

**ORIENTAÇÕES E INSTRUMENTOS
PARA A ORGANIZAÇÃO DO ACESSO
À ATENÇÃO PRIMÁRIA À SAÚDE**

Eugênio Vilaça Mendes

Trabalho realizado para o
Conselho Nacional de Secretários de Saúde (CONASS)

Outubro de 2016

I. OS MICROSSISTEMAS CLÍNICOS

Esse conceito deriva de observações feitas por Quinn (1992) que mostraram que as empresas de maior sucesso progressivamente aprenderam a focar suas ações em menores unidades replicáveis. A teoria de Quinn sugere que as organizações se constituem dessas menores unidades replicáveis para conectar a competência central da empresa com os seus clientes.

Godfrey et al (2004) transplantaram o conceito de pequenas unidades replicáveis para o campo das organizações de saúde denominando-as de unidades de microssistemas clínicos com foco nos microssistemas clínicos da atenção à saúde que se caracterizam por terem uma população de pessoas usuárias, os prestadores de serviços, os processos finalísticos e de apoio e a tecnologia de informação com um propósito comum.

Os microssistemas clínicos proveem um referencial para organizar, medir e melhorar a oferta dos cuidados de saúde. Uma característica essencial dos microssistemas clínicos é que eles estão onde se dão os encontros clínicos entre as unidades de saúde e as pessoas usuárias.

Os microssistemas clínicos são definidos como uma unidade entre um pequeno grupo de pessoas que trabalham juntas de forma regular para prestar cuidados de saúde e uma população que recebe estes cuidados. Esses microssistemas clínicos buscam objetivos claros, estruturam processos, compartilham um ambiente de informações e produzem serviços e cuidados que podem ser medidos por resultados de desempenho.

Os microssistemas clínicos evoluem ao longo do tempo e se articulam em sistemas ou organizações maiores. Como quaisquer sistemas adaptativos vivos, os microssistemas clínicos devem desenvolver o trabalho, satisfazer as necessidades de seus colaboradores e se manterem como unidades clínicas.

Os microssistemas clínicos são parte de organizações maiores ou macrossistemas. Eles têm no seu centro as pessoas usuárias. A camada seguinte são os microssistemas em que se dão as interfaces entre os profissionais de saúde e as pessoas usuárias e a terceira camada é dos macrossistemas (GODFREY et al, 2004).

Os microssistemas clínicos operam segundo um modelo que descreve a sequência de passos para desenvolverem um senso de consciência: cria a consciência de um trabalho realizado num microssistema; identifica alguma disfuncionalidade para compreender que

uma mudança é possível; conecta o trabalho com aqueles que se beneficiariam dele; implanta alguma mudança estratégica ou de melhoria; estabelece medidas de desempenho para aqueles que se beneficiam do microsistema e para o seu funcionamento e transparência; e procura trabalhar integradamente com insumos e produtos, com microsistemas pares, com a população e com o macrosistema.

Um microsistema clínico começa a operar com senso de consciência quando alguém faz a pergunta: você poderia me descrever a forma como seu microsistema funciona na perspectiva dos seus profissionais e da população? Um novo senso de consciência de como os profissionais trabalham juntos (ou não) começa a desenvolver-se e essas pessoas começam a identificar as disfuncionalidades de seus processos de trabalho e ações são tomadas para eliminá-las ou reduzi-las.

O microsistema clínico compreende, então, que as mudanças são possíveis e que podem ser feitas ali mesmo, sem necessitar de autorização de alguém externo a esse microsistema. Ou seja, é possível mudar ali mesmo o microsistema clínico. Algumas perguntas são fundamentais como por que estamos fazendo o que fazemos? Tornar o objetivo do microsistema clínico explícito é importante porque constitui a sua base de trabalho.

Os microsistemas clínicos têm sido considerados o lugar em que as pessoas usuárias, as famílias e os profissionais de saúde se encontram e em que o cuidado é prestado, a qualidade e a segurança são garantidas, a eficiência e a inovação são alcançadas e o moral da equipe e a satisfação das pessoas usuárias são altos. Esses microsistemas clínicos incluem o pessoal administrativo, as tecnologias, os processos e os padrões de informação, comportamentos e resultados.

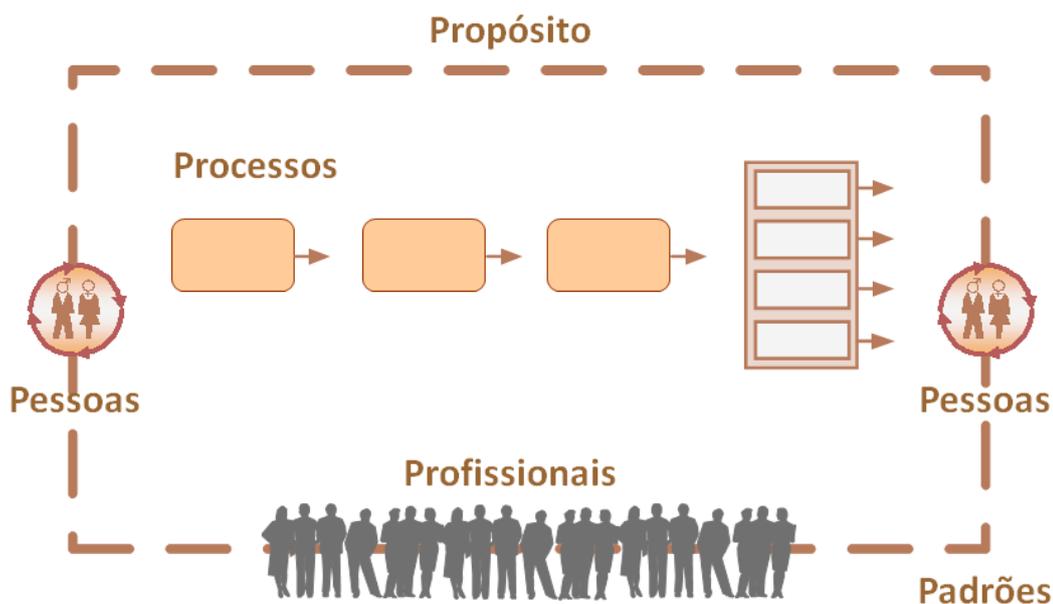
Os microsistemas clínicos são as partes que constituem os sistemas de atenção à saúde. A qualidade do sistema deriva da qualidade de seus microsistemas que entregam os cuidados, conforme a equação: qualidade do sistema de saúde = qualidade do microsistema 1 + qualidade do microsistema 2 + qualidade dos microsistemas 3-n.

Cada microsistema clínico constitui um complexo e dinâmico sistema adaptativo em que interagem vários elementos que atuando de forma interdependente permitem produzir resultados superiores. Esses elementos são: no campo da liderança: liderança competente e suporte organizacional; no campo da equipe de saúde: foco na equipe, educação permanente e interdisciplinaridade; no campo das pessoas usuárias: foco nas pessoas e na comunidade; no campo do desempenho: melhoria de processos e resultados de

desempenho; no campo da informação, tecnologia da informação e comunicação (GODFREY et al, 2004).

O diagrama apresentado na Figura 1 retrata um microsistema clínico.

Figura 1. O diagrama de um microsistema clínico



Fonte: Godfrey et al (2004)

Como se vê na Figura 1 os microsistemas clínicos têm cinco componentes, os 5 P's: propósito, pessoas usuárias, profissionais, processos que o microsistema utiliza na prestação dos cuidados e padrões que caracterizam a funcionalidade do microsistema.

II. OS PRINCÍPIOS DA ORGANIZAÇÃO DO ACESSO AOS SERVIÇOS DE SAÚDE

Um Comitê do Institute of Medicine definiu seis princípios básicos para o acesso aos serviços de saúde (INSTITUTE OF MEDICINE, 2015).

O primeiro princípio é o balanceamento entre a oferta e a demanda. Esse princípio implica uma avaliação quantitativa, formal e contínua, da oferta e da demanda para a provisão do acesso oportuno aos serviços de saúde.

O segundo princípio é o atendimento imediato das pessoas usuárias e suas famílias. Cada pessoa ou família que busca atenção à saúde deve ser questionada sobre qual é o problema e o que o serviço de saúde poderia ajudar naquele momento (Como podemos ajudá-lo hoje?). O atendimento imediato implica, no acesso à atenção primária à saúde, a definição de uma meta de atendimento no mesmo dia ou em dois dias no máximo, a menos que a pessoa deseje ser atendida em outro momento e nos atendimentos programados pela equipe (os bons backlogs).

O terceiro princípio são as preferências das pessoas usuárias. As pessoas usuárias são estimuladas a expressar suas preferências durante o cuidado interativo. As estratégias do acesso aos serviços de saúde devem objetivar a melhoria da experiência das pessoas usuárias e a responder às suas preferências. A captação das preferências das pessoas usuárias pode ser feita por análise de dados dessas preferências usando métodos como narrativas das pessoas e surveys (TOUSSAINT e BERRY, 2013; BRANDENBURG et al, 2015).

O quarto princípio é a utilização de uma atenção definida, sob medida, pelas necessidades das pessoas usuárias. As opções de atendimentos no mesmo dia deveriam incluir vários métodos testados para atender às necessidades e aos interesses das pessoas usuárias. Isso pode incluir atendimentos alternativos aos encontros clínicos individuais face a face como atendimentos à distância ou virtuais e os atendimentos por profissionais não médicos.

O quinto princípio é a utilização de planos de contingência. Cada serviço de saúde deve desenvolver planos de contingência para responder aos desequilíbrios, previstos e imprevistos, entre oferta e demanda.

O sexto princípio é a avaliação contínua. As métricas de acesso aos serviços de saúde devem ser coletadas, avaliadas e publicadas pelas unidades de saúde. Os dados coletados devem servir para avaliar as atividades diárias e monitorar os agendamentos em períodos de tempo específicos. Além disso, para avaliar o desempenho comparativo, padrões e benchmarks em acesso e tempos de espera na atenção primária à saúde devem ser desenvolvidos, testados e implementados (LITVAK, 2009).

III. ORIENTAÇÕES E INSTRUMENTOS PARA A ORGANIZAÇÃO DO ACESSO À ATENÇÃO PRIMÁRIA À SAÚDE

Mendes (2016) elaborou um roteiro metodológico a ser seguido para a organização do acesso à atenção primária à saúde. Esse roteiro é detalhado aqui utilizando-se de orientações e instrumentos operacionais e obedecendo-se aos diferentes enfoques utilizados no acesso aos cuidados primários.

Esses instrumentos e orientações foram buscados na literatura sobre atenção primária à saúde e sobre o acesso aos serviços de saúde, especialmente em Mendes (2012), Institute of Medicine (2015), Litvak (2015), Murray (2015), Institute of Healthcare Improvement (2016) e Mendes (2016).

1. Análise situacional do acesso numa equipe de saúde da família

Um bom exemplo de um microssistema clínico é uma equipe da estratégia da saúde da família. Essa equipe constitui uma pequena unidade replicável que é parte de uma unidade de atenção primária à saúde que, por sua vez, é parte de um macrossistema, por exemplo, uma Secretaria Municipal de Saúde. Essa equipe tem um propósito, tem vários profissionais de saúde, tem uma população adstrita a ela, opera por processos e adota padrões que marcam a funcionalidade do sistema.

Não necessariamente a realidade de um microssistema clínico é a sequência de passos apresentados no modelo descrito para o desenvolvimento do senso de consciência. Por exemplo, durante a organização de um microssistema clínico para implementar uma nova forma de acesso os seus membros identificam uma nova disfuncionalidade e estabelecem atividades de melhoria adicionais.

Os programas de melhoria dos microssistemas clínicos devem começar por uma análise situacional dos 5 P's.

Tomando-se o exemplo de um microssistema clínico constituído por uma equipe de estratégia da saúde da família, com vistas a organizar o acesso, a análise começaria com a definição da equipe, local em que atua, gerente responsável, localização e objetivos principais (propósito).

