

Métodos de Avaliação de Serviços e Programas de Saúde

Methods of Evaluation of Services and Health Programs

MARISTELA CHITTO SISSON¹

RESUMO

A incorporação de processos de avaliação no trabalho em saúde pode significar mudanças no estado de saúde das pessoas e na direção da consolidação do modelo assistencial do SUS. No setor público, existe um interesse maior pela avaliação de programas e recentemente, através da realização do pacto de indicadores da atenção básica, é possível identificar a monitorização, na população, das principais situações de saúde que demandam intervenções. Com o objetivo de relacionar conceitos avaliativos com os principais desenhos epidemiológicos utilizados em processos de avaliação, desenvolveu-se um levantamento documental. A análise identificou o crescimento do uso instrumental da avaliação, por poder alcançar mais diretamente os que investem nas intervenções em saúde. A maior premissa a ser considerada é que o objetivo de uma avaliação é influenciar uma decisão. Identificou também uma tipologia que se distribuiu entre desenhos experimentais, quasi-experimentais e não experimentais, indicando, por exemplo, o uso do método epidemiológico em estudos avaliativos para medida de atributos de qualidade de ações de saúde, como efetividade, eficiência, eficácia e outros, além de possibilitar a construção de indicadores utilizados para medidas de avaliação. Há convergência entre os autores no sentido de esclarecer que a seleção do tipo de estudo depende necessariamente dos interesses da avaliação a ser realizada, não havendo, portanto, um desenho melhor que o outro em todas as circunstâncias.

DESCRIPTORIOS

Avaliação em saúde; métodos. Serviços de saúde. Metodologia.

SUMMARY

The inclusion of processes of evaluation in health work can mean changes in a person's state of health and also towards the consolidation of the model of assistance provided by the SUS (Unified Health System). In the public sector a great interest exists in the evaluation of these programs. Recently, by means of the execution of the basic attendance indicators agreement, it has become possible to monitor the population and identify the main health situations that require interventions. A documental survey was developed with the purpose of associating evaluation concepts to the main epidemiological outlines used in evaluation processes. The analysis identified the growth of instrumental use with the evaluation, since it is able to directly reach those who invest in the interventions in health. The greatest assumption to be considered is that the purpose of an evaluation is to influence a decision. It also identified a typology that was distributed through experimental, quasi-experimental and non-experimental designs. This gave a greater indication of the advantages in using the epidemiological method to evaluate the quality of attributes used in health actions, such things as: effectiveness, efficiency, efficacy and many others. In addition to allowing the construction of indicators used for evaluation measures. There is convergence among the authors in clarifying that the choice of the type of study depends on the interests of the evaluation to be performed and therefore, there is no single design, better than the others in any circumstances.

DESCRIPTORS

Evaluation in health; methods. Health services; Methodology.

¹ Professora do Programa de Mestrado em Ciências da Saúde - Universidade do Sul de Santa Catarina - Florianópolis - Santa Catarina - Brasil.

A escolha em estabelecer um sistema de cuidados de saúde é uma responsabilidade da sociedade, e em função disso, diferentes conceitos de saúde têm realizado aplicações diferentes na construção destes sistemas. Planejar para que este sistema realmente beneficie os cidadãos, se relaciona com a discussão da qualidade como direito básico de saúde e pode adquirir maior visibilidade através da implementação de processos de avaliação em saúde, cuja incorporação pode significar mudanças no estado de saúde das pessoas e na consolidação do modelo assistencial (VUORI, 1991).

Com o objetivo de relacionar alguns conceitos avaliativos e sua aplicação aos principais desenhos de investigação utilizados em processos de avaliação de programas e serviços de saúde, se desenvolveu este estudo, realizado através de técnicas de levantamento documental em referências da área.

A área de Avaliação de Serviços, Programas e Tecnologias de Saúde tem a década de 50 como marco referencial, em função de ser uma fase de expansão desenvolvida na saúde mundialmente, com grande crescimento da tecnologia médica e dos serviços de saúde, mas também com importante incremento de custos e de responsabilidade por parte dos Estados no financiamento da saúde, demandando a disposição de serviços efetivos e eficazes.

A partir daí, as primeiras investigações foram baseadas principalmente na utilização de métodos epidemiológicos para avaliação da eficácia terapêutica e diagnóstica.

Na década de 60, DONABEDIAN (1998, 1990), autor clássico na área de avaliação em saúde, propôs uma avaliação da qualidade da atenção médica por meio de um modelo que sistematizava os atributos que traduzem a qualidade nos serviços: eficácia, efetividade, eficiência, otimização, aceitabilidade, legitimidade e equidade. Definiu também sua constituição, através de etapas de produção que poderiam ser medidas: de estrutura, processo e resultados.

Nas últimas décadas as investigações na área da saúde tornaram-se mais complexas. Um diálogo entre o biológico e o social, baseado em uma nova compreensão dos determinantes de saúde, estimulou uma nova reflexão sobre os fundamentos do trabalho em saúde, sobre a necessidade de manejo de outros métodos e horizontes de tempo diferentes, e sobre a maneira de conceber os fenômenos que se estuda. As pesquisas ultrapassaram a avaliação de técnicas ou procedimentos isolados e agregaram referenciais de outras disciplinas científicas, a fim de ampliar as abordagens de avaliação,

Establishing a health care system is a responsibility of society and in view of that, different health concepts have implemented different applications in the construction of these systems. Planning so that this system really benefits the citizens, is related to a discussion about quality being a basic right for health and it may acquire greater visibility through the implementation of processes for evaluation in health. The incorporation of these may mean changes in the health conditions of the people and in a consolidation of the model of assistance (VUORI, 1991).

This study was developed through a documental survey, with the purpose of associating some evaluation concepts and their application to the main designs of investigation used in processes of evaluation in health programs and services.

The area of Evaluation of Health Services, Programs and Technologies has the decade of the fifties as its reference mark. This was a phase of expansion which developed in the health area, worldwide and with a great growth in medical technology and health services. In addition, there were important increases in costs and responsibility on the part of the States in the financing of health, requiring the availability of effective and efficient services.

