BMA

Workload Control in General Practice

Ensuring Patient Safety Through Demand Management



British Medical Association
bma.org.uk

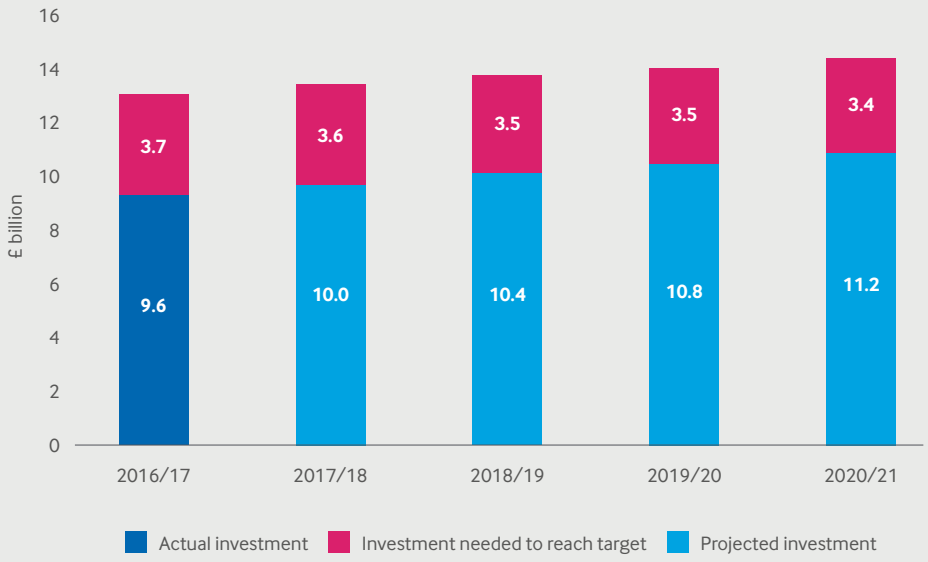
1 Introduction

There is an urgent need for a campaign led by the GPC which focuses on the problems caused by the current GP workload within the NHS. This issue affects most practices and is of particular significance to GP partners, although those under employment contracts often report similar problems. It is certainly a factor behind the rise in the number of "independent" or locum GPs, who have chosen to work flexibly and maintain a better work-life balance, rather than take on unmanageable workloads.

1.1 Why workload has increased

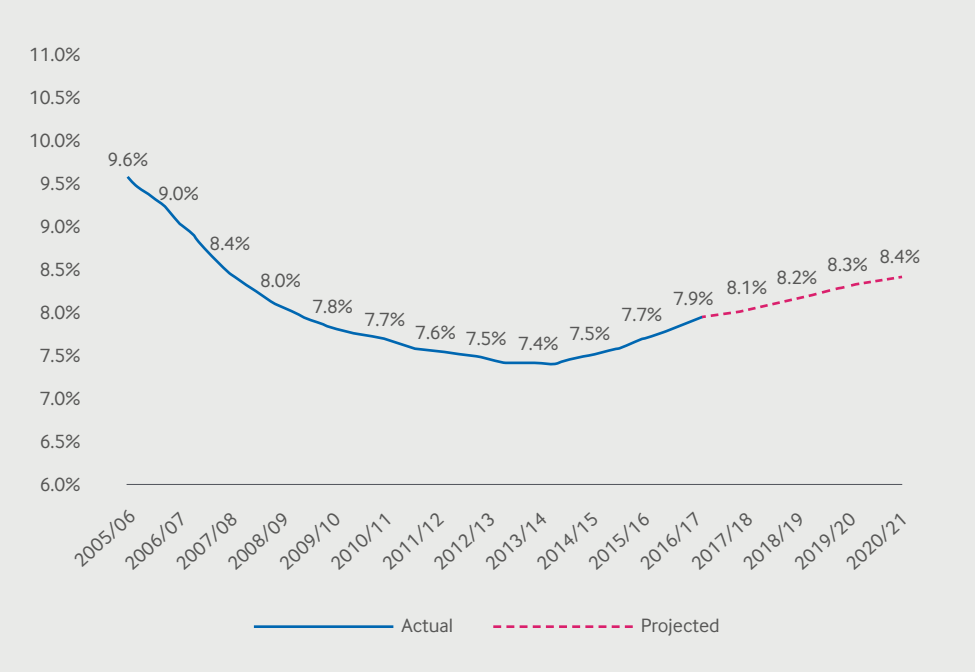
- Genuine and growing patient needs (complex multi-morbidity) within primary care have produced a consultation rate in the UK which is 2-3 times that of comparable EU populations.
  - Between 2007 and 2014 overall consultation rates for GPs in England rose by 13.6% (Oxford University, 2016). Consultations grew by more than 15% between 2010/11 and 2014/15 (King's Fund, 2016).
- Long term lack of investment.