A análise do segundo P faz-se por meio da afirmativa: conheça profundamente as pessoas usuárias. É preciso criar um quadro preciso da população adstrita a aquela equipe: Quem são essas pessoas? Que recursos elas utilizam? Quando as pessoas buscam os serviços elas os recebem? Alguns pontos de análise são essenciais: Qual é o tamanho do painel de pessoas adstritas a esta equipe? Qual a distribuição etária dessa população adstrita (por exemplo, 0-10; 11-18; 19-45; 45-64; 65-79; 80 ou mais anos) Qual é a composição por sexos feminino e masculino? Quais são as pessoas hiperutilizadoras e quais as razões para essa demanda elevada? Quais são os 10 problemas/diagnósticos mais comuns? Qual o percentual de pessoas com condições crônicas referidas a unidades ambulatoriais especializadas? Qual o percentual de pessoas referidas a unidades de urgência ou emergência? Qual o percentual de pessoas com eventos agudos classificados por riscos? Qual o percentual de pessoas com condições crônicas estratificadas por riscos? Indicadores específicos por condições de saúde segundo as prioridades estabelecidas: por exemplo, hemoglobina glicada, pressão arterial, LDL etc. Qual a experiência de acesso? Forma de acesso (presencial, telefone, internet, retornos programados pela equipe ou programados após rastreamento ou busca ativa) Quanto tempo demorou em dias para ser atendido? Quanto tempo levou desde a chegada a unidade até ser finalizado o atendimento? Foi atendida pelo profissional a que está vinculada? A demanda apresenta variação sazonal (por exemplo, dengue no verão, infecção de via aérea superior no inverno)? Quantas pessoas são atendidas por dia, por turno e por hora de cada turno diário? Quantas pessoas foram atendidas na última semana? Quantas pessoas foram atendidas no último mês? Quantas pessoas foram atendidas por cada membro da equipe?

A análise do terceiro P faz por meio da afirmativa: conheça as pessoas que compõem a equipe de saúde. É preciso ter uma visão ampla dos trabalhadores de saúde. Quem faz o que? Quantas horas estão disponíveis para atendimento às pessoas? Quanto tempo destinam às atividades clínica e não clínicas? Qual a capacidade de oferta (número de atendimentos disponíveis por dia para cada profissional multiplicado pelo número de dias de trabalho por dia/semana/mês/ano)? Qual é a oferta real expressa no número de pessoas que diariamente buscam e recebem atenção dos profissionais? Qual é a demanda real de pessoas atendidas num período de tempo como dia/semana/mês/ano? Qual é a demanda indireta ou externa de atendimentos programados pela equipe? Qual é a demanda direta ou interna feita por telefone ou internet? Qual é a demanda feita hoje para hoje? Qual é a demanda feita hoje para atendimento futuro? Quantos são e que percentuais são os atendimentos segundo perfis de oferta (eventos agudos, condições crônicas não agudizadas, atendimentos a pessoas hiperutilizadoras e às enfermidades, atendimentos às demandas administrativas, atendimentos preventivos, atendimentos domiciliares e

apoio ao autocuidado? Qual a duração média dos atendimentos por diferentes perfis de oferta? Quantos espaços de atendimentos clínicos estão disponíveis? Qual é o moral da equipe de saúde? Quem são os profissionais clínicos e de apoio administrativo segundo sua ocupação? Há a oferta regular de educação permanente para os profissionais da equipe? Há capacitação da equipe para utilizar ferramentas do modelo de melhoria e os ciclos de PDSA? Quão estressante é o trabalho? Recomenda este lugar para se trabalhar?

A análise do quarto P envolve o conhecimento dos processos de trabalho. Há o uso rotineiro de diretrizes clínicas baseadas em evidências (linhas-guia e protocolos clínicos)? Há um mapeamento de processo de cada tipo de perfil de oferta (eventos agudos, condições crônicas não agudizadas, atendimentos a pessoas hiperutilizadoras e às enfermidades, atendimentos às demandas administrativas, atendimentos preventivos, atendimentos domiciliares e apoio ao autocuidado)? Há uma avaliação do tempo que as pessoas chegam à unidade até que a deixam utilizando a ferramenta de análise do ciclo do tempo da pessoa segundo cada tipo de perfil de oferta (eventos agudos, condições crônicas não agudizadas, atendimentos a pessoas hiperutilizadoras e às enfermidades, atendimentos às demandas administrativas, atendimentos preventivos, atendimentos domiciliares e apoio ao autocuidado)? Há oferta regular de consultas médicas? De consultas de enfermagem? De consultas odontológicas? De consultas de outros profissionais da equipe (assistente social, farmacêutico clínico, nutricionista, psicólogo, profissional de educação física, fisioterapeuta e outros)? Os trabalhos da equipe multiprofissional fazem-se de forma interdisciplinar? Há oferta regular de cuidado compartilhado a grupo? De atendimento contínuo? De grupos operativos? De outras tecnologias de trabalho em grupo? De atendimentos por correio eletrônico? De atendimentos por telefone? De atendimentos por videoconferências? Há a utilização rotineira da tecnologia de gestão da condição de saúde? Há a utilização rotineira da tecnologia de gestão de caso? Há a utilização rotineira da tecnologia de auditoria clínica?

A análise do quinto P faz-se por meio do conhecimento dos padrões. Há encontros regulares entre os membros da equipe? Quão frequentes são esses encontros? Esses encontros monitoram os processos e resultados alcançados por meio de um painel de bordo? Esses encontros monitoram a qualidade da atenção? Há uma preocupação em fazer mudanças? Essas mudanças têm melhorado a atenção à saúde e a eficiência do sistema?

2. Faça hoje o trabalho de hoje

Um sistema que faz hoje o trabalho de hoje atende a demanda de cada dia no mesmo dia em que é gerada. Esse é um ponto fundamental do acesso à atenção primária à saúde que deriva do enfoque do acesso avançado.

Nas equipes que utilizam esse sistema os únicos agendamentos que estão marcados no início de um dia de trabalho são aqueles que foram programados previamente pelas equipes para retornos ou por ser uma data mais confortável para as pessoas usuárias (os bons *backlogs*). Isso dá uma maior flexibilidade ao sistema para absorver a demanda diária espontânea.

Fazer hoje o trabalho de hoje requer um compromisso de cada equipe para atender às pessoas a ela vinculadas. E requer, também, que oferta e demanda estejam balanceadas e que o mau *backlog* seja eliminado.

3. O manejo dos *backlogs*

O problema central do acesso à atenção primária à saúde está na existência de *backlogs* que são acumulações de trabalho em função de um modelo de agendar para datas futuras os atendimentos de hoje. Contudo, há que se diferenciarem os *backlogs* em maus *backlogs*, expressos em acumulação de solicitações de atendimentos de pessoas que buscavam a atenção pronta e não conseguiram, dos bons *backlogs*, atendimentos de pessoas nas datas mais confortáveis para elas e atendimentos de retorno programados pelas equipes de saúde.

O manejo dos *backlogs* são diferentes nos maus e nos bons *backlogs*.

Deve ficar claro que nenhum sistema de agendamento pode funcionar adequadamente sem a eliminação do mau *backlog*. Isso é uma pré-condição para a organização do acesso à atenção primária à saúde e deve ser colocado como uma prioridade.

A redução do mau *backlog* acontece quando há uma redução dos terceiros próximos atendimentos disponíveis; quando isso ocorre torna-se possível atender toda demanda no mesmo dia.

É importante considerar-se na organização do acesso à atenção primária à saúde do terceiro próximo agendamento disponível que é definido como o número médio de dias

entre o momento em que uma pessoa usuária solicita um atendimento e o terceiro agendamento disponível para de um profissional da atenção primária à saúde para uma consulta que envolva um exame físico, um exame de rotina ou um exame de retorno. O terceiro próximo agendamento disponível é usado, ao invés do primeiro agendamento disponível, porque reflete com maior sensibilidade a disponibilidade de agendamento que está sujeita a eventos inesperados como cancelamentos de atendimentos.

Alternativamente ao terceiro próximo agendamento disponível pode-se trabalhar com o número absoluto de agendamentos futuros que se enquadrem no conceito de mau *backlog*.

A eliminação do mau *backlog* envolve: medir a sua magnitude pelo terceiro próximo atendimento disponível; se necessário elaborar um plano de redução do *backlog* que tenha uma data final que coincidirá com a implantação de um novo sistema de agendamento; esse plano de redução envolverá todos os profissionais da unidade e das equipes e, muitas vezes, exigirá, temporariamente, uma carga adicional de trabalho e/ou uma redução do tempo médio de atendimento. Como exemplo de eliminação de maus *backlogs* estão os mutirões.

O enfrentamento do mau *backlog* envolve a elaboração de um plano deliberado para sua redução e implica alguns passos.

O primeiro passo é aumentar a oferta. Para isso a equipe deve fazer-se algumas perguntas: Há alguma pessoa agendada para o futuro que eu estou atendendo hoje? Posso atender às necessidades dessa pessoa de modo a eliminar a demanda futura? Essa demanda futura pode ser atendida por telefone ou por internet? Essa demanda futura pode ser atendida por outro membro da equipe? Essa pessoa necessita realmente desse atendimento?

O segundo passo é aumentar a oferta de serviços pela equipe e gradativamente ir criando condições para fazer hoje o trabalho de hoje. Uma forma de fazer isso é aumentar temporariamente os atendimentos incrementando horas adicionais ao final da tarde e ao final de semana. Outra forma é buscar uma redução do tempo médio de atendimento.

O terceiro passo é definir uma data de começo do plano e uma data final quando o *backlog* estará eliminado. É preciso cuidar para que esse tempo não seja muito curto para evitar o estresse da equipe.

O quarto passo é submeter esse plano ao gerente da unidade para garantir o suporte administrativo.

O manejo adequado do bom *backlog* derivado de atendimentos programados pela equipe, normalmente, sob a forma de retornos, tem uma lógica operacional distinta.

Como há uma variação muito grande de demanda aos cuidados primários por dias e horas dos dias e por períodos do ano, deve-se, a partir do diagnóstico destas variações, procurar agendar os atendimentos programados pela equipe para aqueles momentos de menor pressão de demanda. Assim, esses atendimentos devem ser agendados para dias e horas em que há menor pressão de demanda à equipe de saúde.

4. O tamanho do painel

O ajuste do tamanho do painel de pessoas por equipe é uma condição necessária para a racionalização da demanda na atenção primária à saúde. O tamanho do painel deve ser estabelecido com equidade entre as diferentes equipes de modo que não haja equipes com painéis excessivos e equipes com painéis baixos. Isso leva a uma distribuição injusta entre os profissionais de saúde e afeta o ambiente de trabalho de uma organização.

Algumas estratégias podem ser utilizadas para se obterem painéis de tamanho certo. Dentre elas estão: definir um painel inicial e observá-lo ao longo do tempo; impedir que os profissionais de painéis excessivos atendam pessoas de outras equipes ou cubram ausências de outros profissionais; quando o painel é excessivo para um médico deve se introduzir outros profissionais não médicos na equipe; e redistribuir pessoas que estão adstritas a uma equipe com painel excessivo para outras equipes que estão com um painel baixo.

A definição de tamanho de painel é o número de pessoas pelas quais uma equipe é responsável.

O gerenciamento do tamanho do painel é fundamental para racionalizar a demanda. É uma medida de equidade na distribuição do trabalho. Um tamanho de painel apropriado é resultado de um sistema de acesso ótimo e, não, uma meta em si mesma. A meta deve ser uma boa gerência de painel que se expressa na equipe sendo responsável e cuidando de uma subpopulação a ela adstrita.

Na perspectiva da equipe ter um painel adequado e equânime assegura que ela será capaz de ofertar boa atenção, em tempo oportuno, a um número razoável de pessoas a ela vinculadas. Um painel balanceado garante que a demanda por serviços não deve superar a capacidade de oferta.

Um plano de gerenciamento do tamanho do painel pode ter alguns passos.

O primeiro passo é limitar o escopo da equipe e isto se expressa numa carteira de serviços que devem ser ofertados e em que quantidade. Os serviços a serem ofertados serão aqueles que efetivamente geram valor para as pessoas usuárias e suas famílias. Assim há que se fazer uma análise da carteira de serviços vigentes para verificar se alguns serviços podem ser descontinuados ou reduzidos. Aqui, também, há que se racionalizar os atendimentos de retorno que se farão, não segundo padrões culturais, mas por parâmetros assistenciais elaborados por estratos de riscos e definidos em diretrizes clínicas baseadas em evidências. Os retornos de pessoas com menores riscos serão mais frequentes que das pessoas com maiores riscos.