From then on, the first investigations were based mainly on the use of epidemiological methods for the evaluation of the therapeutic and diagnostic effectiveness.

In the sixties, DONABEDIAN (1998, 1990), a respected author in the area of health evaluation, proposed an evaluation in the quality of the medical care by means of a model that systemized the attributes that translate the quality in the services: efficacy, effectiveness, efficiency, optimization, acceptability, legitimacy and equality. It also defined its constitution, through stages of production that could be measured in structure, process and results.

In the last few decades the investigations in the area of health became more complex. A dialogue between the biological and the social, based on a new understanding of health determinants has stimulated a new reflection in the grounds of the work being done in health. There is necessity in handling other methods and different time horizons when conceiving the phenomena being studied. The research has exceeded the isolated procedures or evaluation techniques and added referentials from other scientific disciplines in order to increase the evaluation approaches, such as those studies involving the equality or the social

como as que envolvem a equidade ou a legitimidade social de ações de saúde (CONTANDRIOPOULOS, 1998).

Além disso, a tarefa da avaliação se torna hoje também mais complexa, em função da expansão dos alvos das ações de saúde pública para muito além das doenças infecciosas, incluindo doenças crônicas, violência, patógenos emergentes e os contextos sociais que influenciam disparidades na saúde (BARRETO e CARMO, 1994; POSSAS, 1994; CDC/MMWR, 1999).

O uso instrumental da avaliação está em crescimento, por alcançar mais diretamente os que investem nas intervenções em saúde (HABICHT; VICTORA, VAUGHAN, 1999). A maior premissa a ser considerada é que o objetivo de uma avaliação é influenciar uma decisão.

A complexidade e a precisão da avaliação vão depender de quem toma as decisões e dos tipos de decisão gerados como consequência dos achados. É geralmente entendido que outros fatores pesam mais que resultados de avaliações quantitativas nas decisões finais sobre programas. Entretanto, inferências da avaliação quantitativa seriam pertinentes para as decisões, se essas avaliações tivessem uma utilidade aplicável, e assim os avaliadores desenhariam pesquisas para responder questões específicas de interesse de quem toma as decisões.

Os tipos de inferência requeridos na tomada de decisão, de grande importância para os desenhos de avaliação, estão longe de aparecer na literatura epidemiológica, apesar de serem realizadas majoritariamente. Em função disso, cresce o interesse pela metodologia qualitativa de avaliação.

Até bem pouco tempo atrás não se encontrava descrita na literatura uma proposta básica de modelo de avaliação para programas especificamente de saúde pública. Em 1999 o Grupo de Trabalho em Avaliação do Centro de Controle e Prevenção de Doenças (CDC) de Atlanta, U.S.A. apresentou um modelo de avaliação com o propósito de implementar a compreensão da avaliação de programas e para facilitar a integração da avaliação por todo o sistema de saúde pública (CDC, 1999).

Programa é descrito pelo CDC como um objeto de avaliação que pode ser qualquer ação organizada de saúde pública. Preconiza integrar a avaliação à operação rotineira do programa, quando existir ênfase, na prática, em uma avaliação progressiva que envolve toda a equipe de trabalho e os interessados em seu investimento, e não somente *experts* em avaliação. Indica, também, que a prática da avaliação complementa a gerência no aporte das informações necessárias à implementação e efetividade do programa.

legitimacy of health actions (CONTANDRIOPOULOS, 1998).

Besides, the task of doing the evaluation has currently become more complex too, due to the increase of the target actions of public health ranging far beyond infectious diseases and now including chronic diseases, violence, emerging pathogens and the social contexts that influence disparities in health (BARRETO and CARMO, 1994; POSSAS, 1994; CDC/MMWR, 1999).

The instrumental use of the evaluation is growing because it reaches those who invest in the interventions in health more directly (HABICHT; VICTORA, VAUGHAN, 1999). The major assumption for consideration is that the purpose of an evaluation is to influence a decision.

The complexity and precision of the evaluation will depend on those who take the decisions and the types of decision generated as a result of the findings. Generally, it is understood that other factors weigh more than the results of quantitative evaluations in the final decisions on programs. However, inferences of the quantitative evaluation would be relevant for the decisions should these evaluations have an applicable utility and thus, the evaluators would outline specific research to answer the questions of interest of those who make the decisions.

The types of inference required in the taking of decisions are of great relevance for the evaluation designs but are far from appearing in epidemiologic literature, even though they are carried out in great volumes. As a result of that, the interest grows for the qualitative methodology of evaluation.

Until recently there wasn't a basic proposal for a model of evaluation for programs of public health described in literature. In 1999 the Work Group in an Evaluation of the Center of Disease Prevention and Control (CDC) of Atlanta, U.S.A. presented a model of evaluation with the purpose of implementing the understanding of the evaluation of programs and facilitating the integration of the evaluation throughout the entire system of public health (CDC, 1999).

The program is described by the CDC as an object of evaluation that can be any organized action of public health. It recommends integrating the evaluation to the routine operation of the program, when there is emphasis, in practice, in a gradual evaluation that involves the whole work team and the parties interested in its investment, and not only the evaluation *experts*. It also indicates that the practice of the evaluation complements the management in the provision of the information required for the implementation and effectiveness of the program.

O modelo de avaliação do Grupo de Avaliação do CDC se baseia no conceito de que “uma avaliação de programas efetiva é uma maneira sistemática de implementar e de assumir responsabilidades em ações de saúde pública, por envolver procedimentos que são úteis, factíveis, éticos e precisos”. É considerado uma ferramenta prática e não prescritiva, desenhada para sumarizar e organizar os elementos essenciais do processo de avaliação. O modelo consiste em seis etapas principais, que devem ser seguidas em qualquer avaliação. Estas etapas são consideradas interdependentes, portanto não devem ser interpretadas em uma ordem linear, apesar de que existe uma certa ordem de execução de cada uma e de que cada passo anterior fundamenta o progresso subsequente. São elas: envolver os interessados; descrever o programa; focar o desenho da avaliação; coletar evidências confiáveis; justificar as conclusões e assegurar o uso e compartilhar lições. Os padrões de avaliação utilizados são os de Utilidade, Factibilidade, Propriedade e Precisão.

