MIDLANDS AND LANCASHIRE
COMMISSIONING SUPPORT UNIT
Strategy Unit

NHS England

Utilisation review

A scoping study

Prepared for the West Midlands Clinical Senate

October 2014

Produced for the:
West Midlands Clinical Senate
St Chads Court, 213 Hagley Road, Edgbaston, Birmingham, B16 9RG, United Kingdom
Tel: +44 (0)113 825 3257
Email: england.wmcs@nhs.net

Contents

Acknowledgements

We would like to thank the following individuals for giving their time and information to support this study:

Kathleen Coniglio – VHA

Peter Ellis – Medworxx

James Ferris - Medworxx

Alistair Frost – Oak Group

Peter Huskinson - NHS England

Matthew Loxton – WBB Consulting

David Maltz – Oak Group

Jason Nerval – Monmouth Partners

Mandy Salter - Medworxx

Rachel Sharp - Greater East Midlands Commissioning Support Unit

Neil Spragg - Independent Consultant

Serge Teplitsky – Kaiser Permanente

Executive summary

Increasing demand and financial pressures have led to questions on how effectively and efficiently resources are being used. Alongside this, there is a growing sense that some care may be unnecessary and even harmful (Restuccia, 1995). Utilisation review (UR) has emerged as a methodology which explores the extent to which resources are appropriately used, systematically applying a set of objective clinical criteria to admissions, bed days and episodes, and assessing the appropriate level of care. The Strategy Unit was commissioned by West Midlands Clinical Senate to undertake this scoping exercise to review the UR methodology and its utility. This review assesses the published evidence base, reported experience and the current offer in the market.

The concept of UR originates in the US where it was devised in the 1960s as a method of cost containment of state funded (Medicare/Medicaid) programmes and is now standard practice in multiple health systems. During our study, we have explored case studies from the US, Kaiser Permanente and Veterans Health Administration, to identify key learning which could transfer to the NHS.

NHS England (2013) has recently encouraged uptake of UR for specialised services and inpatient services, through the 2014/15 CQUIN (*Patient Flow improvement through Clinical Utilisation Review*).

An underdeveloped evidence base

The published evidence base on the use of UR tools and methods in the UK is very limited and much of what has been published is well over 10 years old; as far back as 1996, Coast (at the University of Bristol) was calling for more robust research to test the cost effectiveness of utilisation review but the evidence base remains relatively underdeveloped. There are examples of interesting work within the NHS, particularly in the East Midlands, but very few projects are formally published. There is therefore an opportunity for formal evaluation and research to further develop the evidence base.

Used effectively, UR can provide evidence in relation to patient flow within an organisation - for example, the prevalence of admissions for potential treatment in primary care and delays. Operationally, embedded UR solutions can support demand and capacity management, through the provision of real-time, standardised data and suggest actions for improving resource utilisation. Anecdotal feedback from case study sites suggests that strategically, such solutions can provide an evidence base to highlight opportunities for service improvements and redesign by identifying inappropriate but justified care, variations in care, delay trends and inefficiencies.

There are some criticism of UR reported in the literature, some of which have since been addressed by changes such as PbR. One of the key limitations of UR, and other interventions aimed at shifting care, is the lack of knowledge on the costs (including opportunity costs) and feasibility of providing care in alternative settings. There is also a possible risk of unintended consequences to other services e.g. community services; a health-economy wide view may help.

Gertman and Restuccia (1981) who developed the Appropriateness Evaluation Protocol, which is one of the key instruments used in UR, acknowledge that criteria such as the AEP are an attempt to establish an absolute measure and will therefore produce higher estimates of inappropriate use than subjective techniques which will factor in adjustments. They note that 100% appropriateness is "neither practical nor desirable" but do not suggest what an optimum level might be; this is likely to be influenced by local

availability of services, case mix and other factors. Criteria are usually assessed to measure validity and reliability. It is difficult to measure validity due to the lack of a real "gold standard" (Phelps, 1993); typically, studies will use the gold standard of an expert panel but a consensus-based panel itself is subject to bias.

UR in the NHS

Numerous NHS organisations have invested in utilisation review, either in the form of standalone audits to diagnose local issues or embedded solutions to provide real-time monitoring. Within the West Midlands, the Heart of England Foundation Trust and Shrewsbury and Telford Hospitals NHS Trust have undertaken reviews recently and a number of other trusts are in discussions with suppliers. The East Midlands has led a systematic and coordinated approach to UR: over 4 years, the team has conducted 35 different reviews, including snapshot audits, a live concurrent embedded programme, readmissions reviews and a mental health pilot in older adults.

Utilisation review is likely to offer greater value if it fits within a broader approach of mutually reinforcing interventions - Kaiser Permanente is an example of such an integrated approach. The Better Care Fund may present an opportunity for a health economy-wide approach.

The current UR market

One of the key arguments for UR focuses on the application of objective, clinical and evidence-based criteria. It is suggested that clinical judgment alone to determine the appropriate level of care is limited, due to the complexity and subjective nature of decision making and the influence of many discretionary factors. Embedded solutions from Medworxx and the Oak Group have built in criteria.

Our scoping study has identified a small number of organisations currently working within the UK, in the field of utilisation review. These organisations vary in their offers: consultancy support for diagnostic audits, change management and/or embedded solutions. Some of the large consultancies (e.g. Capita, EY, McKinsey) have also worked in this area. There are currently two suppliers offering embedded solutions within the UK market: Medworxx and the Oak Group. Within the larger US market, there are more suppliers, notably, McKesson and MCG. Within the NHS, there are two commissioning support units (CSUs) offering utilisation review as a service: Greater East Midlands CSU (GEMCSU) and North West CSU (NWCSU).

Key recommendations

Importantly, our study shows that, as with other improvement interventions, UR is likely to be less successful in isolation. Where UR has proved successful, it has been integrated into an overall approach to managing avoidable admissions and length of stay with realistic expectations, and alongside other interventions which are mutually reinforcing.

UR may be best addressed at a health economy level rather than at an organisational level, to address wider strategic as well as operational issues and to avoid unintended consequences of pushing issues downstream. Health economies should be clear as to why they are undertaking UR and ensure the right people are committed and involved from the outset. Expectations need to be realistic in relation to how long it may take to embed and sustain change.

The applicability of US-derived clinical criteria to a UK context does not appear to be an issue. Although early research suggested this may be a problem, suppliers such as Medworxx and the Oak Group have adapted criteria to comply with NICE guidance; the largest UK contract (East Midlands) stipulates ongoing

annual assurance of NICE compliance as a contractual obligation. This suggests it would not be worthwhile to develop a bespoke tool tailored to the West Midlands, particularly given the resource involved in creating and maintaining such a tool.

The case studies we have explored show that improvement of services and processes depends on behaviour change and a structured change management approach. UR is a significant investment and to demonstrate value for money, should be underpinned by a clear theory of change underpinning decisions. There will be impacts on roles and responsibilities (for example, retraining discharge planners as case managers) as well as how things are done and this will lead to cultural change; learning from case studies suggests that integration of UR into existing processes can help to embed UR within the organisation.

The importance of senior clinical buy-in is reflected in the case studies we have explored and echoed in a recent report from the Advisory Board Company (2014) which emphasises the early involvement of clinicians to ensure success of change initiatives in patient flow. The report points out that support and compliance is likely to be increased by engagement from the outset.

Establishing routine UR will require dedicated resource and appropriate governance. Training and support is important to ensure tools are used consistently; for example, Medworxx test knowledge on an annual basis and McKesson has established routine testing of inter-rater reliability.

For organisations or health economies aiming to complete a standalone audit to give a snapshot view, the methodology needs to be carefully planned to avoid potential bias. There are a number of options regarding the methodology which impact on the validity of the results and have implications for subsequent modelling.

