad de la atención prestada en el hospital, pero que las diferencias entre los médicos se eliminan en los mejores y los peores hospitales y se expresan, en distintos grados, en los de calidad intermedia.

Un método preferido por los estudiantes de organización de la atención médica consiste en examinar las relaciones existentes entre la estructura y el resultado, sin referirse a los complejos procesos que los vinculan. Ya se han citado algunos ejemplos de esos estudios⁶⁻⁹. Otros incluyen estudios de los efectos que tiene la reorganización de una clínica de atención ambulatoria para el estado de salud⁵⁰, los efectos de la atención hospitalaria intensiva en la recuperación⁵¹, los efectos de la atención en el hogar en la supervivencia⁵² y el efecto de un programa de rehabilitación en el estado físico de los pacientes atendidos en sanatorios particulares^{53,54}. La falta de una relación con el resultado en los dos últimos estudios sugiere que, en algunos casos, la opinión actual sobre la forma de organizar la atención no está bien definida.

En esta breve reseña se indican las clases de pruebas relativas a la validez de los diferentes métodos de evaluar la calidad de la atención. Obviamente, no se entiende a cabalidad la relación entre el proceso y el resultado ni entre estos dos últimos y la estructura. Con respecto a ello, los requisitos para confirmar la validez se pueden expresar mejor con el concepto, ya citado, de una cadena de acontecimientos en la que cada uno marca el final del precedente y representa una condición necesaria para el siguiente. Eso indica que la relación de los medios con el fin entre pares adyacentes exige que se confirme la validez en cualquier cadena de acontecimientos hipotéticos o reales⁵⁵. Por supuesto, este es un proceso laborioso. Como se ha demostrado, los vínculos entre uno y otro son los que más comúnmente se pasan por alto. Como resultado, las inferencias causales se atenúan en proporción de la distancia que separa a los dos acontecimientos en la cadena.

Lamentablemente, existe muy poca información sobre la evaluación real de la calidad en la que se emplea más de un método de evaluación simultáneamente. Los estudios de Makover se han concentrado en particular en la relación existente entre la evaluación multifactorial de la estructura y la del proceso en los mismos grupos médicos. "Se descubrió que los grupos médicos que lograron clasificaciones altas en cuanto a calidad por el método usado en el presente estudio fueron los que, en general, se ciñeron más estrictamente a los patrones médicos mínimos fijados por el Plan de Seguro Médico. Sin embargo, las excepciones fueron suficientemente notorias, en número y grado, para poner en tela de juicio la fiabilidad⁵⁶ de uno u otro método de clasificación cuando se aplicó a cualquier grupo médico. Al parecer, es obvio que se necesita hacer otras comparaciones de estos dos métodos de clasificación"²³.

Índices de atención médica

Puesto que una evaluación multidimensional de la atención médica es una empresa costosa y laboriosa, se siguen buscando datos específicos fácilmente mensurables que propor-

cionen información sobre la calidad de la atención médica. Los datos usados pueden referirse a varios aspectos de estructura, proceso o resultado. El principal requisito es que se puedan medir con facilidad, a veces, en forma regular, y que tengan una validez razonable. Ente los estudios de calidad en los que se emplea este método están el de actividades profesionales³⁶ y los de Ciocco⁵⁷ y Furstenberg *et al.*³⁷.

Esos índices tienen la ventaja de ser convenientes; pero las inferencias derivadas de los mismos pueden ser de dudosa validez. Myers ha señalado las muchas limitaciones de los índices tradicionales de la calidad de la atención hospitalaria, incluso tasas de mortalidad total y postoperatoria, complicaciones, infección postoperatoria, cesárea, consultas y extirpación de tejido normal en una operación⁵⁸. La precisión e integridad de la información básica puede estar sujeta a dudas. Más importante aún es la posibilidad de que surjan serios interrogantes sobre lo que significa cada índice, ya que son muchos los factores que intervienen para que se produzca el fenómeno que mide. Por otra parte, Eislee ha señalado que por lo menos ciertos índices pueden ser útiles, si se emplean con cuidado³⁶.

La búsqueda de formas fáciles de medir un fenómeno tan complejo como la atención médica puede ser un sueño imposible. El uso de índices sencillos en lugar de medidas complejas se puede justificar demostrando una elevada correlación entre los mismos¹. No obstante, por no haberse podido demostrar que existen vínculos causales, ésta puede ser una base inestable como punto de partida. Por otra parte, cada índice puede ser una medida de una dimensión o un elemento de atención. Por tanto, los índices múltiples seleccionados acertadamente pueden constituir el equivalente de perforaciones de reconocimiento en un estudio geológico, que dan suficiente información sobre las partes para poder reconstruir el todo. Por supuesto, la validez de las inferencias sobre el todo dependerá de la medida de la continuidad interna en el ejercicio particular o institucional de la medicina.

Algunos problemas para evaluar la atención ambulatoria

Ya se han mencionado algunas de las dificultades especiales para evaluar la calidad de la atención ambulatoria. Incluyen la escasez de información registrada y los conocimientos previos del médico encargado en lo relativo a los antecedentes médicos y sociales del paciente. El primero de esos problemas ha llevado a usar los servicios de observadores adiestrados y el segundo, a observar casos en que los conocimientos previos no constituyen un factor de importancia para el tratamiento en curso. También se ha puesto en tela de juicio el grado de importancia que para el ejercicio general tienen los patrones y estrategias de atención establecidos por médicos centrado en los hospitales y orientados hacia el ámbito académico.