A prática dos trabalhadores da saúde está repleta de ações informais de avaliação, que dão conta das pequenas modificações necessárias no dia a dia. Quando se amplia, porém, a necessidade de tomar uma decisão, ou quando são necessárias mudanças nos rumos de um programa, o emprego de procedimentos de avaliação explícitos, formais e justificáveis se tornam essenciais (CDC, 1999) e, neste caso, a epidemiologia desempenha um papel fundamental.

Como forma de ampliar o quadro avaliativo e obter análises mais completas de outras dimensões dos programas e serviços de saúde, utilizam-se pesquisas avaliativas com diferentes métodos.

MÉTODOS

A ênfase colocada na avaliação permanente de processos e resultados nos serviços de saúde demanda a contribuição dos instrumentos considerados do campo específico da Epidemiologia (CECÍLIO, 1995).

Raramente são incorporados nos desenhos das investigações dos estudos epidemiológicos tradicionais, as características dos sistemas e serviços de saúde e dos processos diagnósticos e terapêuticos (enquanto elementos participantes da construção de seu ponto de partida, as doenças). Quando a avaliação aparece como um dos instrumentos de intervenção, torna-se clara a articulação entre clínica, epidemiologia, administração e planejamento, devendo dar conta da qualidade dos resultados obtidos pelos serviços sobre os processos de saúde-doença.

The model of evaluation of the CDC Evaluation Group is based on the concept that “an effective evaluation of programs is a systematic way of implementing and undertaking responsibilities in the actions of public health, since it involves procedures that are useful, feasible, ethical and precise”. It is considered a practical, not prescriptive tool, designed to summarize and organize the essential elements of the evaluation process. The model consists of six main stages that must be followed in any evaluation. These stages are considered interdependent, therefore they should not be interpreted in a linear sequence, notwithstanding there is a specific order of execution of each one and that each previous step is the base for the subsequent progress. They are: involving the interested parties; describing the program; focusing the design of the evaluation; gathering reliable evidences; justifying the conclusions and assuring the use and sharing lessons. The standards of evaluation used are the Utility, Factibility, Property and Precision standards.

The practice of health workers is full of informal actions of evaluation that take care of the small modifications required in daily routine. However, when the need to take a decision is increased, or when changes are required in the way the program is developed, the use of explicit, formal and justifiable procedures of evaluation become essential (CDC, 1999) It is in this circumstance that the epidemiology plays a basic role.

Evaluation researches with different methods are used as a means of increasing the evaluation framework and getting more complete analyses of other dimensions of the health programs and services.

METHODS

The emphasis placed on the permanent evaluation of processes and results in health services demand the contribution of the instruments considered in the specific field of Epidemiology (CECÍLIO, 1995).

The characteristics of the health systems and services, as well as those of the diagnostic and therapeutic processes are rarely incorporated into the designs of the investigations of traditional epidemiologic studies (as elements that participate in the construction of their starting point, the diseases). When the evaluation appears as one of the instruments of intervention, the articulation between clinics, epidemiology, administration and planning becomes clear and this should deal with the quality of the results obtained by the services on the health-disease processes.

O papel da metodologia como saber que informa o diagnóstico e a avaliação exerce um fundamental papel nestes processos. Sua contribuição para a avaliação de serviços ou programas de saúde está colocada principalmente em dois campos, que serão comentados a seguir em alguns aspectos: o campo das medidas dos resultados da avaliação, quase sempre indicadores epidemiológicos de saúde da população e o desenvolvimento de desenhos epidemiológicos de investigação para avaliação.

1. Indicadores e seu uso na Avaliação da Atenção Básica em Saúde

No Sistema Único de Saúde brasileiro, desde 1998 estão sendo selecionados indicadores para serem trabalhados como medidas de acompanhamento de metas em saúde planejadas nos municípios, componentes do chamado Pacto de Indicadores da Atenção Básica. Foi concebido e instituído pela Portaria GM/MS 3.925 de 1998 como um instrumento nacional de monitoramento das ações e serviços de saúde referentes à atenção básica.

Os indicadores definem os atributos pertinentes às questões ou ao foco da avaliação. Em função de que traduzem conceitos gerais relativos ao programa, seu contexto, e seus efeitos esperados, em medidas específicas que podem ser interpretadas, proporcionam uma base de coleta de evidências que é válida e confiável para as intenções de uso da avaliação.

Os indicadores adicionam critérios que serão usados para julgar o programa e, além disso, refletem aspectos do programa significativos de serem monitorados.

Algumas questões, porém, apontam a insuficiência dos indicadores como medidas avaliadoras únicas. Do ponto de vista da utilização dos indicadores para medida do estado de saúde da população, existe um questionamento a respeito da necessidade estatística de sintetização da informação, que os reduz a uma unidade avaliativa. Esta sintetização pode não refletir a diversidade contida no que se denomina "o novo paradigma da saúde" (HANSLUWKA, 1987). A epidemiologia, ao discutir as causas das doenças e sua incidência, tem a possibilidade de identificar e definir um quadro mais complexo de variáveis que representem um espectro mais amplo dos determinantes do processo saúde-doença. O conceito de saúde utilizado é que vai dirigir esta definição e, portanto, o paradigma da saúde e não o da doença é o que deve ser promovido (DEVER, 1988).

The role of methodology as informed knowledge from the diagnosis and the evaluation exerts a basic role in these processes. Its contribution to the evaluation of health services or programs is placed mainly in two fields, which in some aspects will be commented on subsequently. The fields of measures that result from evaluation are almost always epidemiologic indicators of the health of the population and the development of epidemiologic designs of investigation for evaluation.