O segundo passo é garantir a continuidade do cuidado por meio do princípio da suficiência na construção das redes de atenção à saúde. Aqui a redução do retrabalho é muito importante. Uma boa distribuição das tarefas clínicas pela equipe garante a eliminação de retrabalho. Se a pessoa está agendada para atendimento por enfermeiro ela não deve ter outro atendimento com o médico, a menos que seja indicado pelo enfermeiro. Além disso, deve-se obedecer ao princípio da longitudinalidade que visa a prestar o atendimento pela equipe à qual as pessoas estão vinculadas. É importante que a equipe faça o atendimento dos eventos agudos porque se sabe que estes atendimentos feitos em unidade de pronto atendimento gera um atendimento subsequente pela equipe de atenção primária à saúde, o que significa retrabalho.

O terceiro passo é introduzir o autocuidado apoiado nas pessoas com condições crônicas não agudizadas. Há evidências de que o autocuidado apoiado diminui a demanda por consultas individuais face a face.

O quarto passo é estabelecer a equidade dos painéis. Se a equipe trabalha em tempo completo ela terá um painel maior que uma equipe que opera em tempo parcial. Quando duas equipes trabalham em tempo completo e apresentam diferentes painéis e, por consequência, diferentes níveis de demanda, esforços devem ser feitos para equilibrar estas diferenças. Isso pode ser feito redistribuindo pessoas de uma equipe para outra ou fechando temporariamente a vinculação de novas pessoas às equipes de maiores painéis.

5. O agendamento por blocos de hora

O acesso aos serviços de saúde é materializado por meio de sistemas de agendamento das pessoas usuárias aos serviços. Três métodos de agendamento são comumente utilizados: o agendamento em bloco, o agendamento em bloco modificado e o agendamento individual (NATIONAL ACADEMY OF ENGINEERING AND INSTITUTE OF MEDICINE, 2005).

No agendamento em bloco as pessoas usuárias são agendadas em turnos específicos durante o dia, tais como manhã e tarde, e atendidas, em geral, por ordem de chegada. No agendamento em bloco modificado um menor número de pessoas usuárias é agendado em períodos de tempo menores, como blocos de hora. No agendamento individual as pessoas usuárias são agendadas para um ponto específico de tempo determinado pela oferta dos profissionais, em geral, por atendimentos a cada 15 minutos (consultas médicas às 7,00 hs, às 7,15 horas, às 7,30 horas e assim sucessivamente).

No estudo de filas as variáveis se apresentam de modo aleatório. Para essas variáveis existem valores médios e uma distribuição de probabilidades. Isso implica que dizer que o ritmo médio de atendimento é de 4 pessoas a cada 60 minutos não significa afirmar que cada pessoa será atendida em 15 minutos, mas que o somatório dos 4 atendimentos é igual a 60 minutos.

Essa constatação teórica leva a que a forma mais virtuosa de agendamento na atenção primária à saúde é pelo agendamento em bloco modificado, ou seja, em blocos de hora. Nesse método, tendo como base um atendimento médio de 15 minutos, grupos de quatro pessoas são agendadas para cada hora do dia (por exemplo, as 7,00 horas deverão estar presentes quatro pessoas que serão ser atendidas entre 7,00 e 8,00 horas). A vantagem do bloco de horas é que introduz uma flexibilidade na atenção atendendo essas quatro pessoas com tempos diferenciados conforme as naturezas singulares de suas demandas (por exemplo, dois atendimentos de 10 minutos e dois atendimentos de 20 minutos).

6. A agenda aberta

A abertura da agenda é uma contribuição importante do enfoque do acesso avançado na atenção primária à saúde.

A organização racional do acesso à atenção primária à saúde pressupõe um sistema de agenda aberta todos os dias para o atendimento das pessoas que buscam as unidades para a atenção e que desejam ser atendidas naquele mesmo dia. Nos modelos clássicos de acesso avançado essa agenda aberta pode chegar à maioria dos atendimentos, especialmente nas consultas individuais. Contudo, quando se organiza o acesso, superando-se o mau *backlog*, mas valorizando o bom *backlog*, essa relação deve ser mais equilibrada para fugir do modelo da consultação.

A adoção da agenda aberta na atenção primária à saúde tem duas implicações que levam a mudanças em relação aos sistemas tradicionais de organização do acesso.

Em nenhuma circunstância deve-se utilizar a definição de uma equipe especialmente destinada, num turno ou num dia, para atender exclusivamente todas as demandas de eventos agudos. Entre outros problemas, isso lesa o atributo da longitudinalidade e realimenta a demanda.

Além disso, devem ser eliminados mecanismos tradicionalmente aplicados, frutos da permanência de elementos remanescentes dos programas verticais de saúde pública na atenção primária à saúde, como os agendamentos com horários reservados em dias prefixados como horário da gestante, horário para pessoas com hipertensão, horário para as crianças e outros.

Por outro lado, isso não significa reconhecer que há diferentes tipos de demandas que convocam diversos perfis de oferta que se constituem de composições tecnológicas distintas (a atenção aos eventos agudos; a atenção às condições crônicas não agudizadas, às enfermidades e às pessoas hiperutilizadoras; a atenção às demandas administrativas; a atenção preventiva; a atenção domiciliar; e o apoio ao autocuidado). Mas todos esses perfis de oferta serão oferecidos, na perspectiva da agenda aberta, sem que haja, também, fixação de horários especiais destinados a cada um deles.

7. A ordem de atendimento

O processo de fila é caracterizado por três elementos: o regime de chegada, o regime de serviço e a disciplina da fila.

O processo de chegada é quantificado através da taxa média de chegada ou pelo intervalo de tempo entre chegadas sucessivas que ocorrem de maneira aleatória. O processo de atendimento é quantificado através do ritmo médio de atendimento e do tempo ou duração

média do serviço. A disciplina da fila é a regra que define qual a próxima pessoa a ser atendida, é a ordem em que os usuários são selecionados da fila para o atendimento.

As regras de atendimento mais utilizadas são: FIFO (*first in-first out*), em que o atendimento ocorre por ordem de chegada (primeiro a chegar- primeiro a sair); LIFO (*last in- first out*), o último a chegar será o primeiro a ser atendido; PRI (*priority service*), em que o serviço é por ordem de prioridade definido pelos profissionais de saúde; SIRO (*service in random order*), em que o atendimento aos pacientes segue uma ordem aleatória (FOGLIATTI e MATTOS, 2007).

Na atenção primária à saúde, em geral, deve-se aplicar a regra da prioridade (PRI) porque a atenção à saúde deve priorizar as pessoas de maiores riscos, mesmo quando tenham chegado depois de pessoas de menores riscos. Isso é especialmente importante para a organização da atenção aos eventos agudos, o que implica a introdução de um sistema de classificação de riscos desses eventos para estabelecer a ordem de atendimento por uma equipe.

8. O tempo estendido nas unidades de atenção primária à saúde

O princípio da atenção centrada na pessoa, fundamental na organização do acesso à atenção primária à saúde, significa que os sistemas de atenção à saúde devem ser estruturados para responder às necessidades das pessoas e famílias que os utilizam. Infelizmente nem sempre é assim e, muitas vezes, ocorre que esses sistemas são desenhados para responder às necessidades dos prestadores de serviços de saúde.

Isso se expressa no tempo em que as unidades de atenção primária à saúde ficam abertas. É comum que estejam funcionando de 7 às 17 horas; muitas vezes, fecham no horário de almoço; e recentemente, em função da crise econômica dos municípios, uns poucos estão adotando uma prática completamente equivocada de funcionamento somente no horário da manhã, derivando as pessoas usuárias para as unidades de pronto atendimento.

Assim, na organização do acesso à atenção primária à saúde, é necessário que as unidades funcionem, principalmente, de 7 às 19 horas. Isso é necessário para dar maior flexibilidade de atendimento às pessoas usuárias, bem como para permitir que pessoas que trabalham em tempo integral tenham tempo de chegar às unidades no final do dia.

A lógica restritiva do funcionamento das unidades de atenção primária à saúde assenta-se na visão de senso comum de que os cuidados primários podem ser prestados, de forma

alternativa, pelas unidades de pronto atendimento. As unidades de pronto atendimento destinam-se somente à atenção aos eventos agudos - uma parte da demanda à atenção primária -, assim mesmo nos horários em que as unidades de atenção primária à saúde estão fechadas.

As evidências indicam que um atendimento numa unidade de pronto atendimento é feito exclusivamente para o problema da queixa principal e que cada atendimento nesta unidade de urgência gera outro atendimento na atenção primária à saúde.

9. As tecnologias de alisamento dos fluxos de atendimento

Melhorar o fluxo de trabalho e eliminar desperdícios aumenta a eficiência e a efetividade da clínica da atenção primária à saúde. Estima-se que mais de 40% do trabalho clínico é redundante ou apresenta desperdícios. Para contrapor-se a isso várias estratégias podem ser utilizadas como o mapa de fluxos, o ciclo de tempo do atendimento, o planejamento das consultas, a reunião da equipe, a co-locação, a padronização dos consultórios e dos equipamentos, a identificação e remoção de gargalos, a remoção de passos intermediários, o uso de processamento *just-in-time* e o processamento de tarefas em paralelo (INSTITUTE FOR HEALTHCARE IMPROVEMENT, 2016).

9.1 O mapa de fluxo

O mapa de fluxo e o ciclo de tempo do atendimento são ferramentas importantes para realizar o alisamento dos fluxos de atendimento e para aumentar a eficiência do atendimento (BACKER, 2002).

Mapa de fluxo significa acompanhar uma pessoa numa unidade e anotar detalhadamente o que ocorre neste circuito desde que a pessoa o inicia até que o termine, anotando todas as observações e impressões. O objetivo é identificar gargalos e compreender suas causas para superá-los. Os principais costumam gargalos costumam estar em uso excessivo de papéis, equipamentos inadequados, espaços mal dimensionados, duplicações de tarefas, retrabalhos, interrupções no trabalho do profissional (por exemplo, para atender chamadas telefônicas ou buscar um formulário) e atividades que não geram valor para as pessoas usuárias. É importante registrar as observações das pessoas usuárias que são acompanhadas em seus circuitos na unidade.

O mapa de fluxos permite uma compreensão dos processos reais, de seus problemas e de suas causas, permitindo imaginar como os processos podem ser redesenhados para garantir eficiência. O redesenho deve ser feito na perspectiva das pessoas usuárias e não dos profissionais de saúde. Também o conceito de valor é estabelecido em relação às pessoas usuárias.

O mapa de fluxos permite identificar processos que geram valor e outros processos que não geram valor para as pessoas e que devem ser eliminados. Tem se constatado que certos fluxos não agregam valor para as pessoas como certas pré-consultas e o autopreenchimento de históricos de saúde pelas pessoas usuárias (Backer, 2002).

O mapa de fluxos deve ser feito para perfis de oferta singulares: fluxos da atenção aos eventos agudos; fluxos da atenção às condições crônicas não agudizadas, às enfermidades e às pessoas hiperutilizadoras; fluxos de atenção preventiva; fluxos de demandas administrativas; fluxos de atenção domiciliar; e fluxos de apoio ao autocuidado.

9.2 O ciclo de tempo do atendimento

O ciclo de tempo do atendimento baseia-se no mapa de fluxos. Ele mede a quantidade em minutos que uma pessoa despende desde que chega à unidade até que a deixa. O ciclo do tempo não inclui, na atenção primária à saúde, o tempo gasto no laboratório ou em serviços de apoio diagnóstico.

Essa tecnologia busca distinguir entre o tempo que a pessoa gasta diretamente com os profissionais da equipe (tempo que gera valor) do tempo que gasta em espera de atendimento com esses profissionais (tempo que não gera valor). A meta não é necessariamente reduzir o ciclo de tempo total, mas maximizar o tempo que a pessoa gasta com o atendimento direto pelo profissional. Isso é importante porque se pode aumentar o acesso das pessoas aos atendimentos, mas com incremento dos tempos de espera e das filas no interior da unidade. A meta deve ser otimizar o tempo que gera valor, diminuindo o tempo que não gera valor.

A metodologia para aplicar essa tecnologia consiste em selecionar uma amostra de 15 pessoas por semana num determinado dia e numa determinada hora. Use o mesmo dia e a mesma hora de cada semana. Esse momento deve ser selecionado durante dia e hora de grande pressão de demanda na unidade, de modo a que se possa captar a verdadeira capacidade do sistema. Comece a registrar o tempo quando a pessoa inicia o atendimento

até quando o termina. Se a pessoa chegar muito cedo, considere somente o tempo em que o atendimento se inicia. A coleção de dados deve parar quando as 15 pessoas são avaliadas.