Otro problema es la dificultad de definir el segmento de la atención que puede ser el objeto de evaluación en materia de atención ambulatoria⁵⁹. Cuando se trata de atención hospitalaria, una solo ingreso es comúnmente la unidad apropiada⁵⁹. Cuando se trata del ejercicio en un consultorio particular o una clínica, una secuencia de atención puede cubrir un número indeterminado de consultas, de modo que la identificación de la unidad apropiada representa un interrogante. Por lo común, la

respuesta consiste en escoger un período en forma arbitraria para definir el episodio de atención pertinente. Ciocco y colaboradores⁵⁷ lo definieron como la primera consulta más 14 días de seguimiento. Huntley *et al.*⁴³ usan un período de cuatro semanas después del examen inicial.

Conclusiones y propuestas

En esta reseña se ha tratado de dar una idea de los diversos enfoques y métodos empleados para evaluar la calidad de la atención médica y de señalar ciertos asuntos y problemas destacados por los mismos, que exigen consideración.

Cabe afirmar que los métodos empleados han sido de dudoso valor y carecen a menudo de rigor y precisión. Pero, ¿cuán precisas tienen que ser las estimaciones de calidad? Al menos los mejores métodos han sido adecuados para los fines de política administrativa y social que les han dado origen. La búsqueda de perfección no debería cegarnos respecto del hecho de que las actuales técnicas de evaluación de la calidad, por rudimentarias que sean, han revelado una escala de calidad que va de sobresaliente a deplorable. Hoy en día se dispone de medios para emitir juicios de amplio alcance con bastante seguridad. Este grado de seguridad se apoya en resultados a los que ya se ha hecho alusión, que sugieren que existe un grado aceptable de homogeneidad en la práctica individual y de poder duplicar los juicios cualitativos basados en un método de evaluación de estructuración mínima. Eso no quiere decir que no quede todavía mucho por hacer para lograr la mayor precisión que exigen ciertos otros fines.

Se podría comenzar un catálogo de los puntos que se necesita perfeccionar, considerando la naturaleza de la información que constituye la base de cada juicio emitido respecto de la calidad. Se debe saber más sobre el efecto que tiene el observador en la práctica analizada y sobre el proceso de observación propiamente dicho y su fiabilidad y validez. Es preciso hacer comparaciones entre la observación directa y la información registrada con y sin el complemento de una entrevista con el médico encargado. No basta registrar los casos en que hay acuerdo o desacuerdo. Se necesita un estudio más detallado de la naturaleza y las razones de la discrepancia en varios ámbitos. Asimismo, debe someterse a prueba el uso de resúmenes de fichas frente al de las fichas mismas.

El proceso de evaluación propiamente dicho exige un estudio más detallado. Se dedica un gran esfuerzo a la formulación de criterios y patrones que presuntamente dan estabilidad y uniformidad a cualquier juicio emitido sobre la calidad y, sin embargo, ese presunto efecto no se ha demostrado empíricamente. No se sabe hasta dónde debe ir la normalización explícita para lograr adelantos considerables en materia de fiabilidad. También se debe considerar si con una mayor normalización se pierde tanto la capacidad de dar cuanta de los elementos imprevistos en la situación clínica que se logra fiabilidad a costa de validez. La evaluación del mismo conjunto de fichas con patrones y criterios cada vez más estructurados debe dar información valiosa sobre estos puntos. La afirmación de que los evaluadores menos idóneos que usan criterios minuciosos pueden emitir un juicio fidedigno y válido también se puede someter a prueba de esa forma.

Anteriormente se ha indicado lo poco que se sabe respecto de la fiabilidad y del sesgo cuando se comparan dos o más jueces y de la fiabilidad de los juicios emitidos en repetidas ocasiones sobre los mismos puntos de atención por el mismo evaluador. Asimismo, se sabe muy poco sobre los efectos que en la fiabilidad y validez tienen ciertas características de los jueces, incluso su experiencia, sus campos de particular interés y los factores de su personalidad. Se puede aprender mucho sobre estos y otros asuntos afines haciendo explícito el proceso de emisión de juicios y sometiéndolo a un cuidadoso estudio. Eso debe revelar las dimensiones y los valores empleados por los diversos jueces y mostrar cómo se resuelven las diferencias cuando dos o más jueces discuten sus puntos de vista. Hoy en día existe alguna duda respecto de la validez de los actos de conciliación realizados en grupo en los que puede predominar un punto de vista, no necesariamente por ser más válido¹. El efecto de enmascarar la identidad del hospital o del médico que presta atención se puede estudiar de la misma manera. Aquí se propone no sólo que se demuestren las diferencias o similitudes en cuanto a juicios en general sino que, al hacer explícita la forma de pensar de los jueces, se trata de determinar cómo surgen las diferencias y similitudes y cómo se resuelven las primeras.

Además de los defectos del método, en la mayoría de los estudios se ha adoptado una definición de calidad demasiado limitada. En general, se ocupan del tratamiento técnico de la enfermedad y prestan poca atención a la prevención, rehabilitación, coordinación y continuidad de la atención o al manejo de la relación entre el paciente y el médico. Presuntamente, la razón de ello se centra en que los requisitos técnicos del tratamiento se reconocen más ampliamente y están más normalizados. Por tanto, se necesita una exploración conceptual y empírica más completa de la definición de calidad.

El significado de "exploración conceptual" se puede explicar considerando la dimensión de eficiencia que se pasa por alto a menudo en los estudios de calidad. Se podrían distinguir dos tipos de eficiencia: lógica y económica. La eficiencia lógica tiene que ver con el uso de información para llegar a una decisión. En este caso, el problema podría ser si la información obtenida por el médico es pertinente o no para el asunto clínico que será objeto de transacción. En caso afirmativo, se podría considerar el grado de repetición o duplicación de los datos obtenidos y la medida en que excede de los requisitos de la adopción de decisiones en una situación dada. Si la parsimonia representa un valor en la atención médica, la identificación de redundancia se convierte en un elemento de la evaluación de la atención.