1. Indicators and their use in the Evaluation of Basic Health Care

Since 1998, indicators are being chosen in the Brazilian Unified Health System, to work as follow-up measures for planned health goals in the cities, components of the so-called Basic Attendance Indicators Agreement. These indicators were conceived and instituted by Ordinance GM/MS 3,925 from 1998 as a national instrument for the monitoring of the actions and services regarding the basic health care.

The indicators define the attributes related to the issues or the focus of the evaluation. Since they translate general concepts related to the program, their context and their expected effects. They use specific measures that may be interpreted, providing a basis for the collection of evidence both valid and reliable for use within the evaluation.

The indicators add criteria that will be used to judge the program and in addition to that, they reflect meaningful aspects of the program to be monitored.

Some issues however, point to the insufficiency of the indicators as being the only measure of evaluation. From the point of view of using indicators for measuring the health conditions of the population, there is a question regarding the statistical necessity of synthesizing the information, which reduces them to one evaluation unit. Such synthesis may not reflect the diversity contained in what is referred to as "the new health paradigm" (HANSLUWKA, 1987). When the epidemiology discusses the causes of the diseases and their incidence, there is the possibility of identifying and defining a more complex framework of variables that would represent a wider range of determinants of the health-disease process. The health concept used will guide such definition and therefore, the health paradigm, not the disease paradigm, is the thing that should be developed (DEVER, 1988).

A definição e identificação das dimensões da saúde e a definição da equidade se constituem em aspectos fundamentais onde deveriam centrar-se as medidas de saúde. Na prática, porém, estas medidas são dificilmente desenvolvidas.

2. O desenho da investigação

Uma vez que o avaliador sabe o que vai estudar, o passo seguinte consiste em como estudar. Traça um plano para escolher as pessoas que serão estudadas, distribui a investigação no tempo e estabelece procedimentos para a coleta de dados. Como forma de ampliar o quadro avaliativo e obter análises mais completas de outras dimensões dos programas e serviços de saúde, pode utilizar pesquisas avaliativas com diferentes metodologias.

Ao traçar seu desenho, o pesquisador poderá propor um experimento tradicional controlado ou um dos desenhos quasi-experimentais, menos formalizados. Poderá circunscrever sua investigação a um projeto ou examinar os resultados de certo número de projetos que tenham as mesmas metas fundamentais. Poderá trabalhar com variáveis da ciência social ou efetuar uma análise econômica dos custos do programa e dos benefícios que produz.

A escolha do desenho tem implicações para o que contará como evidência, como esta evidência será coletada e que tipos de inferências podem ser feitas, inclusive a validade interna e externa das conclusões (MCQUEEN; ANDERSON, 1999). Além disso, as decisões metodológicas esclarecem como a avaliação vai operar, ou seja, em que extensão os participantes do programa estarão envolvidos; como as fontes de informação serão selecionadas; que instrumentos de coleta de dados serão usados; quem coleta os dados, que softwares serão utilizados e quais serão os métodos mais apropriados para análise, síntese, interpretação e apresentação dos dados. Em função de que cada opção tem seu viés metodológico e suas limitações, as avaliações que trabalham com um mix de métodos são geralmente mais efetivas (COOK; REICHARDS, 1979); (GREEN; CARACELLI, 1997).

A seguir, representando a maioria dos estudos desenvolvidos no país, são apresentados os vários desenhos metodológicos de aproximação quantitativa possíveis de serem utilizados em processos de avaliação.

The definition and identification of the dimensions of health and the definition of equity are the basic aspects in which health measures should focus. In practice however, these measures are hardly developed.

2. The design of the investigation

Once the evaluator knows what he will study, the following step is composed of how to study it. A plan is outlined to choose the people who will be studied; the investigation is distributed and procedures are established for the gathering of data. As a means of increasing the evaluation framework and getting more complete analyses of other dimensions of the health programs and services, evaluation researches with different methodologies can be used.

While drawing its design, the researcher may propose a controlled traditional experiment or one of the less formal quasi-experimental designs. He/she may restrict his/her investigation to one project or examine the results of certain number of projects that have the same basic goals. He/She may be able to work with variables from the social sciences or make an economic analysis of the costs of the program and the benefits it produces.

The choice of the design has implications for what will count as evidence, how this evidence will be gathered and what types of inferences may be made, including the internal and external validity of the conclusions (MCQUEEN; ANDERSON, 1999). Moreover, the methodological decisions clarify how the evaluation will work, that is, to what extent the participants of the program will be involved; how the sources of information will be chosen; what instruments will be used for the gathering of data; who gathers the data; what software will be used and which will be the most appropriate methods for analysis, synthesis, interpretation and presentation of the data. Due to the fact that each option has its methodological bias and limitations, the evaluations that work with a mix of methods are generally more effective (COOK; REICHARDS, 1979); (GREEN; CARACELLI, 1997).

Next, the majority of the studies developed in the country are represented by several methodological designs of quantitative approach that are likely to be used in evaluation processes.

2.1 Desenho experimental

O desenho clássico para as avaliações tem sido o modelo experimental, ainda que mais como exortação do que na prática. Este modelo utiliza um grupo experimental e outro de controle. Calculam-se as diferenças e se qualifica o êxito do programa quando o grupo experimental melhora mais que o grupo controle.

Os desenhos aleatórios são uma maneira efetiva de descartar a possibilidade de que algo que não seja o programa esteja causando as melhorias que se observam. Entretanto, o experimento controlado frequentemente não se pode realizar na prática, nos campos de ação concreta. Muitas vezes os controles são difíceis de encontrar, e os únicos possíveis podem estar muito espalhados, muito distantes ou não estão dispostos a cooperar com um programa que não vai lhes oferecer nada em troca. Alguns autores atacaram nos últimos anos o modelo experimental, por não ser factível e também por ser contraproducente: requer que o programa se mantenha constante, não facilitando sua contínua melhora; é somente útil para tomar decisões depois do final do ciclo do projeto, e não durante seu planejamento e execução; como controla muitas condições, resulta asséptico para se generalizar ao mundo real (STUFFLEBEAM, 1966); (GUBA e LINCOLN, 1989).