Figure 1 – Investment excluding drug reimbursement (cash terms) – 2016/17 – 2020/21



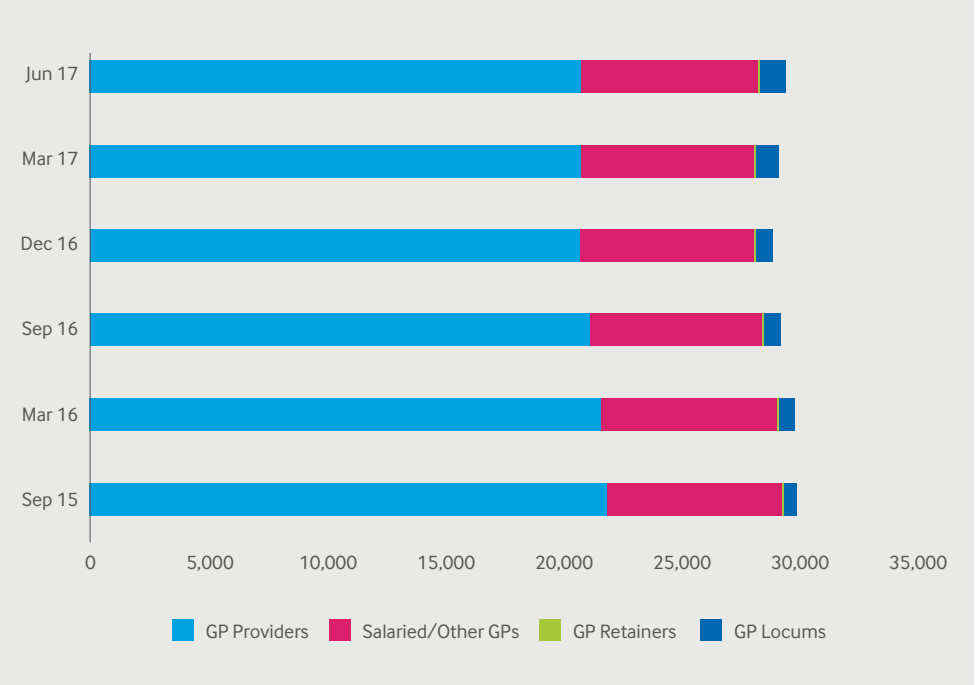
- BMA has a target of general practice receiving 11% of the NHS budget (excluding drug reimbursement)
- Current investment falls £3.7 billion short of this target.
- The proportion of the NHS budget going to general practice, excluding the reimbursement of drugs, has fallen from 9.6% in 2005/06 to 7.9% in 2016/17. An extra £2 billion could have gone to general practice this year if funding met the levels of 2005/06.