La eficiencia económica se refiere a las relaciones entre insumos y productos y lleva a preguntarse si se puede obtener un determinado producto a menor costo. Por supuesto, recibe la influencia de la eficiencia lógica, ya que la acumulación de información innecesaria o no utilizada es un procedimiento costoso que no reporta ningún beneficio. Típicamente, se extiende más allá del individuo y guarda relación con el producto social del esfuerzo de atención médica. Entraña consideración de la posibilidad de que la "mejor" atención médica para el individuo tal vez no lo sea para la comunidad. Peterson y colaboradores citan un ejemplo que resume este asunto. "Dos médicos habían

delegado la supervisión de las consultas prenatales ordinarias en las enfermeras del consultorio y uno de ellos atendía a la paciente sólo si tenía alguna queja particular"¹⁸. Por un lado, esta puede haber sido atención subóptima para cada madre embarazada. Por otro, puede haber sido una brillante estrategia para proporcionar la pericia conjunta de un equipo de atención médica al mayor número posible de mujeres. Cordero, en un trabajo que incita a la reflexión, ha documentado la tesis de que, cuando los recursos son limitados, la atención médica óptima a la comunidad puede exigir un nivel subóptimo para cada uno de sus miembros⁶⁰.

Además de la exploración conceptual del significado de calidad, en función de las dimensiones de atención y de los valores que se les asignan, se necesitan estudios empíricos de lo que son las dimensiones y los valores predominantes en grupos de población pertinentes⁵. Por ejemplo, poco se sabe respecto de la forma en que los médicos definen la calidad y no se conoce la relación ente el ejercicio del médico y su propia definición de calidad. Este es un campo de investigación importante tanto para la educación médica como para la calidad. Los estudios empíricos del proceso de atención médica también deben contribuir mucho a la identificación de las dimensiones y los valores que se deben incorporar en la definición de calidad.

Una reseña de los estudios de calidad muestra una cierta repetitividad desalentadora en los conceptos, enfoques y métodos básicos. Cualquier otro adelanto sustantivo, fuera del perfeccionamiento de la metodología, es factible que provenga de un programa de investigación del proceso de atención médica propiamente dicho más bien que de ataques directos al problema de la calidad. Se cree que eso es así porque, antes de poder emitir un juicio respecto de la calidad, se necesita entender la interacción del paciente con el médico y la actuación de este en el proceso de prestación de atención. Una vez que se entiendan los elementos del proceso y su interrelación, se puede juzgar su valor en función de su aporte al logro de metas intermedias y finales. Suponga, por ejemplo, que el conjunto formado por el autoritarismo y la permisividad representa una dimensión de la relación del paciente con el médico. Un estudio empírico puede mostrar que los médicos se diferencian, de hecho, por ese atributo. Se podría preguntar entonces si el autoritarismo o la tolerancia debe ser el criterio de calidad. La respuesta podría derivarse de los valores generales de la sociedad que puede apoyar a uno u otro como el atributo más deseable en las interacciones sociales. Esta es una forma de emitir un juicio respecto de la calidad y es perfectamente válida, siempre y cuando su razón justificativa y sus bases sean explícitas. Sin embargo, el estudio del proceso de atención médica en sí puede ofrecer un enfoque distinto y más pragmático. Suponga, de momento, que el cumplimiento de las recomendaciones del médico es una meta y un valor en el sistema de atención médica. El valor del autoritarismo o la permisividad se puede determinar, por su aporte al cumplimiento. La validez de ese cumplimiento en sí se debe confirmar con el criterio de los resultados de la atención, que es de orden superior. Tal vez el verdadero estado de cosas sea más complejo que el ejemplo hipotético dado. Se puede dar el caso de que el criterio de calidad sea la congruencia con las esperanzas de los pacientes o una adaptación más compleja a determinadas situaciones

clínicas y sociales, en lugar del autoritarismo o la permisividad como modalidad predominante. Además, tal vez ciertas metas del proceso de atención médica no sean compatibles con otras y uno no pueda referirse a la calidad en términos generales sino dentro de dimensiones y fines específicos. Por tanto, la evaluación de la calidad no resultará en un juicio sumario sino en un perfil complejo, según lo ha sugerido Sheps¹.

Por supuesto, una gran parte de la investigación sobre el proceso de atención médica se centrará en la manera en que los médicos acopian información de importancia clínica y adoptan decisiones en cuanto a diagnóstico y terapia. Este no es el lugar apropiado para presentar un marco conceptual de investigación de esta parte del proceso de atención médica. Sin embargo, se pueden mencionar determinados estudios y dar algunas indicaciones para realizar otras investigaciones.

Las investigaciones sobre el acopio de información incluyen estudios sobre la percepción e interpretación de los signos físicos^{61,62}. Por ejemplo, Evans y Bybee han demostrado que, al interpretar los sonidos cardíacos, se cometieron errores de percepción (del ritmo y la frecuencia), junto con otros de interpretación sobre lo percibido. El diagnóstico equivocado, juzgado en comparación con un criterio, fue el resultado de estos dos errores⁶². Eso señala la necesidad de incluir en las estimaciones de calidad información sobre la fiabilidad y validez de los datos producto de las funciones de los sentidos, sobre los que se basa parcialmente el tratamiento.