Outros estudos (HABICHT, VICTORA, VAUGHAN, 1999) referem que estes fatos críticos não são inevitáveis ao experimento controlado, que não requer um programa estável, pois se desejamos conhecer os efeitos de um programa em condições de desenvolvimento, as técnicas experimentais devem ser aplicadas com imaginação, em vez de estudar um projeto por vez, fazer estudos comparados, repetidas medições periódicas e desenho e análise mais sutis e complexos.

O problema fundamental consiste em ajustar o desenho da investigação aos fins do estudo. O objeto do estudo de avaliação deverá determinar seu desenho. O desenho experimental serve, sobretudo, para realização de fins que ficam fora do programa imediato e que necessitam de um grande rigor na investigação, como por exemplo, recomendar à escala nacional um projeto.

Mesmo em condições favoráveis, a experiência sugere que este tipo de desenho não se dá frequentemente. Rara vez os programas se realizam para a conveniência do avaliador e suas necessidades e são um fator importante nos ajustes (e no orçamento do programa). Em tais circunstâncias, é alentadora a tendência a um desenho quase experimental.

2.1 Experimental Design

The classic design for the evaluations has been the experimental model, even though more as concept than in practice. This model uses an experimental group and a control group. The differences are calculated and the success of the program is defined as being when the experimental group has more improvement than the group of control.

Random designs are an effective means to rule out the possibility that something not connected to the program caused the observed improvements. However, the controlled experiment frequently cannot be carried out in practice within the fields of concrete action. Many times the controls are difficult to find, and the only possible ones can be dispersed, very distant or are unwilling to cooperate with a program that will not offer them anything in return. Some authors have attacked the experimental model in recent years because it is not feasible and also because it is counterproductive; it requires the program to stay constant, thus not facilitating its continuous improvement; it is only useful to take decisions after the end of the cycle of the project, but not during its planning and execution as it controls many conditions, and many consider it idealistic to generalize the real world (STUFFLEBEAM, 1966); (GUBA and LINCOLN, 1989).

Other studies (HABICHT, VICTORA, VAUGHAN, 1999) mention that these critical facts are not inevitable to the controlled experiment, that it does not require a stable program. If we intend to know the effects of a program under conditions of development, the experimental techniques must be applied with imagination, instead of studying one project at a time, studies should be compared, periodic measurements repeated and a more complex and subtle approach to design and analysis.

The basic problem consists of adjusting the design of the investigation to the purposes of the study. The object of the evaluation study must determine its design. The experimental design serves, over all, to achieve goals that are outside the immediate program and that require rigid strictness in the investigation, as for example; when recommending a project to be implemented at the national level.

Even in favorable conditions, the experience suggests this type of design does not occur frequently. Rarely programs are performed for the convenience of the evaluator and his/her needs and they are an important factor in the adjustments (and in the budget of the program). In such circumstances, the inclination to a quasi-experimental design is encouraging.

2.2. *Desenhos quasi-experimentais*

São desenhos que não cumprem os requisitos estritos dos experimentos. À diferença do desenho experimental, que protege contra quase todas as ameaças possíveis à validade interna, os desenhos quasi-experimentais, geralmente deixam de controlar uma ou várias delas. Têm a vantagem de ser práticos quando as condições impedem uma verdadeira experimentação; mas não são, de forma alguma, pura e simplesmente experimentos improvisados. Possuem forma e lógica próprias. Ao reconhecer previamente o que fazem e o que controlam e as possíveis más interpretações dos resultados, permitem ao avaliador tirar conclusões com todo cuidado. Os procedimentos devem ser tão rigorosos quanto o dos experimentais. São eles:

Desenho de séries temporais: É um dos quasi-experimentais mais atrativos. Consiste em uma série de medidas, a intervalos periódicos, antes que o programa se inicie e medidas contínuas depois que tenha se terminado. Desta forma é possível averiguar se as medidas tomadas imediatamente antes e depois do programa representam uma continuidade de pautas ou normas anteriores ou se mostram uma mudança decisiva.

Séries temporais múltiplas: É um desenho ainda melhor que o anterior. Consiste em encontrar um grupo ou instituição semelhante e fazer medidas periódicas ao longo do tempo. Assim poderá se proteger também contra efeitos de confusão da história.

Grupo controle pareado: Neste caso não existe seleção aleatória ao programa, nem controle como haveria em um verdadeiro experimento. Utilizam-se como controles, indivíduos ou grupos sadios, mas com características semelhantes. As medidas se fazem antes e depois em ambos os grupos e se comparam os resultados. Este desenho talvez seja o desenho mais comum, na prática. A questão primordial é como fazer para que o grupo de comparação seja o mais semelhante possível ao grupo experimental.

Apesar do prestígio, poder e simetria que goza o desenho experimental, o desenho quasi-experimental possui freqüentemente a virtude de que é factível de ser realizado. O avaliador que realiza um experimento controlado, e tropeça com obstáculos e recorre a práticas viciosas, é menos produtivo que o que adapta seus desenhos, de acordo com a boa prática de investigação, às possibilidades.

2.3. *Desenhos não experimentais*

Às vezes é impossível inclusive o uso de

2.2. *Quasi-experimental designs*

These are designs that do not meet the strict requirements of the experiments. Unlike the experimental design, which protects against almost every possible threat to the internal validity, the quasi-experimental designs generally do not control one or several of them. They have the advantage of being practical when the conditions prevent a true experimentation; but they are not by any means, purely and simply, improvised experiments. They have their own logic and format. As they recognize beforehand what they do and control and the possible poor interpretations of the results, they allow the evaluator to make his/her own conclusions with all due care. The procedures must be as rigorous as those of the experimental model. They are:

Time series design: It is one of the most attractive quasi-experimental designs. It consists of a series of measures, at periodic intervals, before the program commences and continuously measures after it has finished. This way, it is possible to verify if the measures taken immediately before and after the program represent a continuity of patterns or previous rules or if they show a decisive change.

Multiple time series: It is an even better design than the previous one. It consists in finding a group or similar institution and taking periodic measures throughout the duration of time. This way it may also be protected against the effects of confusion of the history.