Developing the evidence base

The challenge we have faced in synthesising the evidence base is that much of the published research is now dated and mostly relates to one-off retrospective reviews – our search did not identify any primary research assessing the effectiveness of embedded utilisation review solutions. New primary research would further develop the evidence base, enabling a better understanding of the utility and impact of utilisation review. There are two areas in particular which we suggest should be addressed by new primary research:

The validity of UR tools. As indicated in our report, it has been challenging to test the validity of UR tools due to the lack of a gold standard. Typically, UR tools have been compared to expert consensus panels, which are subjective. Phelps (1993) suggests a more statistically valid approach using maximum-likelihood methods which enable an estimate of the true prevalence of inappropriate admissions/days. It would be helpful to test this method in an NHS setting.

The effectiveness of UR. The gold standard for evaluating an intervention is an RCT, including an economic analysis to assess the cost of the intervention and any savings achieved. An RCT could involve randomising different wards in a trust: to the intervention group, implementing UR or to the control group, who continue with standard practice. This method reduces the risk of bias and facilitates statistical analysis. However, trials can be expensive and resource-intensive, and can take significant time. A mixed methods approach would be useful in collecting qualitative data as well as quantitative data, such as stakeholder (patients, staff, commissioners, GPs etc.) responses to utilisation review, which could capture key lessons which could be shared across the health economy and NHS.

As with any new intervention, it will be important for health economies to evaluate the local impact. This should be planned at the start of any change initiative to ensure the baseline position is measured and to ensure adequate resource (time, people, money) to complete the evaluation.

Critical success factors

Strong and consistent senior leadership to ensure UR is used to make improvements

Senior clinical sponsor to lead UR in each organisation

System-wide approach to address wider strategic issues and unintended consequences

A focus on quality not solely efficiency

A systematic approach to identifying and managing issues impacting on patient flow

Explicit change management to support behaviour change behaviours

Early involvement of clinicians and multidisciplinary teams to ensure ownership

Training and support

Careful consideration of underlying assumptions and tradeoffs

Locally agreed objectives and outcomes

Consideration of the end to end patient journey

Processes to ensure consistent use of UR tools

Commitment to a continued and sustained approach

Contracts which reflect specific requirements e.g. assurance of NICE compliance

Context

It is increasingly apparent that insight is required on whether people are being cared for in the most appropriate settings, and for this insight to be granular and clinically meaningful (to allow for proper consideration of planning alternatives).

Experience is suggesting that the Clinical Utilisation Review (UR) methodology is seen as a way to do this, both for one-off stock-take reviews but also, potentially for monitoring systems or decision support systems that include feedback loops for education and development (a continuous improvement system).

UR is a service that a number of trusts and commissioners have either bought or are considering buying in the near future (often by companies such as The Oak Group, Medworxx and McKesson). It is often seen as a way of gaining quantified insights into the questions 'are there people occupying hospital beds who could be better treated elsewhere ' and 'what would it require to enable that shift to occur'. It typically requires a substantial outlay in terms of cost.

However, there are questions about the currently available tools:

- Are they sufficiently attuned to the English context (in terms of clinical algorithms)?
- Is the sampling methodology sufficiently robust to support planning and modelling requirements?
- Do they support the kind of prospective clinical use that some economies envisage and do they support 'continuous improvement' systems adequately?
- Do they offer best VFM or are they of limited utility despite considerable local expenditure?

The Strategy Unit was commissioned by West Midlands Clinical Senate to undertake this scoping exercise to review the methodology and its utility. This review assesses the published evidence base, reported experience and the current offer in the market; and considers whether there is a case to seek an improved offer.

Introduction

Increasing demand and financial pressures have led to questions on how effectively and efficiently resources are being used. Alongside this, there is a growing sense that some care may be unnecessary and even harmful (Restuccia, 1995).

Utilisation review (UR) has emerged as a methodology which explores the extent to which resources are appropriately used, systematically applying a set of objective clinical criteria to admissions, bed days and episodes, and assessing the appropriate level of care. Criteria typically consider the severity of the condition, risk of mortality, medical history and co-morbidities.

Defining appropriate levels of care

"Inappropriate use of a facility comes about through providing unnecessary care, through providing necessary care through a resource not suited for the level of care actually provided or required, and through less than complete use of time during the course of care. Such inappropriate use is responsible for hidden capacity ("latent service reserve") which, if released, would reduce or obviate the need for resource expansion".

Donabedian (1973)

Origins

The concept originates in the US where it was devised in the 1960s as a method of cost containment of state funded (Medicare/Medicaid) programmes; by 1976 90% of hospitals had a UR programme (Smeets et al, 2000) and the approach was later adopted by health insurance companies.

Within Medicare/Medicaid programmes, UR teams are standard, comprising registered nurses and physician advisors who coordinate transfer of patients not deemed to be at the right level of care (Foster and Prairie, 2014); these teams are expected to use industry standard criteria or develop "evidence-based policies".

Typically, UR is based on a standard instrument, often tweaked to suit local context, designed to assess the appropriateness of admission or a bed day. Many of the instruments in use have been developed in the US and adapted for a European or UK context.

UR has been recommended policy since 2006 (Department of Health, 2006) with the aim of enabling:

- more effective planning;
- facilitating the shift from inpatient to community-based care;
- limiting waste;
- and reducing admissions.

NHS England (2013) has encouraged uptake of UR for specialised services and inpatient services, through the 2014/15 CQUIN (*Patient Flow improvement through Clinical Utilisation Review)* with funding available for:

- initial baseline diagnostic reviews (using a recognized tool);
- implementation set up;
- implementation roll out;
- and benefits realisation.

The evidence base

The published evidence base on the use of UR tools and methods in the UK is very limited and much of what has been published is well over 10 years old; as far back as 1996, Coast (at the University of Bristol) was calling for more robust research to test the cost effectiveness of utilisation review but the evidence base remains relatively underdeveloped. There are examples of interesting work within the NHS, particularly in the East Midlands, but very few projects are formally published. There is therefore an opportunity for formal evaluation and research to further develop the evidence base.

There is some scepticism in the literature about the validity of criteria (e.g. Kalant et al, 2000) and of their application to a UK context (e.g. McDonagh et al, 2000; Werneke and MacFaul, 1996) but it should be noted that this literature is now quite dated and is focused on standalone audits rather than embedded systems. One of the key challenges in evaluating UR is the lack of a robust gold standard; in earlier studies, the comparison intervention is typically consensus of a clinical group, which itself is subject to bias. Our search did not identify any published empirical evidence relating to the effectiveness of embedded UR solutions.

The literature reports a shortage of validated public domain criteria and a small market of proprietary criteria. The evidence base on proprietary tools is limited due to commercial sensitivities, particularly in the more competitive US market. Documentation is typically in the form of marketing materials and case studies making it more difficult to assess validity and applicability. Several studies have attempted to evaluate the validity and reliability of criteria; these studies focus on open-access criteria (mainly the Appropriateness Evaluation Protocol) and in general, reliability tends to be reported as fair to good and validity tends to be reported as poor to fair.

Benefits of UR

Used effectively, UR can provide evidence in relation to patient flow within an organisation - for example, the prevalence of admissions for potential treatment in primary care and delays (for example, in diagnostic tests), associations between referrals and appropriateness of admissions, patient “process times” and associations between demand for emergency care and impact on elective care (Coast, 1996).