El trabajo de Peterson y Barsamian^{38,39} representa el método más cercano a una rigurosa evaluación de las decisiones adoptadas en materia de diagnóstico y terapia. Como tal, es quizá el adelanto reciente de mayor importancia en cuanto a métodos de evaluación de la calidad. No obstante, ese método se basa en reseñas de fichas y se destina casi exclusivamente a justificar el diagnóstico y la terapia. Como resultado, en la evaluación se excluyen muchas dimensiones importantes de la atención. Algunas de estas son asuntos relativos a eficiencia y estilos y estrategias para la resolución de problemas.

Los estilos y estrategias para la resolución de problemas se pueden estudiar por medio de la observación real de la práctica, como lo hicieron con tanta eficacia Peterson y colaboradores en su estudio de la práctica general¹⁸. Una gran parte de lo que no se ha observado puede hacerse explícito al pedirle al médico que diga en voz alta lo que está haciendo y por qué. Este método de *reflexión en voz alta* se ha empleado en estudios de resolución de problemas aunque, en sí, puede modificar la conducta⁶³. Otro método consiste en crear situaciones experimentales, como las establecidas por Rimoldi *et al.*³⁴ y por Williamson³⁵ para observar el proceso de adopción de decisiones. Aunque esas situaciones tienen ciertas limitaciones que emanan de su artificialidad⁶⁴, su mayor simplicidad y el mayor control que ofrecen pueden ser de gran utilidad.

A primera vista, el estudiante de atención médica podría esperar ayuda de los conocimientos teóricos y prácticos provenientes del campo general de la investigación en lo que respecta a solución de problemas. Lamentablemente, no existe una base teórica bien establecida que se pueda explotar con facilidad en los estudios de atención médica. Algunos de los estudios empíricos en materia de solución de problemas podrían sugerir métodos e ideas aplicables a situaciones de aten-

ción médica⁶³⁻⁶⁷. Ciertos estudios de resolución de problemas relativos a equipo electrónico, en particular, muestran una intrigante similitud con el proceso de diagnóstico y tratamiento médico. En estos y en otros similares se han identificado características de comportamiento que se podrían emplear para clasificar diversos estilos de tratamiento clínico. Incluyen el volumen de información recolectada, la intensidad de la búsqueda de información, el valor de los puntos de información buscados y modificados según su lugar en una secuencia e interacción con otros, varias clases de redundancia, el surgimiento de estereotipos, los patrones de investigación en relación con la parte defectuosa, las tendencias a obrar antes de conseguir suficiente información o a buscarla más allá del punto en que se puede garantizar razonablemente una solución, la "distancia con respecto a los errores", el éxito en el logro de una solución y así sucesivamente.

La teoría de adopción de decisiones también puede ofrecer instrumentos conceptuales de investigación en el proceso de atención médica. Ledley y Lusted^{68,69}, entre otros, han tratado de aplicar modelos basados en probabilidades condicionales al proceso de diagnóstico y terapia. Peterson y Barsamian^{38,39} decidieron no emplear probabilidades en sus sistemas lógicos por no tener a su disposición los datos necesarios (la probabilidad independiente de enfermedades y síntomas y la de determinados síntomas de ciertas enfermedades). Sin embargo, Edwards y colaboradores⁷⁰ señalan que todavía se puede someter a prueba la eficiencia en materia de adopción de decisiones empleando los datos estadísticos que uno preferiría tener, en sustitución de la probabilidad subjetiva (la de la persona que adopta las decisiones o la de expertos seleccionados).

Una pregunta básica que surge con frecuencia a lo largo del presente artículo es la medida en la cual la actuación en materia de atención médica es un fenómeno homogéneo o heterogéneo. Por ejemplo, se observó que esto era de importancia para el muestreo, el uso de índices en lugar de medidas multidimensionales y la construcción de escalas con las que se pretende juzgar la actuación en su conjunto. Cuando se hace esta pregunta con respecto a médicos individuales, el objeto del estudio es la integración de varias clases de conocimientos y habilidades relativos a la personalidad y la conducta del médico. Cuando se formula esa pregunta respecto de las instituciones y los sistemas sociales, los factores son completamente distintos. En este caso uno se preocupa por los mecanismos formales e informales para organizar, moldear y dirigir el esfuerzo humano, en general, y la práctica de la medicina, en particular. Se espera que la investigación en todos estos campos contribuya a una mayor complejidad en lo que respecta a la medición de la calidad.

Algunas de las convenciones aceptadas en la presente reseña constituyen, por su propia naturaleza, obstáculos para un estudio más importante de la calidad. En el mundo real, los servicios de los médicos no se prestan aparte de los de otros profesionales de salud ni de los de una gran variedad de funcionarios de apoyo. La separación de la atención hospitalaria y ambulatoria también es bastante artificial. Las unidades de atención que son los objetos de estudio propiamente dichos incluyen los aportes de muchas personas durante una secuencia que puede incluir atención en varios ámbitos. La forma en que se definen e identifican esas secuencias tiene repercusiones para

el muestreo, los métodos de acopio de información y los patrones y criterios de evaluación.

Un último comentario se relaciona con la actitud con que se abordan los estudios sobre calidad. Ya se han mencionado los imperativos sociales que llevan a evaluar la calidad. A menudo guardan relación con ellos el celo y los valores del reformador social. En los estudios de calidad se necesita mayor neutralidad y objetividad. A menudo hay que preguntarse "¿qué sucede aquí?" en lugar de "¿hay algo que anda mal?" y "¿cómo se puede mejorar?". Eso no quiere decir que el investigador renuncie a sus propios valores u objetivos sociales, sino que se reconoce y mantiene la distinción entre los valores y los elementos de estructura, proceso o resultado y que ambos están sujetos a un estudio igualmente crítico. En parte, para lograr esa clase de orientación, hay que dejar de preocuparse por evaluar la calidad par concentrarse en entender el proceso de atención médica propiamente dicho.