Figure 2 – Proportion of NHS budget to general practice



– A widespread recruitment and retention crisis

Figure 3 – Number of Full Time Equivalent GPs in England since the introduction of the (primary care) Workforce Minimum Dataset (2015)



Source: NHS Digital General and Personal Medical Services, England As at 30 June 2017, Provisional Experimental statistics (August 2017)

- From September 2016 to September 2017, the total number of FTE (full time equivalent) GPs fell by 1193 (-3.4%). (NHS digital)
- Too few doctors are choosing general practice as a career and many GPs are reducing their time commitment. The BMA's 2016 survey found that 34% of partners are exploring alternative working options and 75% of sessionals are put off partnerships due to excessive workload.
- Others are leaving altogether. Experienced clinicians are choosing to leave the profession earlier than previous generations of GPs, often due to the unrelenting demands of the service.
- Increased demand fuelled by politically inspired targets around access eg. weekend and evening access by 2020 (GPFV).
- Lack of coordinated system reform leading to stagnation in many partnerships. Practice failures, contract terminations and GP performance concerns can largely be attributed to an impossible workload and the resultant inability to recruit.

1.2 Current and planned changes within the NHS in England

- Increase in medical school placements and recruitment from overseas to increase capacity and address rising agency/locum costs. Due to the length of training, the benefits of increased student and trainee numbers will not be realised for some years.
- Focus on a wider skill mix in general practice with the introduction of new professionals, such as physician associates and paramedics, as well as a renewed emphasis on the wider GP-led multi-disciplinary team.
- Planned movement of work into community settings.
- Continued support of list based general practice.
- Universal plans for at-scale working.
- Competition and a pseudo market environment in primary care.
- Rapid movement to New Models of Care and population based contracts.
- Growth of super-partnerships and practice-led companies and federations.

1.3 The case for GP workload control

- The need for a cultural change within general practice from one of 'quantity overload' amid shrinking resources to one of efficient demand and workload management.
- The need to attract new recruits into general practice and retain the existing workforce – too many GPs report being burnt out and are being pushed into reducing their clinical commitment or early retirement due to the demands of the job.
- The need to maintain the partnership model to protect the future of independent general practice – many practices currently find it impossible to recruit into partnerships.
- The need to create capacity within the community
- Lack of capacity in general practice impacts other parts of the NHS, for example patients who cannot get a GP appointment may choose to attend A&E, putting more pressure on secondary care.
- All other parts of the health and social care system have the ability to limit workload and activity and this often leads to an unfunded shift of workload to GPs. The GMS contract has no ability to keep pace with such increased demand.

2 A Workload Control Strategy

2.1 Context

The strategy outlined below will be introduced into an environment which is moving towards locality commissioning and integration within the NHS in England, and where work has already been carried out by the GPC, NHS England and others to demonstrate the need to reduce unnecessary or misplaced workload and to illustrate the contribution of other clinicians.

This strategy is intended to seamlessly fit into these ongoing changes and such work will therefore not be described here in any detail.

2.2 Aim

The aim of this strategy is to enable general practice to improve quality and safety, and to address the recruitment and retention crisis, by agreeing and publicising reasonable safe workload limits, and by providing practices with practical tools with which to achieve workload control.

2.3 Objectives

- Agree a range of clear quantitative limits to help individuals identify what safe practice looks like for them.
- Produce guidance on the implementation of safe practice across scenarios, illustrating common practice working patterns such as telephone consultations and triage.
- Endorse or promote the implementation of system change which allows the provision of safe general practice.
- Propose contractual innovations for practices where rurality or other factors hamper system change.
- Introduce an "OPEL Alert" system for use by practices and LMCs. **See Appendix.**

2.4 Projected Benefits

- Improved patient safety and care in general practice:
  - Improvement in the management of long-term conditions through more focused, less rushed appointments and greater levels of continuity
  - Reduced hospital admissions
  - Patients increasingly valuing the service as demand management requires more patient acceptance and co-operation.
- Long-term recruitment and retention benefits by making general practice a safer and more manageable career.
- Improved GP morale and wellbeing.
- Practices and CCGs should together see the benefits of safe working at a locality level.
- Locality working becomes supportive and practice focussed.
- Practices increase their perceived and real value to the NHS.
- An integrated primary care system gives general practice a stronger voice in any planning for an integrated care arrangement or similar strategic change.