Operationally, embedded UR solutions can support demand and capacity management, through the provision of real-time, standardised data and suggest actions for improving resource utilisation. Anecdotal feedback from case study sites suggests that strategically, such solutions can provide an evidence base to highlight opportunities for service improvements and redesign by identifying inappropriate but justified care, variations in care, delay trends and inefficiencies. The availability of an objective evidence base based on consistent measures helps to quantify the potential benefits of service improvements e.g. rapid reporting of test results (Smith et al, 1997) and feedback suggests can improve the quality of clinical dialogue.

Poulos et al (2011) also suggest that UR can provide a structured way to assess medical stability and identifying patients who can be discharged for rehabilitation, thus assisting patient flow. Indirectly, UR can help an organisation to reduce inappropriate admissions and length of stay, thus reducing iatrogenic risk of care in hospital (Poulos and Eagar, 2007), ensuring safer transitions of care. Organisations could also use UR to assist planning, to identify the characteristics of patients more likely to have an inappropriate stay (Soria-Aledo et al, 2012) and predict system pressures.

Limitations of UR

One of the key limitations of UR, and other interventions aimed at shifting care, is the lack of knowledge on the costs (including opportunity costs) and feasibility of providing care in alternative settings (Smith HE et al, 1997). We cannot be sure lower technology alternatives are in fact more cost effective (Coast, 1995; Lang et al, 1999; Poulos and Eagar, 2007) in terms of outcomes (Kalant et al, 2000; Vetter, 2003). Coast (1996) points out "appropriate care is not the same as efficient care and inappropriate care could be more cost effective than the alternative". UR tends to be based on theoretical rather than actual availability of lower-technology alternatives (McDonagh et al, 2000; Vetter, 2003) and disinvestment can be politically difficult (Coast, 1996). Smith et al (1997) also point out a possible risk of unintended consequences to other services e.g. community services; a health-economy wide view may help. As stated earlier, much of the published research is now quite dated and predates PbR.

There are some concerns noted in the literature as to the validity of published criteria, such as the Appropriateness Evaluation Protocol. Black (1995) noted international differences in the definitions of appropriateness (for example, the frequency of vital sign monitoring) which can limit the applicability of criteria in different contexts (Poulos and Eager, 2007). There are difficulties in generalising across settings, demographies, specialties, countries, time due to differences in numbers of beds, specialties and case mix across organisations (Smeets et al, 2000; McDonagh et al, 2000). There is also a question of how valid criteria may be across different sociodemographic groups - for example, poor housing/access to home support is associated with higher levels of inappropriateness (Houghton et al, 1996). It should be noted that the research cited here pre-dates the development of embedded tools available in the UK and there is very limited current evidence on the validity of such tools.

There are some criticisms in the literature of the value of standalone UR audits. Smeets et al (2000) suggested that changing test order behaviour and improving discharge planning were likely to have a greater cost benefit than UR. Poulos and Eagar (2007) estimate that only 50% of inappropriate days could be avoided without incurring additional resource.

Applications in the US

To inform this scoping study, we have liaised with Veterans Health Administration and Kaiser Permanente in the US to understand how learning from their approach to utilisation review may inform application within NHS settings. The following case studies are based on conversations with UR leads.

Kaiser Permanente

UR is used to screen admissions (in emergency departments) and across inpatient services. UR is seen as a part of a broad approach to quality and service improvement. Kaiser Permanente use McKesson's Interqual criteria to assess medical necessity and appropriate levels of care; the application of the criteria is viewed as one part of a multi-pronged approach to improving hospital utilisation. UR Chiefs are appointed in all hospitals to oversee the UR approach and to resolve instances where there may be disagreement as to the appropriateness of an admission or stay.

Since 2009, as part of a care coordination model, case managers (licensed, registered nurses) have been partnered with physicians to assess the appropriateness of inpatient stays, on a daily basis. Each pairing of case manager and physician has between 10-14 patients assigned to them and they perform twice-daily

rounds. Between 2008 and 2013, Kaiser Permanente achieved a reduction of 376 beds across Northern California through better care coordination. Case managers work on a 24/7 basis in all hospitals.

Within emergency departments, clinical decision areas are set up, providing dedicated beds with care led by general medicine hospitalists. Patients not meeting admission criteria (e.g. chest pain, syncope) are observed, with hourly assessments. Timely care (from diagnosis through to discharge) is a priority, reducing the risk of harm through falls, pressure sores and hospital-acquired infections. Social workers are assigned to emergency departments, acting as "patient financial advisors" to address issues which may delay discharge, advising on placements, availability of services and financial problems.

Veterans Health Administration

Within VHA, UR is seen as a key component of quality management, to ensure "the right care, in the right setting, at the right time, for the right reason utilizing evidence-based practice and continuous measurement and improvement" (Department of Veterans Affairs, 2014). Currently, UR is focused on inpatient care but VHA is considering expanding to include high-cost diagnostics, outpatient care and community/residential care. UR is seen as a tool in daily patient flow management, identifying appropriateness of the level of care and services; providing information to assist decision making; and identifying delays in treatment and services. The process is supported by the National Utilization Management Integration (NUMI) system which integrates with the electronic health record; reviews are conducted by UR nurses who are also encouraged to attend "huddles" (IHI initiative) and ward rounds. VHA requires facilities to appoint Physician Utilization Management Advisors (PUMA) who will make recommendations where there is disagreement between the UR nurse and admitting/responsible clinician. UR is standard practice for elective and non-elective care; UR nurses are being increasingly used in Emergency Departments to screen admissions.

VHA requires its hospitals to review all days of acute inpatient care, stipulating a contractual minimum of 75% of admissions, observation stays and subsequent days of care to be reviewed; high performing facilities may apply for a waiver to focus on other UM activities. Over 8 million reviews have been logged within the last 5 years, providing valuable learning on the barriers and blockages to patient flow. Facilities receive routine data reports on a monthly basis which includes some benchmarking.

UR nurses are assessed on an annual basis and are supported through monthly educational calls. In terms of skills, interpersonal skills and the ability to build rapport are considered to be essential. Some UR nurses also act as case managers; there has been some debate as to whether this may result in a conflict of interest, however, anecdotal feedback suggests this expanded role increases job satisfaction.

The UR programme is overseen by the National Utilization Management Administration Committee, whose role includes approval of standardised evidence-based criteria; currently, VHA use McKesson's Interqual criteria. Facilities are responsible for routinely measuring inter-rater reliability; however, the central team will also run ad-hoc audits as part of their coaching support offer to facilities. This coaching support tends to focus initially on the system issues to identify quick wins; cultural issues, such as physician preferences, can prove more difficult to resolve.

VHA has commissioned a study on the critical success factors of high performing organizations which is due to report in Autumn 2014.

UR in the NHS

Much of the published literature reports on the applicability of standardised criteria developed internationally to a UK setting. Many of these studies are now quite dated and studies using open source criteria have often modified the criteria to suit local circumstances and opted for different sampling approaches, making it difficult to draw robust comparisons (Ash, 1995). A systematic review by McDonagh et al (2000) looked at the use of UR in different specialties. In mental health, the proportion of inappropriate admissions was assessed at around 30%, and of bed days in the range 24-56%. In general adults, inappropriate admissions typically ranged between 15-50%; in older people, there were usually higher rates of inappropriateness due to social needs and co-morbidities with rates varying 6-78% depending on tools, methodology, sampling etc. Within paediatrics, estimates ranged from 10-20% for admissions, and around 30% for subsequent bed days. A summary overview of studies identified is listed in Appendices 2 and 3.