Nota de agradecimiento

Entre los autores, aquí incluidos, que leyeron la versión preliminar e hicieron correcciones o comentarios están Georgopoulos, Makover, Morehead, Peterson, Riedel, Rosenstock, Rosenfeld, Sheps y Weinerman. El autor tiene una profunda deuda de gratitud con la Dra. Mildred A. Morehead y con los profesores Basil S. Georgopoulos, Herbert E. Klarman y Charles A. Metzner por haber tomado tiempo para hacer extensos comentarios. El Sr. Sam Shapiro y el Dr. Jonas N. Muller, críticos oficiales, fueron de gran ayuda en la definición precisa de algunos problemas de evaluación de la calidad. Puesto que el autor no pudo emplear en su totalidad los excelentes consejos recibidos, asume plena responsabilidad por cualquier falla del presente documento.

Bibliografía y notas*

1. Sheps MC. Approaches to the quality of hospital care. *Public Health Reports* 1955;70:877-86.

Este documento representa una materialización insólitamente fructífera de las ideas relativas a la evaluación de la calidad. Contiene discusiones breves pero sumamente completas de los fines de evaluación, los problemas de definición, los criterios y patrones, los distintos sistemas de medición, la fiabilidad de los juicios cualitativos y los índices de calidad. La bibliografía es excelente.

2. Peterson L. Evaluation of the quality of medical care. *New England Journal of Medicine* 1963;269:1238-45.
3. Lerner M, Riedel DC. The teamster study and the quality of medical care. *Inquiry* 1964;1:69-80.

El valor principal de este documento está en que formula preguntas sobre métodos de evaluación, incluido el muestreo de poblaciones y las diversas clases de diagnóstico, el uso de fichas y la necesidad de suplementación con entrevistas, el valor de los patrones detallados, la necesidad de entender el proceso de verificación, la definición de términos y conceptos (por ejem-

plo, de "hospitalización innecesaria") y los problemas de definición del episodio de atención pertinente.

4. Lee RI, Jones LW. *The Fundamentals of Good Medical Care*. Chicago. University of Chicago Press, 1933.
5. Klein MW, et al. Problems of measuring patient care in the outpatient department. *Journal of Health and Human Behavior* 1961;2:138-44.
6. Kohl SG. *Perinatal Mortality in New York City: Responsible Factors*. Cambridge, Harvard University Press, 1955.

Este estudio, patrocinado por la Academia de Medicina de Nueva York, fue un examen de las fichas relativas a una muestra representativa de defunciones perinatales en la ciudad de Nueva York, hecho por un comité de expertos. Se reconocieron las defunciones prevenibles y se identificaron los "factores de responsabilidad", incluidos los errores de juicio y técnica cometidos por el médico. La incidencia de esos errores tuvo que ver también con el tipo de servicio hospitalario, servicio profesional y hospital e indicó la relación existente entre la estructura y el resultado, modificada por las características de la población servida.

7. Shapiro S, et al. Further observations on prematurity and perinatal mortality in a general population and in the population of a prepaid group practice medical care plan. *American Journal of Public Health* 1960;50:1304-17.
8. Lipworth L, Lee JAH, Morris JN. Case fatality in teaching and nonteaching hospitals, 1956-1959. *Medical Care* 1963;1:71-6.
9. Rice CE, et al. Measuring social restoration performance of public psychiatric hospitals. *Public Health Reports* 1961;76:437-46.
10. Lembcke PA. Medical auditing by scientific methods. *Journal of the American Medical Association* 1956;162:646-55. (Los apéndices A y B fueron suministrados por el autor).

Éste es quizá el documento en que mejor se describen los conceptos básicos y métodos del sistema de un alto grado de estructuración formulado por Lembcke para revisar las fichas de los hospitales. También se incluye un ejemplo del extraordinario efecto que puede tener una "revisión externa" de esta clase en la práctica quirúrgica en un hospital.

11. Kelman HR, Willner A. Problems in measurement and evaluation of rehabilitation. *Archives of Physical Medicine and Rehabilitation* 1962;43:172-81.
12. McDermott W, et al. Introducing modern medicine in a navajo community. *Science* 1960;131:197-205 y 280-7.
13. Simon HA. *Administrative Behavior*. Nueva York: The Macmillan Company, 1961;62-6.
14. Hutchinson GB. Evaluation of preventive services. *Journal of Chronic Diseases* 1960;11:497-508.
15. James G. *Evaluation of Public Health. Report of the Second National Conference on Evaluation in Public Health*. Ann Arbor: Universidad de Michigan, Facultad de Salud Pública, 1960;7-17.
16. Weinerman ER. Appraisal of medical care programs. *American Journal of Public Health* 1950;40:1129-34.
17. Goldmann F, Graham EA. *The Quality of Medical Care Provided at the Labor Health Institute, St. Louis, Missouri*. St. Louis: The Labor Health Institute, 1954.

Éste es un buen ejemplo de un método de evaluación basado en características estructurales. En este caso, estas incluyen la disposición y el equipo de las instalaciones físicas, la competencia y

*Las referencias bibliográficas se reproducen en su formato original con las notas del autor

estabilidad del personal médico, las disposiciones tomadas con fines de continuidad del servicio centrado alrededor del médico de familia, la programación y duración de las consultas clínicas, el asunto tratado en el examen inicial, la atención prestada a la medicina preventiva y la idoneidad de las fichas médicas.