Um método de medir a quantidade de tempo para cada etapa do atendimento (tempo que gera valor e tempo que não gera valor) é usar o instrumento do ciclo da pessoa usuária.

O ciclo da pessoa usuária pode ser medido de três diferentes formas: a pessoa recebe o formulário e vai o preenchendo ao longo do atendimento; o preenchimento é feito pela equipe ao longo do atendimento; ou uma terceira pessoa acompanha o atendimento e faz as anotações.

Instrumento ciclo da pessoa usuária

HORÁRIO DO ATENDIMENTO:	HORA:
Instrução: Por favor preencha o formulário anotando cada item	
1. Hora de início do atendimento (p.ex, 13:53 hs)	
2. Hora que o staff administrativo o atendeu	
3. Hora que deixou o membro do staff administrativo	
4. Hora que o profissional de saúde começou o atendimento	
5. Hora que o profissional de saúde terminou o atendimento	
6. Hora em que chegou ao <i>check out</i>	
7. Hora em que deixou a clínica	
Comentários:	

Fonte Adaptado de Godfrey (2003).

O ciclo de tempo total do atendimento ideal é de 1,5 vezes o tempo do atendimento profissional. Assim se uma pessoa passa 20 minutos com um profissional o ciclo de tempo do atendimento ideal é de 30 minutos. Em geral, ciclo de tempo total do atendimento superior a 60 minutos indica a presença de problemas que devem ser sanados.

9.3 O planejamento das consultas

O planejamento das consultas, além de ser muito efetivo, é fácil de ser implementado. Esse planejamento envolve uma verificação anterior para verificar se toda a documentação necessária à consulta está disponível. Assim se uma consulta médica depende do resultado de um exame determinado e se ele não está disponível, faz-se um reagendamento para não se perder o tempo da consulta anteriormente agendada.

9.4 A co-locação

As pessoas de uma equipe, por trabalharem juntas, devem ser localizadas em espaços contíguos ou em uma situação de proximidade. É o que se denomina de co-locação que ajuda a reduzir os ruídos de comunicação e todos os seus efeitos perversos como erros, retrabalhos, atrasos e outras formas de desperdícios.

A distância física entre passos de um processo gera atrasos e a localização física pode afetar o tempo de processamento e causar problemas de comunicação. Assim, se a localização física de passos contíguos num sistema é movida de modo que estes passos estejam próximos, o trabalho pode fluir melhor de um passo a outro, reduzindo atrasos e diminuindo a necessidade de sistemas de comunicação adicionais.

Um bom exemplo de co-locação é manifesta-se na equipe de atenção primária à saúde. Uma comunicação eficiente, oportuna e frequente é essencial para o funcionamento dessa equipe e, por consequência, equipes que necessitam de se comunicarem regularmente devem estar fisicamente próximas.

9.5 A padronização dos consultórios e dos equipamentos e materiais

A padronização implica reduzir variações desnecessárias num processo.

Os consultórios devem ser padronizados em termos de materiais e equipamentos para gerar eficiência nos atendimentos. Na prática clínica reduzir a variação nos consultórios e nos equipamentos e materiais significa que os profissionais não perdem tempo procurando algum item importante antes ou durante o atendimento.

A padronização dos consultórios faz-se em consonância com o conceito de consultório aberto pelo qual um consultório pode ser utilizado por qualquer profissional em qualquer tempo em que esteja livre. Isso é diferente do trabalho clínico tradicional que aloca consultórios para uso exclusivo de determinados profissionais. A proposta de consultório aberto aumenta a flexibilidade para girar as pessoas usuárias por consultórios disponíveis melhorando o fluxo dessas pessoas e diminuindo as perdas de tempo.

Para implantar a proposta de consultório aberto os equipamentos e materiais utilizados devem ser padronizados. Além disso, é importante manter os consultórios com estoque de materiais a fim de evitar interrupções durante a interação clínica entre o profissional e as pessoas usuárias. Para tal, é necessário ter a lista padrão e disponibilizá-la no consultório e ter alguém com a responsabilidade de manter o consultório abastecido com os materiais padronizados. Da mesma forma, deve-se assegurar que cada consultório tenha o equipamento padronizado disponível (termômetro, esfigmomanômetro, balança, otoscópio, oftalmoscópio e outros) e em bom funcionamento para evitar interrupções no atendimento.

É preciso que haja consultórios suficientes para que todos os profissionais possam utilizá-los e que os fluxos das pessoas usuárias sejam otimizados. Em geral, os atendimentos se fazem com o profissional fixo num consultório e as pessoas girando. Mas vai se fixando a ideia de que os profissionais devem girar por vários consultórios, o que aumentaria a eficiência do trabalho clínico.

9.6 A identificação e a remoção de gargalos

O gargalo é alguma coisa que restringe o fluxo das pessoas usuárias dentro e entre os sistemas clínicos. O gargalo ocorre quando a demanda por um recurso particular (exemplo, um consultório, um profissional ou um exame) é maior que a oferta disponível. Para gerenciar o gargalo deve-se primeiro identificá-lo e, depois, remover o trabalho desnecessário incluído no gargalo.

Todo sistema tem um gargalo denominado de passo limitante da velocidade que determina a velocidade com a qual o trabalho transita pelo sistema. Esse gargalo é usualmente o mais valioso e mais escasso dos recursos. O foco deve estar em otimizar a oferta do gargalo limitante da velocidade e não em buscar a otimização de cada recurso do sistema como um todo.

É muito difícil identificar um gargalo avaliando a demanda e a oferta de cada recurso porque estes elementos podem ser mascarados por gargalos de outras partes do sistema. Para identificar o gargalo, observe onde o trabalho está se acumulando ou onde as filas estão se formando. Para identificar o gargalo é preciso observar certos sinais dentro do sistema tais como faltas de informação ou de materiais, ou onde as pessoas usuárias ou os profissionais estão esperando. Muitas vezes se pensa que o médico é o gargalo, mas sempre há outros fatores presentes.

Em geral, o trabalho da equipe clínica é o passo limitante da velocidade porque esses profissionais fazem coisas singulares que geram valor para as pessoas. Conseqüentemente qualquer trabalho que profissional clínico faz e que não está relacionado especificamente com suas habilidades e competências deve ser delegado a outro membro da equipe segundo o enfoque da otimização da força de trabalho organizada em equipes multiprofissionais interdisciplinares.

9.7 A remoção de passos intermediários

Cada membro da equipe de atenção primária à saúde que está envolvido no fluxo de pessoas ou de informações adiciona tempo ao processo e aumenta o risco de erros. Se cada renovação de receita requerer a assinatura do médico, deve-se eliminar o passo de ter um enfermeiro para revisar a solicitação ou deve-se desenvolver um sistema que não exija a assinatura do médico em cada receita.

As pessoas usuárias devem ser instruídas para, quando possível, se dirigirem diretamente ao profissional de sua equipe no *check-in* ao invés de dirigirem-se primeiramente a uma área de recepção central a fim de eliminar um passo extra no processo.

9.8 O uso de processamento just-in-time

O sistema *just-in-time* (no momento certo) foi desenvolvido originalmente para aumentar a eficiência da indústria automobilística. Com esse sistema os produtos e as matérias primas chegam ao local de utilização somente no momento exato em que for utilizado, eliminando a necessidade de formar grandes estoques.

Um comportamento típico nas organizações de saúde para processar múltiplas tarefas diárias é deixar algumas delas de lado para processá-las mais tarde em bloco. Uma das

consequências desse processamento em bloco é aumentar o número de vezes que as pessoas usuárias contactam a unidade para resolver seu problema ou imaginar que ele foi solucionado.

O tempo que se toma em renovar uma receita agora é o mesmo se isto não for feito agora porque serão x minutos agora ou num momento posterior. Desenvolver tarefas num processamento *just-in-time* requer ir de uma tarefa a outra rapidamente. Inicialmente o processamento *just-in-time* parece tomar mais tempo, mas, de fato, resulta em menos trabalho.

Alguns exemplos de processamento *just-in-time* podem ser mencionados.

Faça hoje o trabalho de hoje, se o trabalho for agendamento de atendimento, resposta a mensagens, renovar receitas ou preencher formulários. Um sistema que faz hoje o trabalho de hoje responde a cada demanda do dia em que foi gerada. Fazendo hoje o trabalho de hoje reduzem-se o absenteísmo, o retrabalho e a repetição de telefones e correio eletrônico.

Registre a consulta durante ou imediatamente após sua realização. Essa atitude previne o *backlog* de registro ao final do dia de trabalho e aumenta a eficiência porque o profissional se lembra mais dos detalhes neste momento que muitas horas depois. Esse sistema exige reservar alguns minutos adicionais ao tempo médio de consulta para que o registro seja feito durante ou no intervalo da próxima consulta. Alternativamente pode-se reservar um intervalo entre cada bloco de hora de 5 a 10 minutos para que o registro seja feito. Esse intervalo pode ser, também, utilizado para falar ao telefone ou responder correio eletrônico, falar com outro membro da equipe ou prescrever ou renovar receitas.

As frequentes interrupções no trabalho dos profissionais criam variações desnecessárias no fluxo de tarefas, perturba a coordenação do trabalho da equipe e contribui para o aumento dos tempos de espera das pessoas usuárias. Consequentemente, é necessário analisar as interrupções durante as consultas porque elas aumentam o tempo de atendimento e podem piorar a qualidade da relação entre o profissional e a pessoa usuária. Essa análise permite evidenciar as causas de interrupções, como a busca de um equipamento ou de um material não disponível no momento, o atendimento telefônico, a interrupção por outro membro da equipe e outras que devem ser minimizadas ou eliminadas. Um método simples de fazer essa análise é registrar em papel de toda a lista de interrupções com seu tempo em minutos no curso de uma manhã ou de uma tarde de um dia normal de trabalho. Um instrumento foi desenvolvido para isso e foi denominado de formulário de atividades não planejadas (BATALDEN, 2003).

Trata-se de um formulário simples que pode ser aplicado durante um período de atendimento como um turno diário ou um dia de trabalho por um profissional. O formulário tem o nome do profissional e a data de aplicação. O preenchimento é feito durante os atendimentos e deve registrar a razão pela qual a atenção à pessoa foi interrompida ou atrasada.

Instrumento de atividades não planejadas

RAZÃO	ANOTAÇÃO	TOTAL
Interrupção por telefone	////////	8
Interrupção por pessoal administrativo	///	3
Interrupção por pessoal clínico	/////	6
Falta de equipamento	////	5
Falta de material	////////	7
Falta de exame	////	4
Outra razão		

9.9 O processamento de tarefas em paralelo

Uma sequência rígida de passos num processo significa que um atraso em qualquer passo interrompe inteiramente o processo. O processamento em paralelo significa que o processo total pode continuar mesmo se uma parte dele sofreu atraso. Por exemplo, uma equipe que utiliza um protocolo para vacinação contra influenza pode aplicar a vacina antes da consulta se o profissional estiver atrasado.

Para organizar o processamento em paralelo se requer que os profissionais entendam perfeitamente os elementos do trabalho que fazem. Um *check list* serve a esse propósito.

10. O balanceamento da oferta e da demanda

O segredo para a melhoria do acesso à atenção primária à saúde está em alcançar o balanceamento da oferta e da demanda numa periodicidade diária, semanal e mensal.

As equipes de atenção primária à saúde têm mais controle sobre suas ofertas que sobre suas demandas. Isso é potencializado pelo modelo de gestão da oferta prevalente nas organizações de saúde que tende a privilegiar a solução para o desequilíbrio entre oferta e demanda pelo incremento da oferta.

Este processo de balanceamento requer uma excelente compreensão da oferta e da demanda nos cuidados primários. E convoca, também, um processo de comunicação eficaz em todos os níveis, em especial utilizando reuniões com equipes clínicas e administrativas.

É importante usar meios de comunicação para gerenciar diária e semanalmente as variações de oferta e demanda, bem como antecipar e planejar eventos sazonais.

O objetivo de fazer hoje o trabalho de hoje consiste essencialmente em balancear a demanda de hoje com a oferta de hoje.