3 Action Plan

3.1 Agree quantitative limits to individual safe practice for GPs

Without a recognised and realistic safe limit to individual GP workload there is no opportunity to quantify limits to the GMS contract and clearly therefore no possibility of 'alerts' being acceptable within the system. The BMA's "Safe working in General practice" document (2016) was the first recent attempt to introduce the idea of a maximum workload capacity.

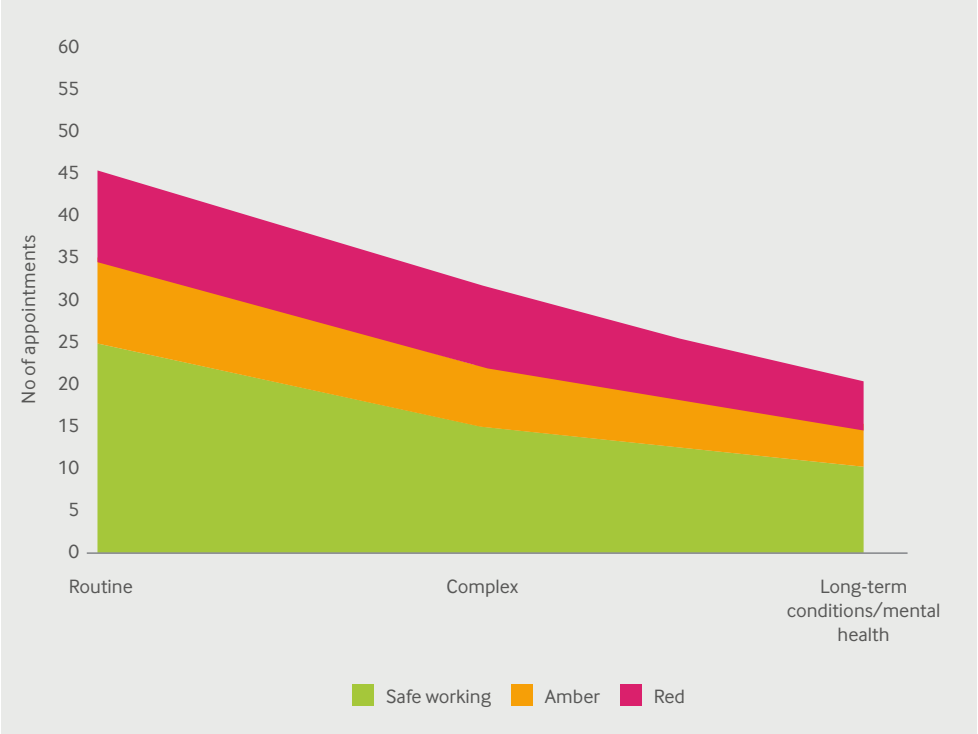
Appropriate limits on workload will depend on the unique circumstances of each practice and the preferences of each individual GP, as well as the complexity of care being provided. There will also be variation in the amount of spinoff work depending on the complexity of the case mix and also on the contractual status of the doctor. This will include:

- Report/letter writing, reading and recording
- Pathology results
- Follow-up of hospital appointments, etc

Therefore, this document presents a range in the number of appointments per day to show when a GP may move into more 'unsafe' practice (red), depending on whether the appointments are routine or complex. For example, up to 25 routine doctor-patient contacts a day could be deemed 'safe', with GPs only reaching 'unsafe' working levels at 35 or more routine patient contacts per day. In comparison, anything over 15 doctor-patient contacts for long-term, complex or mental health conditions could be said to be 'unsafe' due to the more demanding nature of the consultations.

There is surprising little research on "safe" levels of working, although there is evidence around doctor fatigue and an increase in errors or mistakes. Limiting appointment rates, or any other rate limitation method, will require improved triage and care navigation. Triage has been introduced in a largely haphazard and uncoordinated way, rather than as part of planned system change.