Paired control group: In this case there is neither random selection to the program, nor a control as there should be in a true experiment. Individuals or healthy groups, but with similar characteristics, are used as controls. The measures are taken before and after in both the groups and the results are compared. Maybe this design is the most common design in practice. The primordial question is what to do so that the group of comparison is as similar as possible to the experimental group.

Despite the prestige, power and symmetry the experimental design has, the quasi-experimental design frequently has the virtue of being feasible to carry out. The evaluator, who carries out a controlled experiment and then stumbles upon obstacles and appeals against the vicious practices, is less productive than the one who adapts his/her designs, in accordance with good investigation practices, to the possibilities.

2.3. *Non-experimental designs*

Sometimes it is impossible to use the quasi-

desenhos quasi-experimentais e se necessita recorrer a um dos três desenhos não experimentais comuns: *estudo antes-depois* de um só programa, *estudo depois-unicamente* dos participantes do programa e *estudo depois-unicamente dos participantes e de controles* não tomados ao azar. A debilidade inerente a estes desenhos é de que não controlam muitas das explicações rivais (ou que as mudanças causadas foram relativas a algo diferente do programa).

Projeto antes-depois

Este desenho, além de circunscrever-se a medidas anteriores e posteriores ao projeto, pode tomar uma série de medidas dos participantes enquanto se vai desenvolvendo o programa, e avaliar se os passos sucessivos estabelecidos como hipótese vão sendo alcançados. Estes dados podem complementar-se com uma análise qualitativa intensiva do que foi alcançado pelo programa, com o objetivo de entender quais são as relações que existem entre a intervenção do programa e o progresso dos participantes, por exemplo.

Unicamente depois (ex post facto)

Neste tipo de desenho os benefícios são menores, pois saber como as coisas estão indo para os participantes do programa depois da intervenção, somente teria sentido se existirem boas razões para temer qual seria a condição do programa sem a intervenção. Frequentemente este não é o caso. Ainda que possam existir indícios da experiência passada ou de investigações anteriores, rara vez se pode ter em conta todas as contingências do presente.

Unicamente depois com grupo de comparação

Este desenho dá mais força ao tipo unicamente-depois, agregando um grupo de comparação que seja o mais semelhante possível aos usuários do programa.

Apesar dos problemas que possam existir na seleção do grupo de comparação, inevitavelmente o estudo mostrará diferenças importantes com o grupo participante. Tratamentos estatísticos são utilizados para “igualar” os grupos, como por exemplo, a covariância. Existem controvérsias, porém, deste uso, alguns considerando que seu uso crônico *infra-ajusta* e outros considerando que este *infra-ajuste*, se acontecer, é de pouca monta, se considerarmos o grande volume de efeitos do programa. Outros sugerem ampliar a coleta de dados e passar a um desenho quasi-experimental.

experimental designs and it is required to use one of the three common non-experimental designs: *before-after study* of one single program, *only-after study* of the participants of the program and *only-after study with the participants and control groups* not taken randomly. The debility inherent to these designs is that they do not control many of the rival explanations (or that the changes caused concerned something different from the program).

Before-after Project

This design, besides being limited to previous and subsequent measures of the project, may take a series of measures from the participants while the program is being developed, and evaluates whether the successive steps defined as a hypothesis are being achieved. Such data can be complemented by an intensive qualitative analysis of what was achieved by the program, with the purpose, for instance, of understanding what the relations between the intervention of the program and the progress of the participants are.

Only after (ex post facto)

In this type of design the benefits are less significant, since knowing how things are going for the participants of the program after the intervention, would only make sense if there were good reasons to fear what would be the condition of the program without the intervention. That is normally not the case. Even if there are indications of the past experience or of previous investigations, it is rare to find a time when all the contingencies of the present can be accepted.

Only after with comparison group

This design provides more strength to the *only after* type, adding a group of comparison that is as similar as possible to the users of the program.

Irrespective of the problems that may exist in the selection of the group of comparison, the study will inevitably show important differences in relation to the participant group. Statistical treatments are used to “equal” the groups, as for example, the covariance. Nonetheless, there are controversies regarding its use. Some consider its chronic use promotes an *under-adjustment* and others consider that such *under-adjustment*, if any, is of little significance, should we consider the great volume of effects of the program. Others suggest increasing the gathering of data and moving to a quasi-experimental design.

Na avaliação de resultados de um programa, ao buscarmos medir sua efetividade, impacto ou análise de custo, poderemos construir o seguinte quadro resumo:

In the evaluation of results of one program, when we seek to measure its effectiveness, impact or cost analysis, we are able to create the following summary chart:

Quadro 1 - Tipos de desenhos avaliativos de acordo com tipos de análise dos resultados.
Chart 1 – Types of evaluation designs according to the types of analysis of the results.

Análise da avaliação dos resultados Analysis of the evaluation of the results	Desenhos de avaliação Evaluation designs
Efetividade Effectiveness	Longitudinal comparado / Longitudinal, compared Comparação pré-pós / Pre/post comparison Indivíduo como própria comparação / Individual with own comparison
Impacto Impact	Grupo de comparação hipotético / Hypothetical group of comparison Pareado (coorte) / Paired (cut) Experimental/control / Experimental/control
Custo-benefício Cost-benefit	Grupo de comparação hipotético / Hypothetical group of comparison Pareado (coorte) / Paired (cut) Experimental/control / Experimental/control

Fonte: SHALOCK (2001) / Source: SHALOCK (2001).

A seguir, o quadro II mostra sinteticamente o comportamento destes desenhos em relação a seus atributos:

Next, chart II synthetically shows the behavior of such designs in relation to their attributes:

Quadro 2 - Atributos da avaliação e comportamento dos desenhos de investigação
Chart 2 – Attributes of the evaluation and behavior of the investigation designs.