10.1 A medição e a compreensão da oferta e da demanda

A compreensão dos padrões de oferta e demanda semanais, mensais e sazonais permite focar os esforços em racionalizar a demanda para compatibilizá-la com a oferta e/ou aumentar ou diminuir a oferta em períodos de alta ou baixa demanda.

A melhoria do acesso à atenção primária à saúde consiste em equilibrar a oferta e a demanda, o que significa que não há maus *backlogs*, nem atrasos, quando a demanda se faz e quando os serviços são prestados.

O descompasso entre oferta e demanda não somente contribui para um atraso em atender às necessidades das pessoas usuárias, mas, também, costuma ser cara e gerar desperdícios no sistema.

A experiência de várias organizações de saúde demonstra que a demanda não é imprevisível, ao contrário, é inteiramente previsível. De fato, a demanda por qualquer tipo de serviço pode ser acuradamente prevista com base na população, no escopo da organização prestadora e em seu particular estilo clínico.

Períodos de altas e baixas demandas podem ser antecipados, com base numa análise de dados de demanda. Um sistema de acesso eficaz usa essas previsões da demanda para organizar sua oferta segundo as necessidades das pessoas usuárias em relação aos serviços demandados.

O balanceamento da oferta e da demanda envolve medir a oferta, medir a demanda e comparar as duas.

A medida da oferta deve ser feita para todos os recursos clínicos. É importante verificar a capacidade de oferta em relação aos diferentes perfis de oferta: atenção aos eventos agudos; atenção às condições crônicas não agudizadas, às enfermidades e às pessoas hiperutilizadoras; atenção preventiva; atenção às demandas administrativas; atenção domiciliar; e apoio ao autocuidado. Da mesma forma, é preciso medir a demanda por cada tipo de perfil de oferta.

Também importante é medir a demanda externa quando é feita pelas pessoas usuárias e a demanda interna que é gerada pela equipe de saúde, especialmente sob a forma de atendimentos programados de retorno (o bom *backlog*).

Em situações de desequilíbrio mostradas na comparação entre oferta e demanda é preciso tomar medidas para racionalizar a demanda e depois racionalizar a oferta, buscando balancear estes dois elementos centrais do acesso aos serviços de saúde. Essa é uma dica fundamental: sempre começar o trabalho de balanceamento pelo lado da demanda e, não como é comumente feito, pelo lado da oferta.

10.2 A medição da demanda diária

A demanda diária corresponde ao número de pessoas que solicitam atendimento num dia, sem considerar quando esta demanda foi efetivamente transformada num atendimento. A demanda não pode ser confundida com o número de atendimentos realizados. Ela deve ser medida por cada profissional clínico e agregada pela equipe.

A demanda por agendamentos vem de duas fontes: interna e externa. A demanda interna tem como fonte a própria equipe que programa os retornos das pessoas usuárias; a demanda externa é feita pelas próprias pessoas, à distância, por correio eletrônico ou telefone, ou presencialmente na unidade de saúde. Assim: demanda = demanda interna + demanda externa.

A meta é chegar a uma demanda diária que seja igual ou menor que a oferta diária.

O plano de coleta de dados envolve a coleta diária, na mesma semana de cada mês, para cada profissional, com referência a: agendamentos diários por telefone independentemente do dia para o qual o agendamento foi realmente realizado; pessoas que chegaram presencialmente à unidade, agendadas para o mesmo dia, agendadas para dias futuros ou referidas a outras unidades de saúde; agendamentos solicitados por correio eletrônico, por telefone ou por mecanismos centralizados de regulação; pessoas que foram referidas por outras unidades como unidades de urgência, unidades de atenção ambulatorial especializada e hospitais; e pessoas que foram agendadas no dia para retorno programado.

10.3 A medição da oferta diária

A oferta diária é o número de agendamentos clínicos realizados num dia.

O tempo de espera para um agendamento é a diferença entre o momento em que o agendamento foi solicitado e o momento que a oferta se concretizou. É importante compreender a demanda e a oferta por agendamento quando se quer diminuir os tempos de espera.

A oferta refere-se ao total de recursos (pessoas, equipamentos, salas de atendimentos e consultórios) disponíveis numa unidade de saúde para a equipe. Quando os recursos totais são bem gerenciados uma unidade cria oportunidades de agendamentos e/ou espaços para atendimentos das pessoas usuárias. Essas pessoas experienciam a abertura da unidade para os atendimentos na forma da disponibilidade de agendamentos. Como em muitos sistemas o tempo clínico constitui o gargalo, a capacidade de oferta é o tempo total de cada profissional clínico destinado ao agendamento.

A meta é chegar a uma oferta igual ou maior que demanda diária.

O plano de coleta de dados para a oferta diária inicia-se pelo número de profissionais clínicos disponíveis em um dado período. Então meça o número de atendimentos que cada profissional é capaz de realizar na forma como o sistema está operando. Faça isso multiplicando os agendamentos realizados por dia pelo número de dias por semana (excluindo o tempo de trabalho não clínico) para obter os agendamentos por semana. Esses cálculos mostram aproximadamente quantos agendamentos e quantos blocos de horas estão disponíveis para cada profissional.

10.4 A medição de atendimentos a novas pessoas usuárias

Esse número expressa a quantidade total de novas pessoas usuárias atendidas pela equipe clínica no período de um mês.

A meta é que esse dado seja determinado rotineiramente em cada equipe de atenção primária à saúde.

O plano de coleta de dados é medir o número de novas pessoas usuárias utilizando-se o prontuário clínico.

10.5 A estratificação de risco das condições crônicas não agudizadas

O processo de estratificação de risco da população permite identificar pessoas e grupos com necessidades de saúde semelhantes que devem ser atendidas por tecnologias e recursos específicos, segundo um sistema de estratificação.

Quando uma população não é estratificada por riscos pode-se subofertar cuidados necessários a pessoas de maiores riscos e/ou sobreofertar cuidados desnecessários a pessoas de condições de menores riscos produzindo, por consequência, uma atenção inefetiva e ineficiente. Esse problema explica, em grande parte, as dificuldades de organizar a agenda de atendimentos na atenção primária à saúde, o que repercute negativamente no acesso aos serviços.

A repercussão no acesso à atenção primária à saúde dá-se pela construção da previsibilidade e racionalização da demanda, o que é feito por meio da programação de atividades clínicas por pessoas estratificadas por riscos. Assim, a planilha de programação estabelece quantas atividades clínicas, por tipo e por ano por estrato de risco de uma condição crônica não agudizada. Isso tem uma repercussão muito positiva na organização do bom *backlog* de atividades clínicas de retorno.

De um lado porque permite distribuir por riscos os atendimentos, segundo os parâmetros assistenciais estabelecidos nas diretrizes clínicas. Pessoas de baixos riscos terão uma concentração de atendimentos menor do que pessoas com maiores riscos. Isso não ocorre quando não se estratificam os riscos, o que leva a exageros de ofertas nas pessoas de baixos riscos. É comum, na prática da atenção primária à saúde, hipertensos de baixos riscos com 6 consultas médicas e de enfermagem por ano ou, até mesmo, com uma rotina de consultas mensais, quando seriam necessárias apenas 1 consulta com o médico e 1

outra com o enfermeiro. De outro lado, a previsibilidade permite que os atendimentos de retorno sejam distribuídos por um período anual, marcando-os para períodos mensais (fora das variações sazonais atípicas) e para variações diárias de menores pressões de demanda.

10.6 A identificação e a atenção adequada das pessoas hiperutilizadoras

As demandas de pessoas hiperutilizadoras afetam significativamente a racionalização do acesso aos cuidados primários. A utilização dos serviços de atenção primária à saúde pode ser medida pelo número de vezes que uma pessoa utiliza, num ano, os cuidados primários; se esse número se distancia, para mais, da média de utilização se poderia estar diante de um hiperutilizador.

O entendimento do fenômeno das pessoas hiperutilizadoras é importante na atenção primária à saúde, seja na perspectiva de sua economicidade, seja na perspectiva de sua qualidade. Em termos da economicidade, a hiperutilização dos serviços gera uso inadequado e ineficiência dos recursos, seja em tempo da equipe, seja em medicamentos e exames complementares, seja nos custos de oportunidade para as pessoas usuárias.

Têm sido propostos critérios para identificar essas pessoas. Para uns, pessoas hiperutilizadoras são aquelas cujo número de consultas excede um valor médio ou de referência para o respectivo grupo com o mesmo sexo e faixa de idade, num dado período, numa unidade; para outros, elas seriam definidas como as 10% mais frequentes em cada equipe de atenção primária à saúde. Uma forma alternativa de definir as pessoas hiperutilizadoras é pela percepção subjetiva dos profissionais de saúde. Quando um profissional percebe que uma determinada pessoa consulta com uma frequência superior à esperada, tendo em vista a situação clínica e de saúde na perspectiva do profissional, essa pessoa passa ser hiperutilizadora, independentemente de um critério numérico (VESTED e CHRISTENSEN, 2005; RAMOS e CARRAPIÇO, 2012).

As pessoas hiperutilizadoras, comumente, apresentam problemas físicos, mentais e sociais. Essas pessoas têm maiores probabilidades de apresentar enfermidades, problemas gerais e inespecíficos e quadros clínicos pouco organizados e indiferenciados.

As pessoas hiperutilizadoras beneficiam-se do trabalho multiprofissional interdisciplinar, das novas formas de encontro clínico como o atendimento compartilhado a grupo e o grupo

operativo. Portanto, não devem ser atendidas somente por consultas presenciais face a face.

10.7 O uso regular da prevenção quaternária

As equipes de atenção primária à saúde devem estar capacitadas a introduzir na rotina de seus trabalhos a prevenção quaternária.

Em função dos excessos cometidos no campo da medicina, foi proposto o conceito de prevenção quaternária, adotado pela Organização Mundial de Médicos de Família (WONCA), que significa as ações tomadas para identificar pessoas e populações em risco de medicalização excessiva, para livrá-las de intervenções médicas invasivas e para prover serviços que sejam cientificamente aceitáveis e seguros.

A prevenção quaternária destina-se a proteger as pessoas usuárias em relação ao excesso de intervenções de rastreamento, à medicalização dos fatores de risco, à solicitação de exames complementares em demasia, ao excesso de diagnósticos, às medicalizações desnecessárias de eventos vitais e de adoecimentos autolimitados, às solicitações de exames e tratamentos pedidos pelas pessoas usuárias e à medicina defensiva (JAMOULLE et al, 2002).

A prevenção quaternária, além de melhorar a qualidade da atenção à saúde, é um elemento importante na racionalização da demanda aos cuidados primários, com repercussões muito positivas, também, na demanda por cuidados secundários.

10.8 A redução do absenteísmo

A racionalização da demanda beneficia-se muito de uma política de redução do absenteísmo. Lembretes realizados pelos agentes comunitários de saúde ou por internet ou telefone devem ser feitos imediatamente antes do atendimento para diminuir as ausências aos atendimentos agendados. Ademais, pode-se aumentar o número de agendamentos em situações em que o excesso de pessoas é menor que 10% e permitir excesso de agendamentos se a equipe de saúde concorda em trabalhar até que todas as pessoas usuárias sejam atendidas.

O absenteísmo está correlacionado com demoras no acesso e com a fragilidade do atributo da longitudinalidade do cuidado que dificulta as pessoas serem atendidas pelos profissionais da equipe a que estão vinculadas. Assim, quanto menor o tempo que decorre entre a solicitação do agendamento e a data deste agendamento menor será o absenteísmo. Da mesma forma, se as pessoas têm a possibilidade de serem atendidas pelos mesmos profissionais ao longo do tempo, o absenteísmo será reduzido.

A redução do absenteísmo pode ser medida pelo percentual para os quais as pessoas usuárias não compareceram.

A fórmula para essa medição é o número total de agendamentos com falta das pessoas agendadas dividido pelo número total de agendamentos disponíveis, multiplicado por 100.

A meta a ser alcançada é de 50% de redução em todas as atividades clínicas.

O plano de coleta de dados é contar, no último dia útil de cada mês, para cada profissional de saúde ou atividade clínica de grupo, o número de agendamentos totais disponíveis no mês (denominador) e o número total de pessoas que faltaram à atividade agendada (numerador).

A redução do absenteísmo deve ser uma política na atenção primária à saúde. Para implantá-la conseqüentemente algumas atitudes são importantes.