Figure 4 – Safe working in general practice



A GP working in the 'unsafe' red range should trigger a practice action and a practice on red should initiate an 'alert', similar to the OPEL system used in other parts of the NHS.

3.2 Produce guidance on implementation of safe practice across scenarios illustrating common practice working patterns

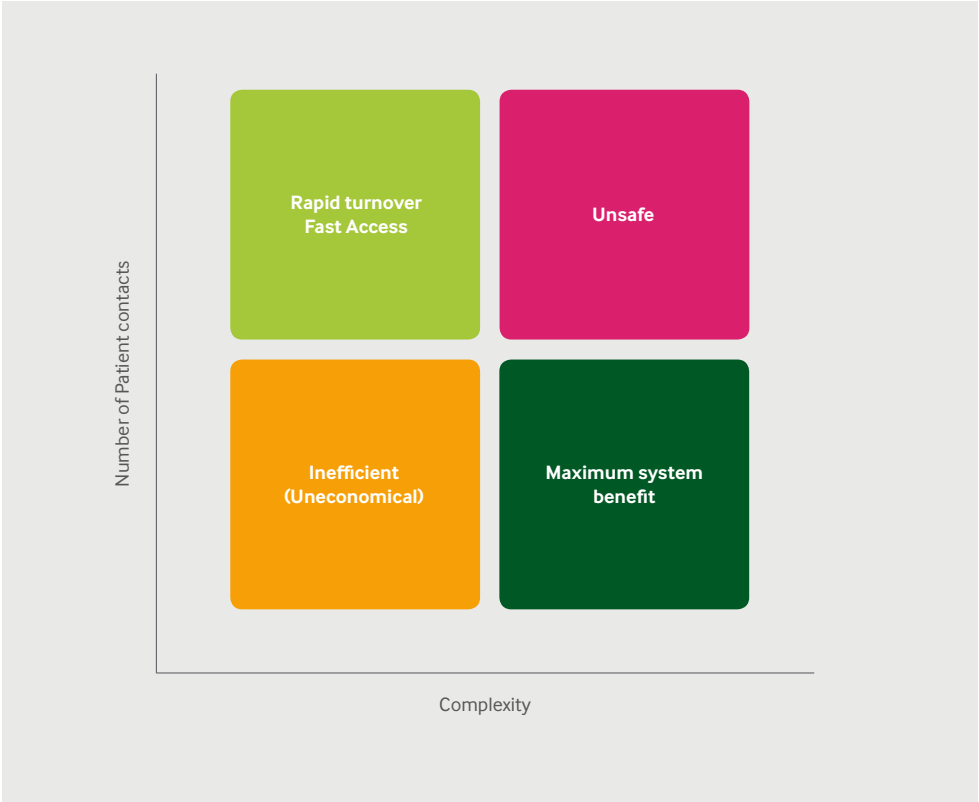
This piece of work will require the GPC to make judgements on how differing work patterns fit within our definition of "safe", and how varying tasks can be effectively traded within a practice in order to ensure both safety and equity. For example, 4-5 face to face consultations in an hour is equivalent to X telephone appointments or Y telephone triage calls. GPC guidance would allow practices to tailor safe practice to their own needs and skills.

An alternative to using appointment numbers as a measure of workload would be to use the hours worked in a day, possibly averaged over the week. A maximum safe list size might also be postulated, although with varying use of other professionals and increasing integration across localities this is increasingly difficult to implement.

The great difficulty in implementing a change to safe working within a practice is the inability to divert patients to other trusted and integrated primary care settings. A practice facing recurrent red situations can close its list but do little else to address the problem.

The profession may also have difficulty with the concept of proposing specific limits to workload due to concerns about external micromanagement.