In 2006, the Department of Health published its position statement on utilisation review (Department of Health, 2006). The North West model was cited as good practice. Delivered over a 10-week timescale, 2-3 times per year across a health economy footprint, the cost was £14-22K per cycle, or around £35k per year for the PCT. Results reported include a 9% reduction in inappropriate admissions and cost reinvestment opportunities of over £1 million at large acute sites. The process incorporated 4 phases:

1. Comparative analysis using standard datasets to show relationships between admission flows and factors such as public holidays and cancellations of elective surgery
2. A utilisation review of all admitted patients over a minimum period of 2 weeks, using a modified version of the Appropriateness Evaluation Protocol (AEP).
3. A feedback and priority setting event to agree actions and next steps.
4. A rapid improvement cycle focusing on high impact changes.

NHS England (2013) has recently published the 2014/15 CQUIN "Inpatient flow improvement through clinical utilisation review" which provides financial incentives for trusts to adopt utilisation review technology to manage patient flow. The documentation states: " A number of English hospitals have implemented UR clinical software applications to reduce admission rates, improve flow and discharge as well as, with commissioners, right-size capacity in step down and community services to match clinical need. The financial and quality case for providers for implementing this change is compelling; this CQUIN provides additional incentive and allows for early adoption without financial risk to providers until the sustainable gains can be shown to be delivering an on-going return on investment that outweighs the costs of change. NHS England direct commissioners are encouraged to apply the CQUIN to priority areas of care for specialised services, & assess potential for a jointly funded scheme with CCG commissioners for greater reach across all inpatient services where this fits local priorities". The CQUIN provides funding over 1-2 years for:

- baseline review and delivery of quick wins (indicative cost of £180,000)
- implementation and set up for ongoing real-time use (£80,000 for Year 1, £8000 for Year 2 and additional funding of £150 per bed per year capped at £120,000 per year)
- training and rollout (£300 per person trained; £120 per bed x % of occupied bed days which have been reviewed)
- benefits realisation (£5000 plus £720 per bed per year)

Numerous NHS organisations have invested in utilisation review, either in the form of standalone audits to diagnose local issues or embedded solutions to provide real-time monitoring. Within the West Midlands, the Heart of England Foundation Trust and Shrewsbury and Telford Hospitals NHS Trust have undertaken reviews recently and a number of other trusts are in discussions with suppliers.

Within the NHS, the East Midlands has led a systematic and coordinated approach to UR. Over 4 years, the team has conducted 35 different reviews, including snapshot audits, a live concurrent embedded programme, readmissions reviews and a mental health pilot in older adults [personal communication, Rachel Sharp, 17/10/14]. The reviews have been conducted in a range of specialties and settings including urgent care, paediatrics, critical care and specialised commissioning.

A key question for decision makers will be to determine if utilisation review is likely to be cost effective. The resource required to set up and maintain utilisation review is not insignificant. There is also a decision to be made at what level utilisation review will be most useful - many examples are trust-based; however, as the review is likely to highlight system-wide issues, it may be more appropriate to address utilisation review at a system-wide level. Buy in of senior clinicians and managers is critical if the insights offered by utilisation review are to be acted on.

There is a recognition that there is significant variation across health economies in the use of hospital beds; for example, Imison et al (2012) found an almost fourfold difference in emergency admissions for patients aged over 65 years across 4 PCTs. The report recommendations include more system working; early involvement of senior clinicians; comprehensive assessment; and planned discharge. The Health Foundation have recently commissioned a number of projects on improving patient flow (Health Foundation, 2013a; Health Foundation, 2013b) which explored various interventions; some of these ideas are being taken forward at the development sites of the Future Hospitals programme (https://www.rcplondon.ac.uk/projects/future-hospital-programme). Health economies, working together via the Better Care Fund, will be developing initiatives to reduce unplanned care and to shift more care towards community-based settings. Utilisation review is likely to offer greater value if it fits within a broader approach of mutually reinforcing interventions - Kaiser Permanente is an example of such an integrated approach.

The current UR market in the UK

UR criteria

One of the key arguments for UR focuses on the application of objective, clinical and evidence-based criteria. It is suggested that clinical judgment alone to determine the appropriate level of care is limited, due to the complexity and subjective nature of decision making and the influence of many discretionary factors. The aim of UR criteria is to provide an objective tool to determine where a patient may best be cared for. Typically, criteria focus on the severity of illness (e.g. body temperature for patients with fever) and the intensity of service - if one criterion is met, the admission or day of stay is considered appropriate. If no criteria are met, there is typically a second set of criteria to understand the reasons why the admission or stay may be inappropriate (e.g. diagnosis or treatment can be delivered on an outpatient basis).

The most commonly used criteria in the published literature is the Appropriateness Evaluation Protocol (AEP), developed in the US and modified for Europe. Gertman and Restuccia (1981) who developed the AEP, acknowledge that criteria such as the AEP are an attempt to establish an absolute measure and will therefore produce higher estimates of inappropriate use than subjective techniques which will factor in adjustments. They note that 100% appropriateness is "neither practical nor desirable" but do not suggest what an optimum level might be; this is likely to be influenced by local availability of services, case mix and other factors.

Most criteria originate in the US and have often been adapted for a European and/or UK context. Criteria often have an override function, which allows the reviewer to override an assessment - that is, if no criteria are met, they may override and deem the admission or day of stay appropriate and vice versa.

Embedded solutions from Medworxx and the Oak Group have built in criteria. These solutions appear to address some of the limitations of earlier criteria, for example, reviewers can assess patients who do not meet criteria on an individual basis, taking into account what would be considered progress (e.g. respiratory flow) for that specific patient.

Criteria are usually assessed to measure validity and reliability. It is difficult to measure validity due to the lack of a real “gold standard” (Phelps, 1993); typically, studies will use the gold standard of an expert panel but a consensus-based panel itself is subject to bias. Phelps suggests a more statistically valid approach using maximum-likelihood methods which enable an estimate of the true prevalence of inappropriate admissions/days. Reliability is measured in two ways: the level of agreement (as a percentage) and using Cohen's kappa score (researchers tend to suggest a score above 0.75 or 0.8 is good) which is considered to be a more robust measure.

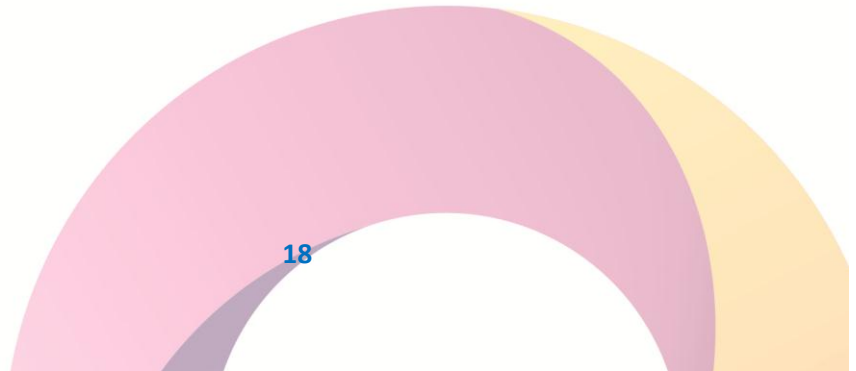
Validity measures how well the tool measures what it is intended to measure, in other words, how good are the criteria at measuring the true prevalence of inappropriateness?

Reliability measures how similar the results are when the criteria are applied by different reviewers.