Conhecer as pessoas usuárias. Em unidades em que a percepção é de que o número de ausências às atividades é relativamente pequeno em relação à população total é necessário examinar mais detidamente as causas destas ausências que podem ser por falta de transporte, lapsos de memória ou outras causas.

Muitas vezes o absenteísmo está correlacionado com o número de agendamentos que são cancelados pela equipe de atenção primária à saúde. Dessa forma, devem ser evitados esses cancelamentos.

O absenteísmo tem a ver com os atendimentos de retorno. Isso não deve ser considerado para eliminar ou diminuir os retornos que constituem o bom *backlog* e que são essenciais no manejo adequado das condições crônicas. Mas o retorno tem de ser discutido com as pessoas usuárias e deve fazer sentido para elas. Se as pessoas não veem valor no retorno elas tendem a ausentar-se deste agendamento. Além disso, os retornos são facilitados pela qualidade do vínculo que existe entre a equipe e as pessoas a ela adstrita. A diminuição do absenteísmo nas atividades clínicas de retorno exige um sistema de lembrete, com 24 horas de antecedência, sobre o agendamento realizado. As pessoas que se ausentam mais frequentemente devem ser examinadas para se conhecer a causa

destas ausências. Alguns atendimentos de retorno, se desnecessários, devem ser eliminados e, outros, se necessários, em certas circunstâncias, podem ser substituídos por atividades não presenciais por telefone, por correio eletrônico ou pela presença no domicílio do agente comunitário de saúde.

Os atendimentos de retorno devem deixar de serem feitos por razões culturais ou porque os retornos são feitos mensalmente porque sempre foi assim nesta unidade. Por exemplo, existe em certas equipes uma cultura de retornos mensais, independentemente dos riscos das pessoas usuárias.

Os retornos para monitoramento de pessoas com condições crônicas não agudizadas devem ser realizados conforme uma programação realizada anualmente, a partir de diretrizes clínicas baseadas em evidência, e por estratos de risco.

Esses atendimentos de retorno, especialmente para pessoas com condições crônicas menos complexas, devem ser feitos com um horizonte de tempo maior (seis meses ou um ano). Isso permitirá distribuir melhor a carga de trabalho e estabelecer planos de contingência para variações sazonais de determinadas condições de saúde. Além disso, os atendimentos de retorno devem ser feitos para dias e horários de menor afluxo de pessoas às unidades, detectados no diagnóstico da demanda efetiva por cuidados primários.

10.9 O reforço de uma política de atender mais de um problema durante um atendimento

A clínica da atenção primária tem características próprias. Essa é uma clínica de problemas e não uma clínica de diagnósticos. Isso significa que na atenção primária à saúde não se trabalha somente com a queixa principal devendo-se, num mesmo momento clínico, explorar outros problemas. Esse é o conceito ligado à oportunidade do atendimento clínico.

A ampliação dos problemas a serem abordados num encontro clínico traz vantagens para as pessoas usuárias reduzindo o custo de oportunidade de se apresentarem muitas vezes a um atendimento e contribui significativamente para a redução da demanda.

Essa política de aproveitar o tempo clínico para abordar mais de um problema deve ser estimulada na equipe. Cada pessoa com evento agudo que deixa de ser vista por uma equipe de atenção primária à saúde para ser atendida numa unidade de pronto atendimento significa um atendimento de um só problema que, ainda, costuma gerar outro

atendimento na unidade de cuidados primários. O que além de empobrecer o cuidado clínico, representa uma sobrecarga de demanda no sistema de atenção à saúde.

Uma boa clínica da atenção primária à saúde sempre tem a pergunta: há algo mais que gostaria de me dizer ou de discutir comigo?

Da mesma forma, deve-se sempre buscar saber se uma pessoa a ser atendida já tem um agendamento para outra data. Nesse caso, há que se verificar a necessidade deste novo agendamento ou se essa necessidade pode ser coberta no primeiro agendamento.

O uso de um *check list* de cuidados preventivos pode antecipar necessidades futuras de pessoas já agendadas.

10.10 A construção de alternativas aos atendimentos presenciais face a face

Neste campo se estarão desenvolvendo medidas sugeridas pelo enfoque de alternativas tecnológicas ao atendimento presencial

A maioria dos atendimentos na atenção primária à saúde é realizada na forma convencional do encontro presencial face a face entre o profissional e a pessoa usuária como consultas médicas e de enfermagem.

Algumas alternativas têm sido usadas como: a utilização de consultas médicas ou de enfermagem por telefone ou por correio eletrônico, sendo que o tempo para isso deve ser reservado para o final de cada dia; uso de telefone para o monitoramento dos planos de cuidado ao invés de exigir a presença física da pessoa na unidade; uso de alternativas de gerenciamento de demandas administrativas, como a renovação de receitas, que minimizem o atendimento presencial face a face; e uso de instrumentos de teleassistência que permitam gerenciar as condições crônicas não agudizadas. Para que isso possa ocorrer é preciso haver disponibilização de correio eletrônico institucional, disponibilização de correios eletrônicos dos profissionais da equipe e disponibilização de chat para diálogo entre a equipe e as pessoas usuárias.

Numa visão prospectiva essas alternativas devem compor uma política de cuidado eletrônico (e cuidado) na atenção primária à saúde que pode envolver outras tecnologias como acesso *online* aos prontuários eletrônicos, acesso *online* aos resultados de exames, prescrições *online* de medicamentos de uso contínuo e acesso *online* a materiais educativos.

Além disso, devem ser introduzidas na rotina da clínica da atenção primária novas formas de atendimento como a atenção contínua, a atenção compartilhada a grupo, os grupos operativos e outras (MENDES, 2012). A oferta desses atendimentos amplia capacidade de oferta da atenção primária à saúde. Essas atividades devem ser agendadas para aqueles horários em que há menor pressão de demanda para a equipe.

A atenção contínua foi desenvolvida para o manejo de pessoas portadoras de condições crônicas. Ela foi proposta para atender pessoas de forma sequenciada, passando por diferentes profissionais de saúde, num mesmo turno de trabalho. O produto da atenção contínua é um plano de cuidado multiprofissional interdisciplinar.

A atenção contínua é indicada para portadores de condições crônicas nas seguintes situações: prevenir ou minimizar complicações para pessoas com incapacidades importantes; gerenciar proativamente a atenção em subpopulações específicas; coordenar os serviços preventivos; monitorar as intervenções que têm sido feitas com base em evidências de sua efetividade ou de padrões ótimos de atenção; elaborar ou revisar os planos de cuidados realizados conjuntamente pela equipe de saúde e pela pessoa usuária; e avaliar os processos e resultados da atenção. Essa tecnologia aplica-se, com especial interesse, em avaliações iniciais de pessoas portadoras de condições crônicas que se beneficiam de diferentes abordagens profissionais.

A programação da atenção contínua é feita para uma coorte de oito pessoas portadoras de condições crônicas, por um turno de trabalho de quatro horas e agendada previamente com os profissionais (GROUP HEALTH COOPERATIVE, 2001). No caso, trata-se de uma atividade clínica feita com o médico, o enfermeiro, o farmacêutico e um coordenador de grupo.

PESSOA	RECEPÇÃO	MÉDICO	ENFERMEIRO	FARMACÊUTICO	GRUPO
Pessoa 1	8,40	8,45-9,00	9,00-9,15	9,15-9,30	10,30-11,15
Pessoa 2	8,55	9,00-9,15	9,15-9,30	9,30-9,45	10,30-11,15
Pessoa 3	9,10	9,15-9,30	9,30-9,45	9,45-10,00	10,30-11,15
Pessoa 4	9,25	9,30-9,45	9,45-10,00	10,00-10,15	10,30-11,15
Pessoa 5	9,40	9,45-10,00	10,00-10,15	11,30-11,45	10,30-11,15
Pessoa 6	9,55	10,00-10,15	11,30-11,45	11,45-12,00	10,30-11,15
Pessoa 7	10,00	10,15-10,30	11,45-12,00	12,00-12,15	10,30-11,15
Pessoa 8	11,30	11,45-12,00	12,00-12,15	12,15-12,30	10,30-11,15

A carta convite padrão escrita às pessoas pode ser a seguinte:

Prezado (a) Sr (a) _____

A fim de melhorar a atenção à sua saúde nós estamos iniciando, em nossa unidade, uma atividade denominada de atenção contínua.

É uma atividade clínica que você poderá, para seu conforto, vir à nossa unidade num turno de quatro horas e receber atenção de vários profissionais que estão envolvidos no seu tratamento. Você terá atendimento pelo médico, pelo enfermeiro e por outros profissionais que são importantes para sua saúde. Haverá, também, uma atividade de grupo para que você possa trocar ideias com outras pessoas com a mesma condição de saúde e informar-se melhor a respeito de como controlá-la.

Você está convidado (a) a participar da primeira atividade de atenção contínua que será realizada em nossa unidade no dia _____. Um profissional de nossa equipe entrará em contacto com você para explicar com mais detalhes o que é a atenção contínua e para agendar sua participação. Se preferir pode nos contatar pelo telefone _____ ou pelo correio eletrônico_____.

Atenciosamente,

(profissional da unidade)

A atenção compartilhada a grupo não se destina a uma pessoa individualmente, nem é provida por um único profissional, mas se foca em grupos de pessoas com condições de saúde semelhantes (atenção a grupo), contando com a participação de vários profissionais da equipe de saúde ao mesmo tempo (atenção compartilhada).

A atenção compartilhada a grupo tem sido indicada para pessoas que necessitam de monitoramento contínuo, pessoas portadoras de condições crônicas estáveis, pessoas que precisam de maior tempo de atendimento e pessoas com necessidades intensas de apoio

emocional ou psicossocial. Essa tecnologia é particularmente útil para as pessoas hiperutilizadoras dos serviços de atenção primária à saúde e para pessoas com enfermidades (*illnesses*). É também um espaço clínico relevante para elaboração e monitoramento de planos de autocuidado.

A atenção compartilhada a grupo tem sido indicada por vários motivos: mostra à pessoa que ela não é a única com aquela condição; gera esperança ao ver casos de sucesso em outras pessoas; compartilha informações; diminui ansiedade; permite o desempenho de papéis de modelagem entre os pares; oferece aprendizagem interpessoal e cognitiva; e aumenta a coesão do grupo (AMERICAN ACADEMY OF FAMILY PHYSICIANS, 2005).

A implantação da atenção compartilhada a grupo supõe um planejamento bem feito pela equipe multiprofissional e um espaço físico adequado para sua consecução, com os meios necessários ao desenvolvimento de trabalhos de grupo.

Os membros da equipe que irão participar dessa atividade clínica são definidos anteriormente e planejam todo o trabalho (MENDES, 2012).

Um membro da equipe deve estar presente na sala antes que as pessoas cheguem para acolhê-las e ajudá-las a se identificarem, nas tarjetas, pelo nome que gostariam de ser chamadas. No horário acordado a equipe multiprofissional se dirige à sala. Um membro da equipe assume as funções de facilitador-líder, algumas vezes o médico, e abre a sessão. Esse facilitador-líder deve ter domínio de dinâmicas de grupo de forma a poder lidar bem com diferentes tipos de comportamentos em grupo. Ele dá as boas-vindas e conduz a apresentação breve de todos os membros da equipe de saúde, bem como dos participantes (15 minutos). Em seguida, fala sobre os objetivos da atividade e abre espaço para questões dos participantes como: Que querem fazer? Que esperam desse trabalho? Que gostariam de discutir sobre sua condição de saúde? Depois fala sobre as normas da atenção compartilhada a grupo e indaga sobre as expectativas, inclusive de confidencialidade.

Depois, abre-se um tempo para interações e questões dos participantes. Essas questões são distribuídas pelo facilitador-líder aos demais membros da equipe para serem respondidas. Isso mostra para os participantes que a equipe trabalha conjuntamente. Os membros da equipe de saúde devem resistir à tentação de dar aulas, de dizer aos participantes o que eles acham que deveriam fazer em relação aos seus sintomas e aos seus tratamentos; essa atitude prescritiva por parte dos profissionais de saúde mina as possibilidades de sucesso da atividade. O papel de cada membro da equipe de saúde é somente facilitar a interação do grupo e criar um ambiente de confiança em que os

participantes possam colocar abertamente suas expectativas, seus medos, suas angústias, suas esperanças e suas possibilidades de controlarem suas condições de saúde. Os profissionais de saúde não devem responder às questões, mas redirecioná-las ao grupo com questões do tipo: Alguém aqui experimentou esse problema? O que funcionou para vocês? Isso aumenta a confiança dos participantes e sua habilidade de resolver, por si mesmos, seus problemas de saúde (30 minutos).