Figure 5

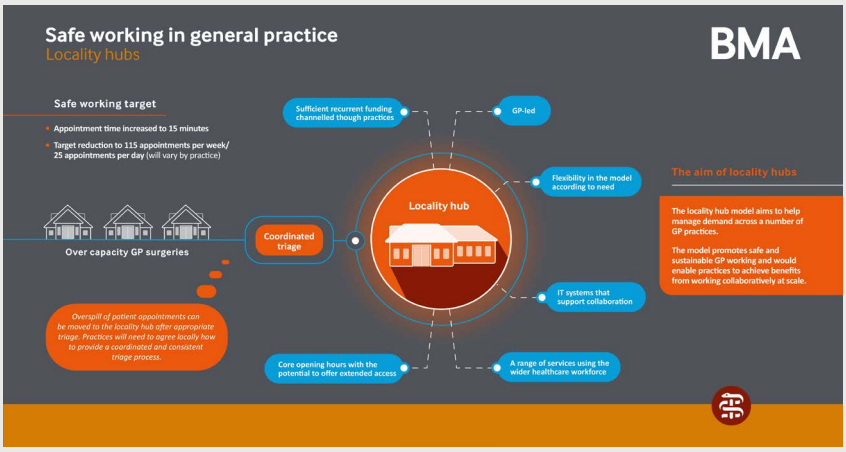


3.3 Endorse or promote the implementation of system change which allows the provision of safe general practice

In the majority of areas in England this will mean collaboration with overflow Hubs. It is difficult to envisage any other way by which identified need can be managed within the community. This can only work successfully if the Hub is the servant of the practices and patients are informed that the Hub is an integral part of the practice and locality/community.

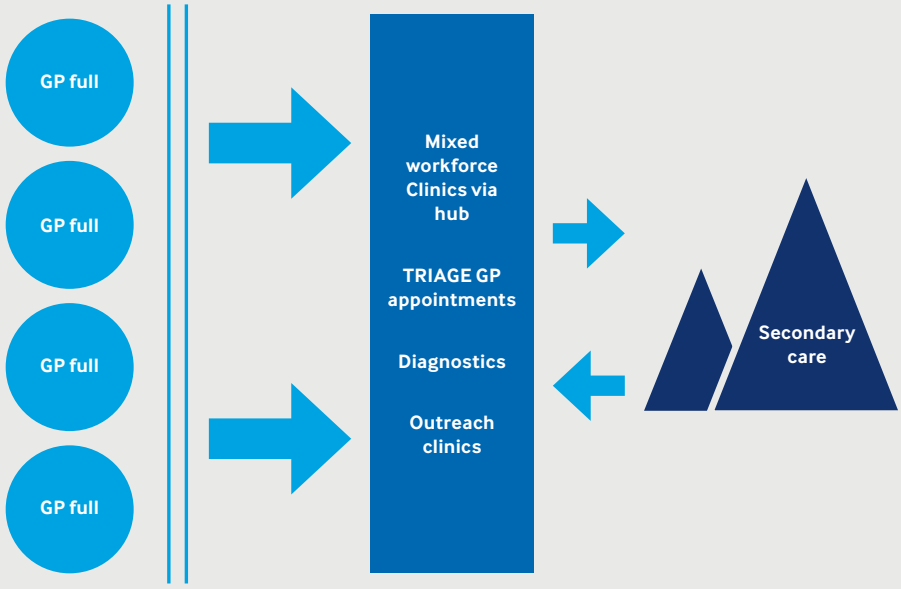
The figures below illustrates the basis under which Hubs could be used to protect practice and function as "overspill "centres for practices in difficulty.

Figure 6 – Safe working in general practice



Click here to enlarge

Figure 7



The advantage of linking to locality working is that it allows a practice problem to become a locality and even a wider ICS (Integrated Care System) problem. The Hub may need to reduce routine out of hospital care and specialist clinics in order to deal with primary care demand. Localities could also be more acceptable environments in which to agree and monitor local workload limits, and access hubs linked to GP surgeries could allow continuity of care to be preserved in an increasingly integrated but complex system.