Table 1 : Summary of UR criteria

Appropriateness Evaluation Protocol (AEP) *Open source*	The AEP was developed in the US in 1981 by Gertman and Restuccia (Gertman and Restuccia, 1981). Under the EU BIOMED programme, a standard European version was developed in 1999. The AEP includes a list of criteria to identify if an admission or day of care is appropriate- if one of the criteria is met, the admission or day of care is deemed appropriate. If not, the AEP also includes a list of reasons for inappropriate admissions and days of care. McDonagh et al (2000) suggest that the AEP has a reliability ranging between 24-75% for admissions and between 64-85% for subsequent days of care; and a validity between 39-80% for admissions and 59-91% for subsequent days of care. McDonagh et al suggest the AEP may overestimate inappropriateness as it assumes 100% appropriateness is possible. The AEP has been adapted for different contexts: for example, community beds (Donald et al, 2001); paediatrics (Werneke and MacFaul, 1996.
Interqual: ISD (Intensity Severity Discharge) *Proprietary*	The ISD criteria were developed by McKesson in the US in 1978 (Mitus, 2008). The ISD comprises diagnosis-independent criteria relating to the severity of the patient's condition and the intensity of care they are receiving. Over 60 clinical staff at McKesson are involved in development of the tool and review is undertaken by a network of over 800 practicing clinicians. McDonagh et al (2000) suggest ease of use is limited due to the number of criteria (22) which takes time. The same criteria are used for admissions and subsequent bed days. Papers relating to the use of ISD in the UK are quite dated and suggested at that time, that the applicability to a UK context was limited, due to the lack of alternatives to admission. Reliability of the tool ranges between kappa=0.18-0.32, and validity ranges between kappa=0.36-0.38. Versions are available for acute care; subacute care; rehabilitation; and home care.
Managed Care Appropriateness Protocol (MCAP) *Proprietary*	The MCAP criteria were developed in the late 1980s by Restuccia and Tarr (Restuccia, 1995), based on the AEP but with the aim of applying to broader range of patients (e.g. obstetrics and psychiatry) and to identify higher levels of avoidable hospitalisation. The criteria are now owned by the Oak Group and used for diagnostic reviews and audits and are embedded in a proprietary solution. The criteria are updated annually following literature reviews, feedback and a review of the previous year's criteria. Proposed revisions are presented to an expert panel comprising international experts.
Medical Patients Assessment Protocol (MPAP) *Open source*	The MPAP was developed in 1996 in Israel (Mozes et al, 1996). The aim of the criteria was to improve on the AEP criteria, recognising the shift towards more outpatient-based care since the inception of the AEP. The criteria focus on bed days (that is, not admissions) of adult medical patients and emphasise the identification of higher risk patients with more intensive needs. The criteria list medical and non-medical reasons for hospital stays and were designed to focus on the patient's condition rather than intensity of service and aim to provide more accurate definitions of non-medical reasons for stay. Unlike the AEP, there is no built-in override function. The criteria developers measured the reliability of the criteria as kappa=0.94 but did not test the validity.

Oxford Bed Study Instrument *Open source*	The Oxford Bed Study Instrument was developed from the AEP (by Anderson et al, 1988). The criteria are applied to days of stay (that is, not admissions). The first part considers patients' needs for medical, nursing and life support services - if no criteria are met, the reviewer works through part 2 to assess reasons for not being at home. There is limited reliability and validity testing of the criteria in the literature. The tool has been criticised for potentially underestimating the inappropriateness of care due to the assumption that diagnostic tests should be performed on an inpatient basis.
Medworxx Patient Flow Platform *Proprietary*	The Medworxx criteria were originally based on the AEP but have evolved through an annual updating process. A team of clinical people scans for evidence annually, making recommendations for amendments to an international panel. Customers are also invited to participate in the process by commenting on the proposed changes and are encouraged to submit their own recommendations. Updated criteria are released in Spring.



UR suppliers

Our scoping study has identified a small number of organisations currently working within the UK, in the field of utilisation review. These organisations vary in their offers: consultancy support for diagnostic audits, change management and/or embedded solutions. Some of the large consultancies (e.g. Capita, EY, McKinsey) have also worked in this area. Table 2 shows the current private sector consultancies specialising in UR.

There are currently two suppliers offering embedded solutions within the UK market: Medworxx and the Oak Group. Within the larger US market, there are more suppliers, notably, McKesson and MCG. McKesson (http://www.mckesson.com/) has previously offered its Interqual tools within the UK (e.g. in the Greater East Midlands area) but announced the sale of the McKesson UK operations in 2013 (http://www.ehi.co.uk/news/EHI/8571/mckesson-uk-put-up-for-sale). MCG (formerly Milliman Care Guidelines and now part of the Hearst Health group) is another key supplier in the US but whilst it offers a care guidelines service in the UK (http://www.careguidelines.com/content/uk), it does not yet offer its utilisation review products here.

Medworxx Patient Flow Platform

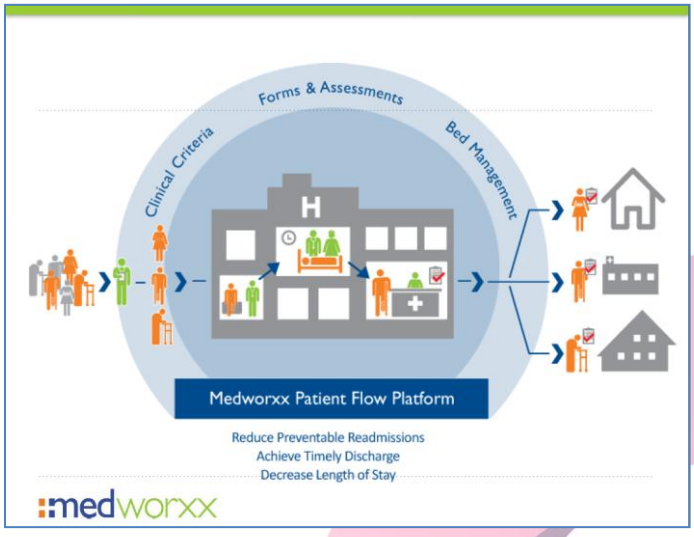


Figure 1: Medworxx

Medworxx (http://www.medworxx.com/en-gb) was established in 2004 in Canada and began working in the UK in 2010, with the first implementation of their embedded solution going live in 2011. The Patient Flow Platform can be used to provide standalone diagnostic audits and as a fully embedded solution for real-time operational management.

Medworxx offer trusts a diagnostic snapshot which can be retrospective or concurrent. A retrospective audit may take around 4 days, using a sample of 100 patients. 2 trained assessors would work with 4-6 key stakeholders, holding short interviews, a ward round and a bed management meeting. A concurrent audit would typically take longer (7-14 days) with more assessors (2-6) and daily bed management meetings/ward

Table 2: Private sector consultancies specialising in UR

	Beacon UK	Monmouth Partners	Model Advice DC Consulting
URL	http://beaconhs.co.uk/	http://www.monmouthpartners.com/home	http://www.modeladvicedc.co.uk/index.html
History	Launched September 2011	Founded December 2012	
About the company	Beacon work with purchasers and providers from the NHS, private and third sectors, specialising in mental health. Beacon works as an 'integrator', coordinating care for people with mental health problems, developing intensive community services, and focusing on reducing both mental health and acute inpatient hospitalisation.	Monmouth Partners is a small consultancy comprising senior directors with experience within the NHS and within consultancies such as Tribal and Capita. In addition to substantive staff, Monmouth has a network of associates comprising clinicians, analysts, consultants and project/programme managers. The Director team is advised by an Independent Advisory Board, membership of which includes senior clinical, academic, informatics and pharmaceutical industry expertise. Monmouth specialise in patient involvement/experience; clinical audit; and health systems improvement.	ModelAdvice DC Consulting Ltd UK has experience in health care process improvement, PbR and PBC program management and has grown from work in the US with Kaiser Permanente, Providence Health System, Premier Inc and VHA. ModelAdvice comprises associates from the UK and North America.
UR offer	Beacon offer retrospective reviews of inpatient and outpatient mental health services, making recommendations for service improvements, length of stay reduction and alternative models for care delivery. The approach is based on data analysis, financial analysis, a review of best practice from the literature and stakeholder interviews	Monmouth Partners are experts in utilisation management and the use of clinical assessment tools to determine the appropriate levels of care required by a given patient population. They have relationships with all of the UR tool providers and can assist NHS clients with optimising care utilisation in the acute secondary care setting; community hospital services; and across the interface between the two.	Model Advice Consulting offer a UR programme incorporating governance support, analysis of utilisation patterns, developing utilisation management methodologies such as case management, and benchmarking.
Implementation sites	NHS MH trust in East of England (Fielden et al, 2014)		Model Advice Consulting have previously worked with Medworxx at the Royal Liverpool and Broadgreen implementation.

rounds. The report and findings include analysis of bed utilisation effectiveness; barriers and sources of delays; and recommendations to improve flow and utilisation.