18. Peterson OL, et al. An analytical study of North Carolina general practice: 1953-1954 (2ª parte). *The Journal of Medical Education* 1956;31:1-165.

Este estudio, que ya es clásico, se distingue por la extraordinaria atención prestada a los métodos y por la minuciosa exploración de la relación existente entre la clasificación de la calidad y las características de los médicos, incluso su formación y sus métodos de práctica. Los resultados de éste y de otros estudios en los que se ha empleado el mismo método llevan a formular preguntas básicas sobre la práctica general que se considera tradicional en este y en otros países.

19. Clute KF. *The General Practitioner: A Study of Medical Education and Practice in Ontario and Nova Scotia*. Toronto: University of Toronto Press, 1963, capítulos 1, 2, 16, 17 y 18.

Puesto que este estudio se basa en el método formulado por Peterson y colaboradores, ofrece una excelente oportunidad de examinar la naturaleza general de la relación entre las características de los médicos y la clasificación de la calidad. Además, el lector de este volumen elegantemente escrito puede disfrutar de una descripción detallada de la práctica general en los dos campos de estudio.

20. Kroeger HG, et al. The office practice of internists, I. The feasibility of evaluating quality of care. *The Journal of the American Medical Association* 1965;193:371-6.

Ésta es la primera de una serie de documentos basados en el estudio de la práctica de los afiliados a la Sociedad de Medicina Interna de Nueva York. En este documento se presentan resultados relativos a la integridad de las fichas de los consultorios, su idoneidad para juzgar la calidad y el grado de concordancia entre los resúmenes de fichas preparados por los médicos y por otro personal bien adiestrado. No se presentan los juicios emitidos respecto de la calidad de la atención prestada. Otros documentos de esta serie publicados en el *Journal of the American Medical Association* tienen que ver con el número de pacientes (23 de agosto), sus características (13 de septiembre), las actividades profesionales distintas de la atención de pacientes particulares (11 de octubre) y los antecedentes y la forma de ejercicio (1 de noviembre).

21. Kilpatrick GS. Observer error in medicine. *Journal of Medical Education* 1963;38:38-43. Véase una bibliografía útil sobre los errores de los observadores en Witts LJ, ed. *Medical Surveys and Clinical Trials*. Londres: Oxford University Press, 1959;39-44.

22. Rosenfeld LS. Quality of medical care in hospitals, *American Journal of Public Health* 1957;47:356-65.

Este estudio comparativo de la calidad de la atención en cuatro hospitales, cuidadosamente formulado, se centra en los problemas de los métodos de evaluación de la calidad. Ofrece importante información sobre el uso de patrones normativos y empíricos, la fiabilidad y el sesgo de los juicios basándose en un análisis por medio de cuadros, la correlación de los defectos del registro con los de la práctica y la homogeneidad de la clasificación de la calidad en cada clase de diagnóstico y entre una y otra.

23. Makover HB. The quality of medical care: Methodological survey of the medical groups associated with the health insurance plan

of New York. *American Journal of Public Health* 1951;41:824-32.

Éste es quizá el primer informe publicado respecto de un programa de estudios de la calidad de la atención en grupos médicos que han suscrito algún contrato con el Plan de Seguro Médico de la Zona Metropolitana de Nueva York, que tiene una estructura administrativa pero está orientado hacia la investigación. Lamentablemente, gran parte de este trabajo se mantiene inédito. Una característica particular de este documento está en que describe y presenta los resultados de una evaluación simultánea de la estructura (políticas, organización, administración, finanzas y actividades profesionales) y el proceso (evaluación de una muestra de fichas clínicas).

24. Daily EF, Morehead MA. A method of evaluating and improving the quality of medical care. *American Journal of Public Health* 1956;46:848-54.

25. Fitzpatrick TB, Riedel DC, Payne BC. Character and effectiveness of hospital use. En: Mc Nerney WJ et al., *Hospital and Medical Economics*. Chicago, Hospital Research and Education Trust, American Hospital Association, 1962;495-509.

26. Morehead MA, et al. *A Study of the Quality of Hospital Care Secured by a Sample of Teamster Family Members in New York City*. Nueva York: Universidad de Columbia, Facultad de Salud Pública y Medicina Administrativa, 1964.

Este estudio y otro de la misma índole²⁸ realizan una función social y administrativa de suma importancia al documentar los casos en que a menudo resulta ser inadecuada la atención recibida por los miembros de un sindicato por medio de fuentes tradicionales. Estos estudios ayudan mucho a comprender las relaciones existentes entre las características de los hospitales y los médicos y la calidad de la atención que prestan. Se tienen en cuenta diversas clasificaciones de los médicos por campo de especialización y privilegios de internado de pacientes, así como la clasificación de los hospitales por dueño, afiliación a una facultad de medicina, autorización para ofrecer formación en residencia y acreditación. Los efectos interactivos de algunas de estas variables también son objeto de estudio. Además, el segundo de los dos estudios²⁶ se centra en asuntos de metodología, incluso el muestreo representativo de los casos de hospitalización frente al muestreo crítico y la fiabilidad de la evaluación de las fichas por diferentes jueces.

27. Morehead MA. Comunicación personal.

28. Ehrlich J, Morehead MA, Trusell RE. *The Quantity, Quality and Costs of Medical and Hospital Care Secured by a Sample of Teamster Families in the New York Area*. Nueva York: Universidad de Columbia, Facultad de Salud Pública y Medicina Administrativa, 1962.

29. Maloney MC, Trusell RE, Elinson J. Physicians choose medical care: A sociometric approach to quality appraisal. *American Journal of Public Health* 1960;50:1678-86.