Confiabilidade Reliability	Precisão Accuracy	Desenho da avaliação Evaluation design	Comparabilidade Comparability	Generalização Generalization
Alto High ↑ Baixo Low	Alto ↑ Baixo	Experimental/control Experimental/control Pareado (corte) Paired (cut) Grupo de comparação hipotético Hypothetical group of comparison Longitudinal comparado Longitudinal compared Comparação pré-pós Pre/post comparison Indivíduo como própria comparação Individual with own comparison	Poucos pressupostos Few assumptions ↓ Muitos pressupostos Many assumptions	Alto High ↓ Low Baixo

Fonte: SHALOCK (2001) / Source: SHALOCK (2001).

As opções apresentadas, no entanto, como já referido anteriormente, não são determinantes do melhor desenho a ser escolhido, pois nenhum é melhor que outro em todas as circunstâncias.

Os métodos de avaliação devem ser selecionados para prover a informação necessária aos interessados, ou seja, vão depender diretamente do objeto da avaliação. O que distingue, assim, uma metodologia de

The options presented, however, as has previously mentioned, are not determinants of the best design to be chosen, since no one is better than the other in all the circumstances.

The methods of evaluation must be chosen to provide the necessary information to the interested parties, in other words, they will directly depend on the object of the evaluation. Thus, what differentiates one

avaliação da outra, não são os métodos, mas as questões que são indexadas e os valores que são promovidos (GREENE, 1994).

APLICAÇÃO

No próprio campo da avaliação em saúde estão as maiores áreas de aplicação da epidemiologia contemporânea: a avaliação dos problemas de saúde da população, o ordenamento de prioridades de programas de serviços de saúde de acordo com as necessidades e a avaliação de programas e serviços em termos de seu impacto na saúde da população (IBRAHIM, 1983).

No Brasil, a integração dos processos de avaliação às operações rotineiras dos programas e dos serviços de saúde recém se inicia, principalmente no campo da atenção básica de saúde. Sua participação nos processos decisórios ainda é bastante marginal, tanto no setor público como no privado.

No setor público tem existido um interesse maior pela avaliação de programas, em função da participação da Saúde Coletiva (NOVAES, 1996) e é reconhecida a necessidade estratégica de avaliar e monitorar o desempenho da Atenção Primária de Saúde nas proposições que buscam aprimorar a qualidade da atenção à saúde (SCHRAIBER *et al.*, 1996).

Mais recentemente, novas necessidades históricas recolocaram as possibilidades de participação efetiva dos serviços de saúde na determinação dos níveis de saúde e de doença da população. Iniciou-se no setor público, a partir daí, o debate sobre a avaliação das ações, instituindo-se em 2003 a Política Nacional de Avaliação para Atenção Básica em Saúde (BRASIL, 2005), ampliando-se o leque de objetivos e de indicadores de saúde.

As ações de avaliação e monitoramento propostas por esta Política, como componente estratégico do SUS, visam compartilhar julgamentos e processos decisórios voltados para a melhoria da qualidade da atenção à saúde, e acompanhamento, na população, das principais situações de saúde que demandam intervenções.

Em 2006 promulgou-se o “Pacto pela Saúde”, instrumento mais complexo, capaz de monitorar, nos municípios, entre as prioridades definidas, a redução da mortalidade infantil e materna, o controle das doenças emergentes e endemias (como dengue e hanseníase) e a redução da mortalidade por câncer de colo de útero e da mama.

Entre 2005 e 2006 desenvolveram-se também estudos de Linha de Base da Estratégia de Saúde da Família (BRASIL, 2004) em algumas capitais brasileiras. Estas pesquisas avaliativas, em geral de base epidemiológica, lançaram mão de variados desenhos de investigação e se voltaram à produção de evidências para mudanças nos rumos da Atenção Básica.

methodology of evaluation from another, are not the methods, but the questions that are indexed and the values that are promoted (GREENE, 1994).

APPLICATION

In the field of health evaluation there are major areas of application for contemporary epidemiology with the evaluation of the health problems of the population, putting program priorities into order for health services, in accordance with the necessities and the evaluation of programs and services in terms of their impact in the health of the population (IBRAHIM, 1983).

In Brazil, the integration of evaluation processes to the routine operations of the health services and programs has just started, mainly in the field of basic health care. Its participation in the decision-making processes is still quite insignificant, both in the public and in the private sector.

There has been greater interest in the public sector for the evaluation of programs due to the participation of the Collective Health (NOVAES, 1996) and it is recognized the strategic necessity of evaluating and monitoring the performance of the Primary Health Care in the propositions that seek to improve the quality of health care (SCHRAIBER *et al.*, 1996).

More recently, new historical needs have restored the possibilities of effective participation of the health services in the determination of the levels of health and disease of the population. From that point on, the public sector started the debate on the evaluation of these actions, creating in 2003, the National Policy of Evaluation for the Basic Health Care (BRASIL, 2005), increasing the range of goals and health indicators.

The actions of evaluation and monitoring proposed by this Policy, as an strategic component of the SUS (Brazilian Unified Health System), aim at sharing decisions and decision processes aimed at the improvement of the quality of the health care and follow-up in the population one of the main situations of health that require interventions.

In 2006 the “Pact for Health” was enacted, a more complex instrument, capable of monitoring in the municipalities, among the priorities set were: the reduction of infant and maternal mortality, the control of the emerging and endemic diseases (such as dengue fever and leprosy) and the reduction of mortality due to colon and breast cancer.

Between 2005 and 2006 Baseline studies of the Family Health Strategy (BRASIL, 2004) were also developed in some Brazilian cities. These evaluation researches, in general, of an epidemiological basis, made use of several designs of investigation and turned towards the production of evidences to perform changes in the directions of the Basic Care.

A implementação e fortalecimento de novos modelos assistenciais (como o Programa de Saúde da Família) supõem uma institucionalização do processo de planejamento e adoção de métodos modernos de gestão, o que poderia levar a uma estruturação possível do modelo assistencial a partir do modelo epidemiológico de diagnóstico das necessidades. A obtenção do melhor impacto na solução dos problemas seria realizada através de uma combinação de recursos e tecnologias capazes de garantir alto grau de efetividade (PAIM, 1992).