The embedded solution comprises three modules: clinical criteria; forms and assessments; and bed management. The solution is cloud-based and integrates with the local PAS (patient administration system) using HL7 messaging. The interface is web-based. Implementation from the point of a contract being awarded to launch typically takes 10-12 weeks. The tool is designed to take between 30 seconds to 2 minutes to complete per patient. The reviewer assesses the patient using a set of criteria. If the patient doesn't meet the criteria, the reason for delay is explored and a decision made as to whether the patient is ready and clinically stable for transfer/discharge. The tool enables the reviewer to assess the steady state on an individual basis as this will vary from patient to patient, thus addressing one of the early criticisms of UR criteria as being too generic and homogenous.

Table 3 : Medworxx costs

Embedded solution
Annual licence costs: approximately £100-150 per bed **Setup, integration, configuration and training**: approximately £50,000 **Hosting**: £20 per bed per annum
Diagnostic audits
Option 1: focused retrospective review for a specific service area on a sample of 75-125 patients: approximately £10,000-£20,000 **Option 2**: more extensive review across the organisation on a sample of 300 or fewer patients: approximately £25,000-£40,000 **Option 3**: full concurrent review over 10 consecutive days: approximately £6,5000-£100,000

An operational dashboard gives an overview of current patients and their status (meets criteria for the level of care; doesn't meet criteria but not clinically stable for discharge; or doesn't meet criteria and is ready for discharge). A management report provides analysis to identify opportunities for service redesign, for example, delay trends. Typically, Medworxx find 30% of patients to be at an inappropriate level of care and around two thirds of the reasons for delays can be addressed internally.

Training (2 days) is provided for reviewers who must pass an exam annually to validate skills and knowledge.

The key markets for the solution are currently the UK, Canada, Australia and the US. Medworxx's first client in the UK was the Royal Liverpool and Broadgreen University Hospitals NHS Trust and they have recently installed their solution in South Tees Hospitals NHS Foundation Trust. Medworxx are also partners with the Greater East Midlands CSU and Central South CSU. The solution has been anglicised to work within the UK (for example, ensuring compliance with NICE guidance and ensuring British spelling throughout).

Case study : Royal Liverpool and Broadgreen University Hospitals NHS Trust

Implemented in 2011, the Patient Flow Platform enables concurrent review of every patient every day with the aim of ensuring a rigorous and consistent approach to provide robust measurement and analysis. The trust reported the following achievements after the first year:

- The average number of patients on the ready for discharge list dropped from 45 to 18
- 15% increase in respiratory patient throughput and length of stay reduced from 10.5 to 9.7 days
- 30% increase in the number of referrals into community beds
- transfer of care delays reduced from 500 to 100 (50% reduction within the first month)

The trust targeted clinical gerontology and respiratory care through an increased focus on case management. Case management leads attended Directorate meetings and targeted action plans were established to reduce length of stay. Root cause analysis meetings were held using data generated through the Medworxx solution. The trust also ran a pilot to test case managers taking on a trusted assessor role, referring directly for packages of care, reducing the need for social work assessments. The increased utilisation of community beds led to questions including:

- What services are there to support patients at home and why are they not being used fully?
- Who can access what services?
- What information is needed to access services?

The Trust uses Sharepoint to store and share reports generated by the solution: daily operational reports; daily snapshots; and executive trends.

In the longer term, the Trust is planning a move towards a "discharge to assess" model and the introduction of real time bed boards at ward level.

Source: http://medworxx.com/sites/default/files/Royal%20Liverpool_MedWorxx-Case-Study-July-13_2012.pdf

The Oak Group

The Oak Group (http://www.oakgroup.com) originates in the US and specialises in utilisation review, offering a solution (Making Care Appropriate to Patients -MCAP) comprising clinical criteria with a data collection and reporting system which can be used in medical/surgical and mental health care settings. The solution aims to place patients for admission or continued stay at the most appropriate level of care and asks 3 questions:

- is the patient at the correct level of care (Figure 2);
- if not, what is the correct level of care;

- and, if not, what are the blockages to the patient being at the correct level of care to meet his/her treatment needs?

The solution works on the assumption that treatment should be as close to home as possible for patients and in the community where possible. The solution can be used to provide standalone diagnostic audits or as an embedded solution to support operational management.

Table 4 : Oak Group costs

Embedded solution
1 year pilot licence (based on single site Trust): £69,200 including: • £55,400 Includes MCAP System licensing, technical implementation, interfaces, training, clinical software support, SaaS Provision, N3 connection, database and application server provision and support. • £13,800 Implementation consultancy The cost for subsequent years varies depending on the duration of the licence. A 3-year contract would cost £55,400 each year per site which includes licensing, technical implementation, interfaces, training, clinical software support, SaaS Provision, N3 connection, database and application server provision and support. A 2 year contract would incur a 10% price increment (£60,940 per year) and a 1-year contract, a 20% price increment (£66,480 for the year). Discounts are available for implementations procured within 6 months of a baseline study.
Diagnostic audits
Utilisation management audit: single point admissions or length of stay review for each patient: from £6,500 based on a sample of 100 to £52,000 for a sample of 1000 Acute flow diagnostic audit: an admissions or length of stay review for each bed for each day over 7 days: ranging from £27,000 for a sample of 100 beds to £140,500 for a sample of 1000 beds Community/Intermediate diagnostic audit: a retrospective admissions review, historic and current length of stay review: ranging from £16,000 for a sample of 50 patients to £40,000 for a sample of 150 patients ED performance review: a retrospective review of the previous days admissions for 7 days: ranging from £15,800 for less than 50 admissions per day to £28,800 for less than 100 admissions per day Acute snapshot diagnostic audit: a retrospective admissions review, a concurrent length of stay review and a mid-point retrospective review: ranging from £9,800 for 100 or less records to £44,400 for 400-500 records. Readmission diagnostic audit: previous admissions review and intermediate discharge: ranging from £10,800 for less than 100 records, to £46,200 for 400-500 records

Within the UK, the Oak Group find that around 20-25% of admissions, 50-60% of bed days and 10-12% of readmissions could either be avoided or managed in alternative settings.

For the embedded solution, an installation will usually take between 6-8 weeks. The solution runs through the N3 network and uses HL7 messaging. Training is provided for reviewers who are also subjected to frequent reliability tests. Additionally, to minimise variation and potential reviewer bias, reviews are managed by teams on wards rather than single individuals. The embedded solution is currently being used by Ealing Hospital, Charing Cross Hospital, St. Mary's Hospital, and West Middlesex Hospital (all around London), and Queen Elizabeth's Hospital King's Lynn.