Este estudio representa un ingenioso método de evaluación por medio de los "juicios emitidos por iguales" en lo que se cree que es una situación particularmente reveladora: la clase de atención que se escoge para el médico o sus propios familiares. Algunas de las características de los médicos y cirujanos seleccionados incluyeron antiguas relaciones personales y profesionales, especialización reconocida y afiliación a una facultad de medicina. Una valiosa información incidental indica que si bien 9 de cada 10 médicos declararon que toda persona debe tener su propio médico, 4 de cada 10 dijeron que tenían a alguien que consideraban

su médico particular y sólo 2 de cada 10 habían consultado a su médico particular en el transcurso del último año.

30. Georgopoulos BS, Mann FC. *The Community General Hospital*. Nueva York: The Macmillan Company, 1962.

El estudio de la calidad notificado en varios capítulos de este libro se basa en la tesis de que si uno desea averiguar cuál es la calidad de la atención prestada, todo lo que hay que hacer es preguntárselo a las personas que la prestan, de manera directa o indirecta. Aunque los médicos pueden considerar que esta es una idea bastante ingenua, la estabilidad y coherencia interna de los resultados notificados en el presente estudio indican que este método merece consideración más cuidadosa. En un segundo estudio de una muestra nacional de hospitales generales se tratará de confirmar la validez de las opiniones de los entrevistados al compararlas con índices seleccionados de actividades profesionales en cada hospital. Los resultados se esperarán con gran interés.

31. El Sr. Arnold D. Kaluzny, uno de los estudiantes del autor, le ayudó a acuñar esta palabra.
32. Georgopoulos BS, Tannenbaum AS. A study of organizational effectiveness. *American Sociological Review* 1957;22:534-40.
33. Evans LR, Bybee JR. Evaluation of student skills in physical diagnosis. *Journal of Medical Education* 1965;40:199-204.
34. Rimoldi HJA, Haley JV, Fogliatto H. *The Test of Diagnostic Skills*. Loyola Psychometric Laboratory, Publicación N° 25, Chicago: Loyola University Press, 1962.

Este estudio es de interés porque se centra en una situación en que se realizan experimentos controlados para estudiar el desempeño de los estudiantes de medicina y de los médicos. Más interesante aún es el intento por abordar la cuestión del valor o la utilidad de las medidas en materia de diagnóstico de una forma sistemática y rigurosa. Si bien este estudio en particular no parece contribuir mucho a la comprensión de la calidad de la atención, al parecer vale la pena seguir este método general.

35. Williamson JW. Assessing clinical judgment. *Journal of Medical Education* 1965;40:180-7.

Éste es otro ejemplo de la evaluación de la actuación clínica tomando como punto de partida una situación basada en un experimento artificial. El aspecto del trabajo más digno de mencionarse es la asignación de cierto valores ("útil" o "nocivo") a un conjunto de medidas de diagnóstico y terapia y la formulación de medidas de "eficiencia", "pericia" y "competencia", basándose en las actividades seleccionadas por el sujeto para abordar el caso experimental. Con este método se detectaron diferencias de desempeño entre un médico y otro. Un resultado inesperado fue la falta de diferencias sistemáticas por edad, formación o clase de ejercicio en los grupos estudiados hasta ahora.

36. Eisleer CW, Slee VN, Hoffmann RG. Can the practice of internal medicine be evaluated? *Annals of Internal Medicine* 1956;44:144-61.

Los autores discuten el uso de índices de los cuales se podrían hacer inferencias sobre la calidad del tratamiento quirúrgico y médico. Los índices descritos incluyen informes de patología de los tejidos en casos de apendectomía, pacientes diabéticos cuya concentración sanguínea de azúcar no se ha determinado y a quienes no se han tomado radiografías y casos de neumonía en los que no se ordenaron radiografías de pulmones. Un resultado sorprendente notificado en este documento, y en otros basa-

dos en el mismo método, se centra en la tremenda variación de la frecuencia de esos índices de "actividad profesional" por médico y hospital.

37. Furstenberg FF, et al. Prescribing as an index to quality of medical care: A study of the Baltimore City medical care program. *American Journal of Public Health* 1953;43:1299-309.
38. Peterson OL, Barsamian EM. *An Application of Logic to a Study of Quality of Surgical Care*. Documento presentado en el Quinto Simposio Médico de la IBM, Endicott, Nueva York, 7-11 de octubre de 1963.

En este documento y otro de la misma índole³⁹ se presenta una descripción bastante completa del método basado en el "árbol lógico" para la evaluación de la calidad. Se dan ejemplos de los sistemas lógicos empleados para el estudio del síndrome de Stein-Leventhal y de fibromioma uterino. No se ofrecen datos sobre los resultados empíricos en cuya consecución se ha empleado este método.

39. Diagnostic Performance. En: Jacquez JA, ed. *The Diagnostic Process*. Ann Arbor: The University of Michigan Press, 1964; 347-62.
40. Lembcke PA, Johnson OG. *A Medical Audit Report*. Los Angeles: Universidad de California, Facultad de Salud Pública, 1963 (documento mimeografiado).

Este documento representa una extensión del método de revisión de la atención médica de Lembcke a las diferentes clases de diagnóstico médico y a un gran número de intervenciones quirúrgicas. Aunque este volumen es un compendio de datos sin elaborar en su mayoría, un cuidadoso estudio puede ofrecer información sobre el método empleado por el autor y sus limitaciones.