Estas estratégias representam o início da integração de processos de avaliação sistemáticos ao SUS, relacionando atividades analíticas com gestão de intervenções programáticas, visando melhorar a qualidade das ações na rede de atenção básica.

The implementation and strengthening of new health care models (such as the Family Health Program) presume an institutionalization of the planning process and adoption of modern management methods, which may take an organization of the health care model based on the epidemiological model of diagnosis of the necessities. The achievement of the most positive impact on problem solution would take place by means of a combination of resources and technologies capable of guaranteeing a high level of effectiveness (PAIM, 1992).

These strategies represent the beginning of the integration of systematic processes of evaluation to the SUS, associating analytical activities to the management of programmatic interventions, aiming at improving the quality of the actions in the system of basic health care.

REFERÊNCIAS

References

1. BARRETO ML, CARMO EH. Situação de Saúde da população brasileira: tendências históricas, determinantes e implicações para as políticas de saúde. *Infor. Epid. SUS*, 3(3/4):7-34, 1994.
2. BRASIL. Ministério da Saúde. Projeto de Expansão e Consolidação do Saúde da Família, Proesf (Componente III). Termo de Referência para o Sub-componente D, Linha de Ação 3. Estudos de Linha de Base "Desenvolvimento de Estudos de Linha de Base nos Municípios Selecionados para o Componente 1"; Brasília, 2004. Disponível em <http://www.saude.gov.br> Acesso em 25 de janeiro de 2005.
3. BRASIL. Ministério da Saúde. Secretaria de Atenção à Saúde. Departamento de Atenção Básica. Coordenação de Acompanhamento e Avaliação. Avaliação na Atenção Básica em Saúde: caminhos da institucionalização. Coordenação técnica: Instituto de Saúde Coletiva da Universidade Federal da Bahia e Instituto Materno Infantil Prof. Fernando Figueira, IMIP.- Brasília: Ministério da Saúde; 2005, 36p.
4. CECÍLIO LCO. A Epidemiologia na Avaliação dos Serviços de Saúde: a discussão da qualidade. *Saúde e Sociedade*, 4 (1/2):115-117, 1995.
5. CDC.CENTER FOR DISEASE CONTROL AND PREVENTION. Framework for program evaluation in public health. *MMWR* 48 (no.RR-11).
6. CONTANDRIOPOULOS AP. Pode-se construir modelos baseados na relação entre contextos sociais e saúde? *Cad. Saúde Pública*, 14(1):199-204. Rio de Janeiro, 1998.
7. COOK, TD, REICHARDS CS, EDS. Qualitative and quantitative methods in avaluation research. *Sage research progress series in evaluation*, vol 1. Beverly Hills: Sage Publictions, 1979.
8. DEVER GEA. Epidemiologia enfocando a prevenção. In: *A epidemiologia na administração dos serviços de saúde*. São Paulo: Pioneira, 1988.
9. DONABEDIAN A. The seven Pilars of Quality. *Arch Pathol. Lab. Med.*, 114: 1115-118, 1990.
10. DONABEDIAN A. The Quality of care. How can it be assessed? *JAMA*. 260(12):1793-1748, 1998.
11. GREEN JC, CARACELLI V, EDS. Advances in mixed-methods evaluation: the challenges and benefits of integrating diverse paradigm. *New Directions for Program Evaluation*. Vol 74; San Francisco: Jossey-Bass, 1997.
12. GREENE JC. Qualiative Program evaluation. In: Denzin, N. K. & Lincoln, Y.S. *Handbook of qualitative research*. Newburry pnk: Sage Publications, 1994.
13. GUBA EG e LINCOLN YS. *Four generation evaluation*. Newbury Park: Sage Publications, 1989.
14. HABICHT JP, VICTORA CG, VAUGHAN JP. Evaluation desings for adequacy, plausibility and probability of public health programmer performance and impact. *Inter. Jour. Epidem*, 28(s.n.):10-18,1999.
15. HANSLUWKA HE. Measuring the health status of a population: current state of the art. *Popul Bull UN* (23-24):56-75,1987.
16. IBRAHIM MA. Epidemiology: application to health services. *J. Hlth Adm. Educ*. 1(1): 37-69, 1983.
17. MQUEEN DV, ANDERSON LM. *What count as evidence? Issues and debates on evidence relevant to the evaluation of community health promotion programs*. In: Rootman I, Goodstadt. Copenhagen, Denmark: WHO (Euro), 1999.
18. NOVAES HMD. Epidemiologia e avaliação em serviços de atenção médica: novas tendências na pesquisa. *Cad. Saúde Pública*, 12(suppl.2):S7-S12,1996.
19. PAIM JS. A reforma sanitária e os modelos assistenciais: a epidemiologia na organização dos serviços de saúde. In: Anais do II Congresso Brasileiro de Epidemiologia. Belo Horizonte: Abrasco, 1992.
20. POSSAS C. Prioridades sanitárias, Ciência e Tecnologia. In: Anais da I Conferência Nacional de Ciência e Tecnologia em Saúde. 24-28 de outubro, Brasília, Ministério da Saúde, 1994.
21. SCHRAIBER LB, NEMES MIB, MENDES GONÇALVES RB. *Saúde do Adulto: programas e ações na unidade básica*. São Paulo: Editora Hucitec, 1996.
22. SHALOCK RL, PRABHAKAR NR. *Outcome-based evaluation*. New York: Kluwer Academic/ Plenum Publishers, 2001.
23. STUFFLEBEAM DL. A depth study of the evaluation requeriment. *Theory into Practice*, 5(3):121-134,1966.
24. VUORI H. A qualidade da saúde. *Divulg. Saúde para Debate*, 3(s.n.):17-25, 1991.

CORRESPONDÊNCIA

Correspondence

Maristela Chitto Sisson
 Travessa Erotides Maria Oliveira, 116
 88062-170 Florianópolis – Santa Catarina – Brasil

E-mail

mcs@mbox1.ufsc.br
 rebrasa@ccs.ufpb.br