Case study: Shropshire, Telford and Wrekin Health Economy
The Oak Group worked with commissioners and providers during Summer 2013 to undertake an audit, with the aim of identifying potential opportunities to improve patient flow. The audit found:

- **Acute care**: 16% of admissions were non-qualified and 48% of continuing stay days were non-qualified.
- The percentage of days of care that could have been provided at home with a variety of services was 59% for PRH and 46% for RSH.
- 23% of the reasons for non-qualified days are outside the control of the acute trust for PRH and 27% for RSH.
- Only 28% of PRH patients and 43% of RSH patients had a discharge plan, almost all were completed after admission.
- **Other levels of care**: 15% of admission days and 47% of continuing stay days were non-qualified at the rehabilitation complex level of care; 4% of admission days and 53% of continuing stay days were non-qualified at the rehabilitation intermediate level of care and 22% of admission days and 69% of continuing stay days were non-qualified at the medical intermediate level of care.
- Discharge planning occurred in about 48% of rehabilitation complex patients, 41% of rehabilitation intermediate patients and 63% of medical intermediate patients.

Source: The Oak Group

NHS services

Within the NHS, there are two commissioning support units (CSUs) offering utilisation review as a service: Greater East Midlands CSU (GEMCSU) and North West CSU (NWCSU). GEMCSU (http://www.gemcsu.nhs.uk/services/service-redesign/) has a history of utilisation review, from the earlier EMPACT (East Midlands Procurement and Commissioning Transformation) programme and has a licence with Medworxx. NWCSU (http://northwestcsu.nhs.uk/about#services) offer analysis, working with a range of organisations to provide real time operational support – identifying issues and providing information, insight and clinical-impact analysis combined with practical support.

The Oak Group

Levels of care

a. For medical-surgical patients - From ICU to home and five levels in-between (whether adult or paediatric)

b. For rehabilitation patients - From complex to home based rehabilitation and 3 levels in-between; there are also criteria for PT, OT, SALT, and chiropractic services

c. For neonatal care- From well-baby to intensive subspecialty and with three levels in-between

d. These levels of care and those below can be used for all of initial admission, step-down or step-up levels of care.

e. For mental health, and divided into adult and child-adolescent areas, which includes geriatric psychiatry – From psychiatric intensive care, acute care, rehabilitation care, residential care and observation as bedded services, and for crisis team care, two types of day treatment programs, extended and short-term secondary care, and consultative services for outpatient care.

f. For Autism Spectrum Disorder (ASD)- From acute units in specialized settings to residential, day therapeutic schools and applied behavioural analysis therapy

g. For the elderly - For placement for patients with neurocognitive disorders.

h. For substance abuse - From medically managed detoxification, medially monitored detoxification, rehabilitation care, residential care, day treatment programs of two types, and outpatient services.

i. While The Oak Group would agree that coordination among physical health, mental health, substance misuse, and social care is critical, there is no established and evidence-

Figure 2 : Oak Group levels of care

Learning and recommendations

For the Clinical Senate

The applicability of US-derived clinical criteria to a UK context does not appear to be an issue. Although early research suggested this may be a problem, suppliers such as Medworxx and the Oak Group have adapted criteria to comply with NICE guidance; the largest UK contract (East Midlands) stipulates ongoing annual assurance of NICE compliance as a contractual obligation. This suggests it would not be worthwhile to develop a bespoke tool tailored to the West Midlands, particularly given the resource involved in creating and maintaining such a tool.

Importantly, our study shows that, as with other improvement interventions, UR is likely to be less successful in isolation – it is not a silver bullet solution. Where UR has proved successful, it has been integrated into an overall approach to managing avoidable admissions and length of stay with realistic expectations. Within Kaiser Permanente, UR is seen as one piece in a larger jigsaw of quality improvement initiatives to improve patient flow along with the recognition that change takes time to embed effectively. Resources such as the new quality standards from WMQRS (http://www.wmqrs.nhs.uk/download/518/WMQRS-Discharge-%26-Int-Care-QS-V1-20140812_1412936167.pdf) and the Patient Flow programme from the Health Foundation (http://www.health.org.uk/publications/improving-patient-flow/) may be useful.

Woodhams et al (2012) consider UR in the context of other interventions to reduce avoidable admissions (including predictive modelling, case management, virtual wards) concluding: "Well-meaning interventions are ineffective at a macro level- whether because the relentless rise in admission is too great; the interventions are not sufficiently powerful in isolation, or they are not being used to their full potential. All three are probably contributory".

UR may be best addressed at a health economy level rather than at an organisational level, to address wider strategic as well as operational issues and to avoid unintended consequences of pushing issues downstream. Health economies should be clear as to why they are undertaking UR and ensure the right people are committed and involved from the outset. Houghton et al (1996) and Donald et al (2001) point out that effective UR needs strong links between all providers within the local health and social economy to ensure timely provision of care at the right level. Anecdotal feedback suggests that the findings of review can help to challenge long-standing assumptions and persistent myths. Commissioners in particular have a key role in ensuring UR implementations don't result in system paralysis (Department of Health, 2006).

Anecdotal evidence from case studies suggests that clinicians and managers value the use of an objective clinical algorithm. Analysis through UR can challenge anecdote and myths; for example, Kaiser Permanente had a particular view of the frequent users of their services but analysis revealed this cohort to be younger than expected, which required a different approach to their care.

Used effectively, UR can provide an evidence base for continued improvement but there is currently a lack of incentive or mechanisms for organisations to share their learning; there is some evidence of a shared learning approach working in the East Midlands. Whilst this report attempts to capture some of the critical success factors and lessons learned, there is an opportunity for an ongoing role to capture the characteristics of high performing organisations and support for lower performing organisations to apply learning to their local context.

For health economies considering UR

Approach and planning

Health economies considering UR should consider a system wide approach as UR is likely to highlight economy-wide problems which cannot be resolved within one single organisation. UR is not a silver bullet solution and will be more valuable as part of an integrated approach to managing admissions and length of stay. Expectations need to be realistic in relation to how long it may take to embed and sustain change.

One of the limitations of UR, and of other interventions and initiatives focusing on the shift care from acute to community, is that we don't really know if lower-technology alternatives are in fact more cost effective than an acute bed. It is therefore critical that underlying assumptions are made explicit in making the case for UR. For example, is there a shared understanding of what constitutes the cost of an acute bed (staffing, drugs etc). A baseline will be essential to understand the impact and consequences of UR across the local health and social care economy.

Coast (1996) asks on what assumptions are potential savings based - favourable assumptions (e.g. same length of stay in a community bed to an acute bed; no readmissions) can suggest savings but more pessimistic assumptions could suggest additional costs. Coast (1996) suggests exploration of the factors influencing cost effectiveness: the context within which care is provided (e.g. pressure on beds and the opportunity costs of treating "inappropriate" patients); the current pattern of care (available services and capacity); proportions of fixed and variable costs; and potential economies of scale (alternatives to acute care are often on a smaller scale.

Kaiser Permanente have focused on coordination and quality of care ("Care Without Delay") rather than on making savings and efficiencies - this has led to an outcomes-based approach rather than a focus on process and delivered a real reduction in beds. The use of UR provides an evidence base enabling commissioners and providers to plan on the basis of sound analysis and there is an opportunity to utilise the learning from research and implementation evidence to plan interventions which have been shown to be effective.

The case studies we have explored show that improvement of services and processes depends on behaviour change and a structured change management approach. UR is a significant investment and to demonstrate value for money, should be underpinned by a clear theory of change underpinning decisions. There will be impacts on roles and responsibilities (for example, retraining discharge planners as case managers) as well as how things are done and this will lead to cultural change; the Royal Liverpool and Broadgreen University Hospitals NHS Trust found that UR raised questions about how

patients are managed and the alternatives available. Learning from case studies suggests that integration of UR into existing processes can help to embed UR within the organisation.