41. Lembcke PA. A scientific method for medical auditing. *Hospitals* 1959;33:65-71(16 de junio);65-72(1 de julio de 1959).
42. Está por determinarse la dimensión del conjunto de variables en las que se incorporan esos patrones.
43. Huntley RR, et al. The quality of medical care: Techniques and investigation in the outpatient clinic. *Journal of Chronic Diseases* 1961;14:630-42.

Este estudio ofrece un ejemplo de la aplicación de un procedimiento ordinario de análisis por medio de cuadros para verificar la calidad de la atención prestada en el departamento ambulatorio de un hospital docente. Con bastante frecuencia se dejaron de realizar los procedimientos ordinarios y se determinó que no se habían seguido los casos de anomalías. Al parecer, un procedimiento modificado de análisis por medio de cuadros ocasionó una importante reducción del porcentaje de anomalías excluidas del seguimiento.

44. Peterson, et al., loc. cit., trataron de confirmar los índices de ponderación por medio del procedimiento de análisis de los factores. Los que son de importancia matemática se citan en su nota de las págs. 14-15.
45. Mainland D. *Calibration of the Human Instrument*. Notes from a Laboratory of medical Statistics, 1964;81 (documento mimeografiado).
46. *Comité Mixto del Real Colegio de Obstetricia y Ginecología y el Comité de Investigaciones sobre Población*. Maternity in Great Britain. Londres: Oxford University Press, 1948.

47. Yankauer A, Goss KG, Romero SM. An evaluation of prenatal care and its relationship to social class and social disorganization. *American Journal of Public Health* 1953;43:1001-10.
48. Wylie CM. Participation in a multiple screening clinic with five-year follow-up. *Public Health Reports* 1961;76:596-602.
49. Comisión de Actividades Profesionales y Hospitalarias. *Medical Audit Study Report 5: Primary Appendectomies*. Ann Arbor: Comisión de Actividades Profesionales y Hospitalarias, octubre de 1957.
50. Simon AJ. Social structure of clinics and patient improvement. *Administrative Science Quarterly* 1959;4:197-206.
51. Lockward HJ, Lundberg GAF, Odoroff ME. Effect of intensive care on mortality rate of patients with myocardial infarcts. *Public Health Reports* 1963;78:655-61.
52. Bakst JN, Marra EF. Experiences with home care for cardiac patients. *American Journal of Public Health* 1955;45:444-50.
53. Muller JN, Tobis JS, Kelman HR. The rehabilitation potential of nursing home residents. *American Journal of Public Health* 1963;53:243-7.
54. Estos estudios también incluyen datos sobre las relaciones existentes entre las características estructurales y los extremos relativos a procedimiento. Son ejemplos de ello el efecto de la estructura clínica en el número de consultas ambulatorias⁵⁰ y el efecto de un programa de atención en el hogar en la hospitalización⁵².
55. Getting VA, et al. *Research in Evaluation in Public Health Practices*. Documento presentado en la 92ª Reunión Anual de la Asociación Americana de Salud Pública, Nueva York, 5 de octubre de 1964.
56. Suponiendo que el criterio empleado sea la evaluación directa del proceso, el asunto se convierte en una de las repercusiones que tienen las medidas de fiabilidad en la validez.
57. Ciocco A, Hunt H, Altman I. Statistics on clinical services to new patients in medical groups. *Public Health Reports* 1950;65:99-115.
- Ésta es una aplicación preliminar a la práctica de grupo del análisis de "actividades profesionales" relacionado hoy en día generalmente con la evaluación de la atención hospitalaria. Los índices usados incluyeron el registro del diagnóstico y tratamiento, la práctica de exámenes rectales y vaginales, la práctica de ciertos análisis de laboratorio y el uso de sedantes, estimulantes y otros medicamentos que son causa de uso indebido. Como sucede en los hospitales, los grupos variaron mucho con respecto a esos indicadores.
58. Myers RS. Hospital statistics don't tell the truth. *Modern Hospital* 1954;83:53-4.
59. Aún en el caso de la atención hospitalaria, la unidad apropiada puede incluir la atención prestada antes y después de la hospitalización, así como varias hospitalizaciones³.
60. Cordero AL. The determination of medical care needs in relation to a concept of minimal adequate care: An evaluation of the curative outpatient services in a rural health center. *Medical Care* 1964;2:95-103.
61. Butterworth JS, Reppert EH. Auscultatory acumen in the general medical population. *Journal of the American Medical Association* 1960;174:32-4.
62. Evans LR, Bybee JR. Evaluation of student skills in physical diagnosis. *Journal of Medical Education* 1965;40:199-204.
63. Fattu NC. Experimental studies of problem solving. *Journal of Medical Education* 1964;39:212-25.
64. John ER. Contributions to the study of the problem solving process. *Psychological Monographs* 1957;71.
65. Duncan CP. Recent research in human problem solving. *Psychological Bulletin* 1959;56:397-429.
66. Fattu NA, Mech E, Kapos E. Some statistical relationships between selected response dimensions and problem-solving proficiency. *Psychological Monographs* 1954;68.
67. Stolurow LM, et al. The efficient course of action in "trouble shooting" as a joint function of probability and cost. *Educational and Psychological Measurement* 1955;15:462-77.
68. Ledley RS, Lusted LB. Reasoning foundations of medical diagnosis. *Science* 1959;130:9-21.
69. Lusted LB, Stahl WR. *Conceptual Models of Diagnosis*. En: Jacquez JA, ed. *The Diagnostic Process*. Ann Arbor: The University of Michigan Press, 1964;157-74.
70. Edwards W, Lindman H, Phillips LD. *Emerging Technologies for Making Decisions*. En: Newcomb TM, ed. *New Directions in Psychology, II*. Nueva York: Holt, Rinehart & Winston, Inc., 1965;261-325.