3.4 Propose contractual innovations for practices where rurality or other factors hamper system change

There will be practices where working in a locality and sharing resources, and therefore workload, is simply impracticable due to isolation or other unusual factors. A contractual solution whereby such practices were identified and classified in terms of their entitlement to additional resources might assist such cases. It would then be for the practice to utilise the additional funding to address safety issues (rural practice allowance).

A similar allowance might also be needed for extreme urban practices such as those in inner London where the relationship with Trusts is unique.

3.5 Introduce an 'OPEL alert' system for use by practices and LMCs

This objective is only effective in the context of an accepted and widely implemented safe working limit. The alert can then be used to inform patients, localities, CCGs, LMC, NHS England, secondary care and other providers, of the situation and its likely duration. (It could mimic the Operational Pressures Escalation Levels Framework already in use in other NHS sectors.)

Local press and Healthwatch are also likely to be interested and CCGs will soon be obligated, under pressure from NHS England, to address the issues which have given rise to the alert.

4. Anticipated Risks with the Strategy

- Resistance from GPs in general and partners in particular – this is by far the greatest hurdle and unless a safe limit can be accepted by the profession then it can never become an objective to be achieved. Everything else flows from this initial step but it will require a cultural change to remove the current noble but potentially self-destructive urge within general practice to simply work harder and longer to meet patients' needs. It is worth noting that public satisfaction with general practice dropped by 7 percentage points in 2017 to 65 per cent, the lowest level in 35 years. Somehow the leaders of general practice must convince their colleagues that to change to safe working practices is not an admission of failure but is instead a crucial step in securing the long-term survival of partnership based general practice.
- Continuity of care will be potentially compromised if localities are not managerially accountable to practices.
- Resistance from commissioners – many commissioners are well aware of our problems and are sympathetic to the need for action. A more integrated system appeals to managers and offers the opportunity to reduce unacceptable levels of variation across primary care.
- Opposition from other providers – the current system allows irregular and unresourced shifting of additional work onto general practice. A more closely integrated locality model might resolve this issue by producing a critical mass within a Hub or similar system. The locality may become the controller of patient flows out of hospital.
- Locality working in this model is dependent on collaboration with community providers.
- Lack of acceptance from patients, pressure groups, politicians and the media that the system needs to change and that this will actually be beneficial for patient care and safety.
- Resistance from secondary and community care sectors due to potential short-term disruption – need to convince them of the longer-term benefits.

Next steps

- Agree the principles behind safe working and work with other organisations to promote its introduction
- Undertake further work to specify precise safe limits to workload in practice settings. (Expressed in appointments, time or list size)
- Produce resources for practices and locality groups with examples of how this model of working can be introduced
- Endorse a locality approach which supports groups of practices, or LMCs, in setting their own safe limits
- Collect and publish examples of hub-based working and workload control from around the country

Appendix

An OPEL Alert system for General Practice

The following system will only succeed within a locality where a range of safe working levels has been discussed and agreed, and where practices are assured that all other providers are operating above agreed minimum quality and capacity measures. Such a system will need peer controlled audit and free exchange of information.

Level 1

Individual GPs or other clinicians are at or above the locally agreed safe level of working.

Action: Internal practice action to divert patients to other clinicians or alternative providers or later sessions/days. Notify position to Locality Hub.

Level 2

One practice reaches unsafe working levels.

Action: Practice reports status to Locality Hub. Hub initiates improved triage and allows practice access to booked overspill appointments.

Level 3

Several locality practices reach unsafe working levels

Action: Hub initiates improved or centralised triage and releases all overspill appointments. Other practices and Hub clinics are alerted to potential "Black Alert" status.

Level 4

All practices in the locality report unsafe working.

Action: Locality Hub switches available workforce to address overspill (Triaged and booked appointments) thereby cancelling routine LTC and specialist clinics. Locality alerts Hospital that discharge management will be suspended until primary care recovers. Planned early discharges are therefore temporarily stopped.

British Medical Association
BMA House, Tavistock Square,
London WC1H 9JP
bma.org.uk

© British Medical Association, 2018

BMA 20180074