Em seguida, faz-se um intervalo, mas durante esse tempo (15 minutos), o médico conversa individualmente com os participantes e pode refazer a receita de medicamento de uso continuado e o enfermeiro toma os sinais vitais, cada qual começando de um lado. Em caso de presença de outros profissionais eles devem ter um momento com os participantes do grupo; por exemplo, o farmacêutico pode verificar a adesão aos medicamentos e o nutricionista fazer uma avaliação nutricional. Os membros da equipe devem perguntar aos participantes se desejam consultas individuais após a atividade. Caso positivo, devem agendá-las.

Há um intervalo em que deve ser oferecido um café ou compartilhados alimentos trazidos pelos participantes.

Na volta, retoma-se a sessão, para um momento de questões abertas sobre a saúde dos participantes ou sobre a dinâmica do grupo, a serem feitas e consideradas por todos os membros da equipe de saúde. Esse momento pode ser estimulado pelo facilitador-líder, a partir de um fato relevante como, por exemplo, um novo tratamento que apareceu na televisão naqueles dias (15 minutos).

Após esse período de interações, inicia-se uma discussão sobre o que se pretende discutir no próximo atendimento, geralmente, um mês após esta primeira sessão. Há uma tendência das equipes de saúde de definirem tecnicamente essa agenda. Mas sabe-se que deixar emergir dos participantes essa agenda de discussão contribui para fortalecer os vínculos entre os participantes e para que os grupos desenvolvam um sentimento de autoconfiança.

Por fim, fecha-se a sessão, com um posicionamento claro e forte por parte do facilitador-líder, com um agradecimento e com a marcação do local e da data do próximo encontro (15 minutos).

Seguem-se os atendimentos individuais que foram agendados (30 minutos) e termina-se com um tempo livre a ser utilizado pela equipe de saúde segundo a evolução da sessão (30 minutos). Assim, uma sessão de atenção compartilhada a grupo típica tem uma duração média de 2 horas e 30 minutos.

10.11 O gerenciamento das variações da demanda e da oferta: os planos de contingência

A demanda à atenção primária à saúde apresenta variações sazonais. No Brasil, ainda que com diferenças regionais, há variações de demanda por problemas das vias aéreas superiores no inverno e por arboviroses no verão (dengue, zika e febre chikungunya). A demanda por essas doenças é significativamente maior em determinados períodos de tempo, o que pode constranger o tempo dedicado ao atendimento de outras demandas, gerando uma acumulação de mau *backlog* com todas as suas más consequências. Como esses períodos estão bem estabelecidos pode-se criar previsibilidade para essa demanda aumentada, sem que esse aumento leve a uma situação caótica na unidade de atenção primária à saúde.

Esse gerenciamento da demanda sazonal tem a ver com uma programação inteligente da demanda direta da equipe por atendimentos de retorno. Como os retornos devem ser programados de acordo com os parâmetros assistenciais estabelecidos nas diretrizes clínicas e, muitos deles, são de prazos mais longos, nesses períodos de aumento sazonal da demanda devem ser diminuídos, ao máximo possível, esses agendamentos feitos diretamente pela equipe da atenção primária à saúde.

É importante gerenciar a variação na oferta que pode variar tanto quanto ou até mais que a demanda. A variação na oferta pode resultar de doenças inesperadas de profissionais, por ausências ao trabalho por outros motivos que não doenças, por férias, por greves, por obras de reformas na unidade e por outras intercorrências.

Essas variações na demanda e na oferta que ocorrem como parte do dia a dia da atenção primária à saúde criam problemas que devem ser respondidos por planos de contingência (INSTITUTE OF HEALTHCARE IMPROVEMENT, 2016).

Mesmo que haja um equilíbrio entre demanda e oferta, haverá momentos de aumento inesperado da demanda, sejam esperados como nas variações sazonais, sejam inesperados como um aumento da procura por serviços num determinado dia. Para melhor gerenciar as variações de oferta e demanda numa base diária há que se desenvolverem planos de contingência para diversas situações.

Certo número de situações ocorre de forma relativamente infrequente, mas não completamente inesperada, como uma mãe que tinha um filho com consulta agendada e que traz outro filho que também está necessitando de atendimento naquele momento. Assim, para evitar distúrbios no fluxo normal da unidade é preciso desenvolver planos de

contingência que estabeleçam protocolos padronizados a serem seguidos nessas situações, bem como definam claramente as responsabilidades de cada membro da equipe nesse processo (como proceder em casos de ausência inesperada de um profissional, em caso de férias, em casos de ausência para uma reunião de trabalho fora da unidade e outras situações).

Costuma acontecer uma ausência súbita de um profissional da equipe e pode ser de curto prazo (ausência por doença de um dia) ou de longo prazo. Por exemplo, numa ausência inesperada de um dia o plano de contingência pode prever uma reunião rápida da equipe para identificar quem vai cobrir o colega ausente para aquelas pessoas que não aceitam ter seu atendimento adiado para outro dia.

A substituição dos profissionais em férias deve ser prevista em plano de contingência.

Uma das ocorrências comuns numa unidade de atenção primária à saúde é a pessoa que chega atrasada ao atendimento. Deve haver um protocolo claro de como portar-se nessas situações.

Planos de contingência devem ser feitos para variações sazonais. É comum que nos momentos de epidemia de dengue a demanda por cuidados em relação a esta doença aumente fortemente, causando uma desorganização do trabalho na unidade. Da mesma forma o inverno traz um aumento das doenças das vias aéreas superiores. Planos de contingência devem existir para enfrentar essas situações previsíveis. Por exemplo, numa epidemia de dengue pode-se organizar, numa unidade de atenção primária à saúde, um fluxo rápido (*fast track*), com uma equipe especial, para atender pessoas com suspeita da doença. Esse fluxo especial deve estar prescrito num plano de contingência expresso num protocolo de atendimento. Com isso, outros profissionais poderiam continuar um atendimento conforme agendamentos previamente realizados.

10.12 O gerenciamento do tempo dos profissionais

Uma das medidas efetivas para a racionalização da oferta é o gerenciamento do tempo contratado com os profissionais.

É comum que profissionais contratados para oito horas de trabalhos diários não cumpram efetivamente esse tempo. Às vezes, em instituições mais voltadas para o atendimento das demandas dos profissionais do que da população, as 8 horas de trabalho são

transformadas em 32 consultas por dia que terminadas levam ao fim do trabalho diário sem o cumprimento total da carga horária contratada.

Para que o acesso à atenção primária à saúde seja organizado é fundamental que os gestores exijam dos profissionais o cumprimento do tempo contratado. Certamente é preferível que toda a equipe trabalhe com uma carga de trabalho de 8 horas diárias. Isso é fundamental para que o sistema de agendamento por blocos de hora seja efetivado.

A prática tem demonstrado que a adoção de ponto eletrônico, utilizado de forma adequada e transparente, ajuda o gerenciamento do tempo dos profissionais.

Além de garantir a presença integral no horário contratado o gerenciamento do tempo dos profissionais implica garantir que um percentual máximo do tempo contratado seja utilizado em atividades clínicas que gerem valor para as pessoas usuárias. Como exemplo, o tempo do médico deve ser maximizado em encontros clínicos, tentando diminuir, ao máximo, outras atividades, especialmente de tipo administrativo.

Certamente que se deve planejar um tempo para educação permanente em horário protegido e para reuniões de equipe.

10.13 O aumento da oferta por novos investimentos

Após a realização de todas as racionalizações do lado da demanda e da oferta, em certas circunstâncias, ainda poderá permanecer um desequilíbrio dentre oferta e demanda.

Nesses casos, justifica-se elaborar um projeto de investimento para incremento da capacidade de oferta de serviços, seja na contratação de novos profissionais, seja na ampliação das instalações físicas ou na aquisição de novos equipamentos.

11. A utilização de equipe multiprofissional interdisciplinar

Neste campo se estará considerando a introdução, na equipe de atenção primária à saúde, de novos profissionais de nível superior, conforme as diretrizes do enfoque da otimização da força de trabalho organizada em equipes multiprofissionais interdisciplinares.

Novos profissionais de nível superior como assistente social, farmacêutico clínico, fisioterapeuta, nutricionista, psicólogo e profissional de educação física devem ser

incorporados como membros efetivos das equipes de atenção primária à saúde, somando-se aos profissionais já existentes nas equipes do SUS.

Por consequência, a concepção dos Núcleos de Apoio Saúde da Família deve ser reconcebida. Esses novos profissionais não serão mais apoiadores, mas membros efetivos das equipes de atenção primária à saúde. A relação desses novos profissionais poderá ser de um deles para cada três equipes da estratégia da saúde da família.

A equipes de saúde da família devem operar de forma interdisciplinar e com otimização do trabalho clínico. Isso é fundamental para racionalização da oferta no acesso à atenção primária à saúde (INSTITUTE FOR HEALTHCARE IMPROVEMENT, 2016).

A atenção multiprofissional interdisciplinar não é, apenas, um grupo de diferentes pessoas com diferentes profissões atuando juntas, na atenção primária à saúde, por um tempo determinado. O trabalho multiprofissional interdisciplinar é importante porque há um valor agregado de mais olhos e mais escutas, dos *insights* de diferentes corpos de conhecimentos e de um espectro mais amplo de habilidades e, por isto, tem sido considerado como um critério de qualidade da atenção primária à saúde.

Os modelos de equipes multiprofissionais interdisciplinares mais efetivos incluem: uma definição compartilhada do problema; o suporte às pessoas usuárias; a focalização compartilhada em problemas específicos; a elaboração compartilhada dos planos de cuidados; a confiança e o respeito mútuos; o monitoramento das condições crônicas; e o acesso pronto aos médicos especialistas (SINGH, 2005).

Para que o trabalho interdisciplinar se instale e para que a coesão se estabeleça são imprescindíveis as seguintes características: a definição de objetivos gerais; a definição de objetivos específicos mensuráveis; a implantação de sistemas clínicos e administrativos; a clara divisão do trabalho; a educação permanente de todos os profissionais; e o processo de comunicação efetivo. Há que haver uma clara divisão de trabalho na equipe.

É importante um sistema de educação permanente de todos os membros da equipe multiprofissional e isso deve ser feito em tempo protegido.

Devem estar bem estabelecidas estruturas comunicacionais como comunicação rotineira por papel, a comunicação eletrônica, a comunicação rápida face-a-face com interação verbal entre os membros da equipe e reuniões da equipe, bem como processos comunicacionais como oferta de *feedbacks* e técnicas de resolução de conflitos.

11.1 A autoavaliação da satisfação da equipe

Uma questão central no trabalho em equipe de atenção primária à saúde é a avaliação das percepções da satisfação dos seus membros. Para isso pode-se usar um instrumento apropriado.

Instrumento de autoavaliação da satisfação de um membro da equipe

1. Eu sou tratad@ com respeito, todos os dias, pelas pessoas que trabalham nesta unidade.

CONCORDO PLENAMENTE CONCORDO DISCORDO DISCORDO PLENAMENTE

2. Eu recebo tudo que necessito, materiais, instrumentos, equipamentos e motivação para desempenhar meu trabalho.

CONCORDO PLENAMENTE CONCORDO DISCORDO DISCORDO PLENAMENTE

3. Quando faço um bom trabalho, alguém na unidade reconhece que eu o fiz.

CONCORDO PLENAMENTE CONCORDO DISCORDO DISCORDO PLENAMENTE

4. Quão estressante é trabalhar nesta unidade?

MUITO ESTRESSANTE MEDIANAMENTE ESTRESSANTE POUCO ESTRESSANTE NÃO ESTRESSANTE

5. Quão tranquilo é perguntar a alguém a respeito de atender às pessoas?

MUITO FÁCIL FÁCIL DIFÍCIL MUITO DIFÍCIL

6. Como você classifica o moral da equipe e suas atitudes no trabalho nesta unidade?

EXCELENTE MUITO BOM BOM RAZOÁVEL RUIM

7. Esta unidade hoje é melhor para se trabalhar do que um ano atrás.

CONCORDO PLENAMENTE CONCORDO DISCORDO DISCORDO PLENAMENTE

8. Eu recomendo esta unidade como um local muito bom para se trabalhar.

CONCORDO PLENAMENTE CONCORDO DISCORDO DISCORDO PLENAMENTE

9. O que poderia tornar esta unidade melhor para as pessoas usuárias?

10. O que poderia tornar esta unidade melhor para os profissionais que trabalham aqui?

Fonte: Godfrey et al (2004)

11.2 A divisão de tarefas nas equipes

A divisão das tarefas por função dos membros das equipes da atenção primária à saúde deve estar, em boa parte, definida nas diretrizes clínicas baseadas em evidências.

A composição de cada equipe surge da discussão de como ela decide balancear a oferta e a demanda. A equipe deve compreender os serviços que provê e decidir quem deveria estar envolvido no trabalho e como o trabalho deveria ser dividido entre seus membros. Esse trabalho começa pela racionalização da demanda e depois da oferta, o que é diferente de estabelecer um padrão arbitrário de equipe e então tentar ajustar a demanda a este padrão.

É necessário que haja um planejamento prévio para cada função-chave de atenção em que fiquem bem claros os papéis de cada membro da equipe. Para isso, sugere-se a utilização de uma folha de distribuição das tarefas nas equipes que deve ser construída, em conjunto, por todos os seus membros.

Instrumento de distribuição da função e suas tarefas pela equipe da atenção primária à saúde

FUNÇÃO	MÉDICO	ENFERMEIRO	TÉCNICO DE ENFERMAGEM	ACS	Pa	Pb	Pn
Tarefa 1							
Tarefa 2							
Tarefa 3							
Tarefa 4							
Tarefa 5							

Pa, Pb, Pn: Outros profissionais

Fonte: Mendes (2012)

Um conceito chave para estabelecer as responsabilidades clínicas que otimizam o trabalho da equipe é ter todos os seus membros trabalhando no mais alto nível de expertise e habilidade. Isso deve presidir a divisão de tarefas entre todos os membros de uma equipe.

A satisfação dos profissionais tende a aumentar quando eles sentem-se confortáveis com seus papéis e responsabilidades.

A otimização do trabalho de uma equipe requer uma clara concepção das necessidades e das demandas das pessoas usuárias de tal forma que as demandas dessas pessoas possam ser balanceadas adequadamente com o cuidado prestado pela equipe (oferta). Isso requer a elaboração e uso de claras descrições de trabalho com definição de competências bem estabelecidas e, possivelmente, com atividades educacionais para viabilizá-las na prática (INSTITUTE FOR HEALTHCARE IMPROVEMENT, 2016).

11.3 O *cross-training* da equipe

O *cross-training* é a oportunidade dada aos profissionais para experimentar áreas diferentes da que ele atua proporcionando uma visão geral de todo o processo, o que permite que a equipe possa assumir diferentes trabalhos quando necessários.

A habilidade de uma clínica em responder a variações inesperadas de demanda e de oferta depende da flexibilidade da equipe em ajustar suas responsabilidades frente a estes eventos. O *cross-training* não nega a importância de que cada membro da equipe estar trabalhando no seu mais alto nível, mas provê uma opção para que um profissional desempenhe papéis de outro em situações inesperadas.

Para desenvolver uma equipe flexível e efetiva deve-se prover algum grau de *cross-training* de modo a que a equipe como um todo possa responder às variações rápidas de demanda e oferta ou a eventos inesperados. Como exemplos, profissionais que podem cobrir responsabilidades na equipe quando necessário, profissionais diferentes para participar de atendimentos compartilhados a grupo ou grupos operativos, profissionais diferentes capacitados a fazer exame de pés nas pessoas com diabetes ou profissionais diferentes capacitados a monitorar planos de autocuidado (INSTITUTE FOR HEALTHCARE IMPROVEMENT, 2016).

11.4 A redução da variação nos estilos e tempos de atendimentos entre os profissionais

Há muitas variações nos estilos e nos tempos de atendimentos entre membros de uma equipe e entre equipes distintas de uma mesma unidade de atenção primária à saúde que

interferem nos fluxos de pessoas usuárias e de atendimentos profissionais. Discussões abertas nas equipes sobre como o trabalho é feito podem ajudar a identificar oportunidades de estabelecer padrões que promovem eficiência no acesso (INSTITUTE FOR HEALTHCARE IMPROVEMENT, 2016).

11.5 As reuniões da equipe

Reuniões periódicas de equipes são fundamentais para aumentar a eficiência do acesso à atenção primária à saúde (INSTITUTE FOR HEALTHCARE IMPROVEMENT, 2016).

Uma estratégia eficaz é fazer uma reunião rápida da equipe antes de se iniciarem os atendimentos do dia, envolvendo, também, pessoal administrativo, para repassar a agenda do dia e identificar maneiras de fazer os fluxos correrem bem e, se necessário, elaborar planos de contingência para demandas inesperadas daquele dia. Essa reunião é feita em frente a um quadro de agendamentos, de maneira informal, inclusive para ganhar tempo, com as pessoas de pé.

Essa reunião tem como objetivo examinar todos os agendamentos do dia, antecipar as necessidades das pessoas marcadas para aquele dia e elaborar rapidamente um plano de contingência para uma situação inesperada. Por exemplo, a equipe pode discutir que pessoas que agendaram e não estarão presentes naquele dia (pessoas que cancelaram o agendamento ou que foram atendidas na semana anterior e não necessitam de retorno), que materiais e equipamentos deverão estar presentes nos consultórios e que serviços adicionais a equipe deve prover a uma pessoa que será atendida hoje de modo a tornar um retorno menos provável no curto prazo.

Outra estratégia, complementar à reunião rápida, é a reunião semanal da equipe. São reuniões de toda equipe, com maior duração, previamente agendadas para todo o ano. Essas reuniões valem-se muito das experiências das reuniões rápidas diárias, e nelas, a equipe concentra-se em avaliar e planejar seus papéis e responsabilidades, bem como em discutir as oportunidades para a melhoria do trabalho. Por exemplo, as reuniões rápidas definiram algum problema crítico no prontuário clínico e isso deve ser tratado, com maior profundidade, na reunião semanal.

Na última reunião semanal deve-se fazer um planejamento dos agendamentos para o próximo mês. Nela se revê a oferta real para o mês seguinte, estabelece-se a demanda e desenvolvem-se planos de contingência para superar os eventuais desequilíbrios entre

oferta e demanda para aquele mês. Por exemplo, planos de contingências redução de oferta por eventos inesperados como adoecimento de um membro da equipe e de faltas de materiais ou equipamentos ou de espaços para atendimentos clínicos.

12. A avaliação da longitudinalidade do cuidado pelas pessoas usuárias

Um dos objetivos do acesso à atenção primária à saúde é reforçar o atributo da longitudinalidade do cuidado e isto pode ser medido pela satisfação das pessoas usuárias com a atenção recebida por profissionais da equipe a que está vinculada. Esse indicador de satisfação refere-se à possibilidade de uma pessoa ter sido atendida por um membro da sua equipe que ela queria ver hoje.

A fórmula é: número de pessoas respondendo sim à questão da pesquisa: você foi atendido(a) pelo profissional da sua equipe que você queria ser atendido(a) hoje?, dividido pelo número total de pessoas pesquisadas, multiplicado por 100.

A meta é ter mais de 80% de pessoas entrevistadas com resposta sim.

O plano de coleta de dados consiste em criar um questionário com uma única pergunta: você foi atendido(a) pelo profissional da sua equipe que você queria ser atendido(a) hoje: sim ou não?

Escolha uma semana de cada mês, selecione randomicamente 50 pessoas e aplique o questionário no check-out da unidade.

13. A avaliação do acesso

É importante medir a satisfação do acesso à atenção primária à saúde. Isso deve ser feito por tipo de acesso: acesso por telefone, acesso por internet e acesso presencial na unidade de atenção primária à saúde.

A fórmula é: número de pessoas que consideram o acesso como excelente, dividido pelo número total de pessoas pesquisadas, multiplicado por 100.

A meta é ter mais de 80% das pessoas entrevistadas com resposta excelente.

Outra meta deve ser aumentar relativamente os atendimentos à distância (por telefone ou internet) diminuindo os atendimentos presenciais. A razão disso é que os atendimentos presenciais na unidade apresentam custos de oportunidade mais altos para as pessoas.

O plano de coleta de dados consiste em criar um questionário com uma única pergunta: como você considera sua experiência recente de agendamento por telefone, por internet ou presencial na unidade: ruim, regular, boa, excelente?

Escolha uma semana em cada mês, selecione randomicamente 50 pessoas para cada tipo de agendamento e aplique o questionário no *check-out* da unidade.

14. Avaliação do tempo de espera por atendimento

O objetivo é medir a satisfação das pessoas com o tempo de espera para obter o atendimento hoje (tempo decorrido entre a solicitação e o atendimento efetivo).

A fórmula é: número de pessoas que consideram o tempo de espera excelente, dividido pelo total de pessoas entrevistadas, multiplicado por 100.

O plano de coleta de dados consiste em criar um questionário com uma única pergunta: como você considera o tempo de espera para o atendimento realizado hoje: ruim, regular, bom ou excelente.

Escolha uma semana em cada mês, selecione randomicamente 50 pessoas e aplique o questionário no *check-out* da unidade.

REFERÊNCIAS BIBLIOGRÁFICAS

- AMERICAN ACADEMY OF FAMILY PHYSICIANS. Group visits: shared medical appointments. 2005. Disponível em:
<http://aafp.org/online/en/home/practicemgt/quality/qitools/pracredesign/january05>
- BACKER LA. Strategies for better patient flow and cycle time. *Fam Pract Manag*, 9: 45-50, 2002.
- BATALDEN PB et al. Unplanned activity card. Hanover, Trustees of Dartmouth College, 2003.
- FOGLIATTI MC, MATTOS NMC. Teoria de Filas. Rio de Janeiro, Editora Interciência, 2007.
- GODFREY MM et al. Patient cycle tool. Hanover, Trustees of Dartmouth College, 2003.
- GODFREY MM et al. Clinical microsystem action guide. Hanover, Center for the Evaluative Clinical Sciences at Dartmouth/Institute for Healthcare Improvement, 2004.
- GROUP HEALTH COOPERATIVE. Visit starter kit. 2001. Disponível em: <http://www.ghc.org>.
- INSTITUTE OF HEALTHCARE IMPROVEMENT. Resources. Disponível em:
www.ihc.org/resources/Pages/default.aspx. Acesso em julho de 2016.
- INSTITUTE OF MEDICINE. Transforming health care scheduling and access: getting to now. Washington, The National Academies Press, 2015.
- JAMOULLE M et al. Working fields and prevention domains in general practices/Family Medicine. 2002. Disponível em: <http://docpatient.net/mj/prev.html>
- LITVAK E. Reengineering flow through the primary care office. In: INSTITUTE OF MEDICINE. Transforming health care scheduling and access: getting to now. Washington, The National Academies Press, 2015.
- MENDES EV. O cuidado das condições crônicas na atenção primária à saúde: o imperativo da consolidação da estratégia da saúde da família. Brasília, Organização Pan-Americana da Saúde, 2012.
- MENDES EV. O acesso à atenção primária à saúde. Brasília, CONASS, 2016.
- MURRAY M. Open access on advanced access scheduling. In: Institute of Medicine. Transforming health care scheduling and access: getting to now. Washington, The National Academies Press, 2015.
- NATIONAL ACADEMY OF ENGINEERING AND INSTITUTE OF MEDICINE. Building a better delivery system: a new engineering/health care partnership. Washington, The National Academies Press, 2005.
- QUINN JB. Intelligent enterprise: a knowledge and service based paradigm for industry.
- RAMOS V, CARRAPIÇO E. Pessoas que consultam frequentemente. In: GUSSO G, LOPES JMC. Tratado de medicina de família e comunidade: princípios, formação e prática. Porto Alegre, Artmed, Tomo I, 2012.
- SINGH D. Transforming chronic care; evidence about improving care for people with long-term conditions. Birmingham, Health Services Management Centre/University of Birmingham, 2005.
- VESTED P, CHRISTENSEN MB. Frequent attenders in general practice care: a literature review with special reference to methodological considerations. *Public Health*, 119: 118-137, 2005.