Sponsorship and governance

The importance of senior clinical buy-in is reflected in the case studies we have explored and echoed in a recent report from the Advisory Board Company (2014) which emphasises the early involvement of clinicians to ensure success of change initiatives in patient flow. The report points out that engagement from the outset is likely to increase support and compliance. This is reflected in the case studies we have explored; for example, Kaiser Permanente initiatives have required strong and sustained clinical leadership.

Vetter (2003) highlights the need for a systematic approach to managing accountability and responsibility for UR to ensure a governance structure to resolve the issues identified. Smeets et al (2000) point out the risk of demotivating staff if reviews highlight problems which are not addressed.

Information quality and information governance

UR relies on accurate and complete data; research (Ramos-Cuadra et al, 1995) suggests that retrospective reviews find higher rates of inappropriateness as there is a greater reliance on incomplete data whereas concurrent reviews can incorporate staff input (Houghton et al, 1996). However, concurrent reviews may be more resource intensive in collecting information required (Santos-Eggimann B et al, 1997) unless using an embedded solution. The quality of medical records impacts on the reliability of retrospective audits; differences in completeness were found between acute and community records (Donald et al, 2001).

From an operational perspective, there are risks around access to and quality of information. Within England, information governance arrangements can present significant barriers, with suppliers reporting wide variations in local policies and processes. The availability of information is also key; missing notes may be more likely in relation to sicker patients thus risking skewing of results. Concurrent review needs to be carefully managed to avoid disruption to wards, as information is gathered (Smeets et al, 2000). Standalone audits can only provide a snapshot which may not be adequate to consider implications; suppliers argue that embedded solutions will enable more significant change. The Advisory Board Company (2014) recommends that hospitals drive improvements in data quality; to challenge current practice, clinicians will need to trust in the data.

Methodology

For organisations or health economies aiming to complete a standalone audit to give a snapshot view, the methodology needs to be carefully planned to avoid potential bias. There are a number of options regarding the methodology of utilisation review: for example, to conduct a retrospective audit or to implement a system of concurrent or prospective review; to review a cross sectional sample of patient days or entire stays for samples of patients; to review admissions or subsequent days of stay or both; to review elective or non elective care. All these choices will impact on the validity of the results and have implications for subsequent modelling.

For standalone audits, the sampling methodology should aim for a representative sample. Inappropriate days are associated more with proximity to discharge than length of stay therefore the day of stay to be audited can inadvertently over or under estimate the prevalence of inappropriate stays (Houghton A et al, 1996; Donald IP et al,2001). There are also seasonal and day of the week variations to consider (Ash, 1995); studies report higher rates of appropriateness when bed occupancy is higher (Smeets et al, 2000; Soria-Aledo et al, 2012). Essentially, as Smeets et al (2000) stress, the outcome can vary depending on sampling, length of stay and the part of the stay being reviewed.

Operational management

Establishing routine UR will require dedicated resource and appropriate governance. From 2009, Kaiser Permanente has teamed up all physicians with an inpatient case manager (each of these teams has 10-14 patients assigned to them at any one time). Utilisation management chiefs have been appointed in all hospitals to manage disagreement between utilisation review nurses and physicians. VHA in the US has found that system and flow issues are considerably easier to address than cultural issues and typically offer sufficient improvements. It is also important to fully understand the cost of lower technology alternatives to acute admission; for example, it could cost more to implement 7 day working to reduce diagnostic delays than to have patients occupying beds. Gertman and Restuccia (1981) note the need for careful consideration of tradeoffs - for example, will it cost more to create new nursing home beds than it costs to simply redefine existing beds? This reinforces the value of a system-wide approach involving commissioners.

Training and support is important. Coaching support is provided to VHA hospitals to help address and prioritise areas for improvement; this team typically comprises 2 individuals so as not to overwhelm the hospital. There is also a need to ensure tools are used consistently; for example, Medworxx test knowledge on an annual basis and McKesson has established routine testing of inter-rater reliability. Where embedded solutions are being implemented, there will be implications for workforce planning, possibly through the development of new roles or refocusing existing roles.

Critical success factors

Strong and consistent senior leadership to ensure UR is used to make improvements

Senior clinical sponsor to lead UR in each organisation

System wide approach to address wider strategic issues and avoid unintended consequences

A focus on quality not solely efficiency

A systematic approach to identifying and managing issues impacting on patient flow

Explicit change management to support behaviour change behaviours

Early involvement of clinicians and multidisciplinary teams to ensure ownership and buy in

Training and support

Careful consideration of underlying assumptions and tradeoffs

Locally agreed objectives and outcomes

Consideration of the end to end patient journey

Processes to ensure consistent use of UR tools e.g. refresher training, routinely testing inter-rate reliability

Commitment to a continued and sustained approach

Consider specific requirements which should be reflected in contracts e.g. assurance of NICE compliance

Recommendations for research and evaluation

The challenge we have faced in synthesising the evidence base is that much of the published research is now dated and mostly relates to one-off retrospective reviews – our search did not identify any primary research assessing the effectiveness of embedded utilisation review solutions. New primary research would further develop the evidence base, enabling a better understanding of the utility and impact of utilisation review. There are two areas in particular which we suggest should be addressed by new primary research:

The validity of UR tools. As indicated in our report, it has been challenging to test the validity of UR tools due to the lack of a gold standard. Typically, UR tools have been compared to expert consensus panels, which are subjective. Phelps (1993) suggests a more statistically valid approach using maximum-

likelihood methods which enable an estimate of the true prevalence of inappropriate admissions/days. It would be helpful to test this method in an NHS setting.

The effectiveness of UR. The gold standard for evaluating an intervention is an RCT, including an economic analysis to assess the cost of the intervention and any savings achieved. An RCT could involve randomising different wards in a trust: to the intervention group, implementing UR or to the control group, who continue with standard practice. This method reduces the risk of bias and facilitates statistical analysis. However, trials can be expensive and resource-intensive, and can take significant time. A mixed methods approach would be useful in collecting qualitative data as well as quantitative data, such as stakeholder (patients, staff, commissioners, GPs etc.) responses to utilisation review, which could capture key lessons which could be shared across the health economy and NHS.

Recommendations for health economies planning UR implementations

As with any new intervention, it will be important for health economies to evaluate the local impact. This should be planned at the start of any change initiative to ensure the baseline position is measured and to ensure adequate resource (time, people, money) to complete the evaluation. An evaluation should aim to address:

- Efficacy – does it work? How far have the intended outcomes been achieved?
- Efficiency – how are resources such as funding, timing and staffing used? Is the intervention delivering value for money (measuring costs against outcomes)?
- Elegance – is the transformation acceptable? What do we need to keep doing and what do we need to stop doing?
- Effectiveness – does the transformation help to achieve the longer term objectives?
- Ethicality – are there any ethical or moral issues with the transformation? Were there any unintended outcomes as a result of this intervention?

For each of the key evaluation questions, health economies will need to determine:

- Indicators – how will this be measured?
- Baseline – what is the starting position?
- Source of data – where will you get this data/how will you collect the data?

There will be some challenges involved in undertaking primary research and evaluation, for example, information governance (e.g. linking health and social care data, linking primary and secondary care data). It would be valuable if resources could be identified to support local health economies to undertake and publish robust evaluations.

Appendix 1: Methodology

This scoping study is based on 3 phases of review and research:

1. *Rapid review of the evidence base*

The following databases were searched: Medline, Embase and HMIC on 4 July 2014. Keywords (sample search strategy included below) were identified from an initial scoping search which identified thesaurus and free text terms from relevant papers. The search was limited to the last 20 years and focused some international studies were included (US, Canada, Australia, Europe).

2. *Review of current suppliers*

Suppliers were identified through existing knowledge of the market, a Google search and a search of trade press (HSJ and eHealth Insider). Information was extracted from supplier web sites and through semi-structured interviews with supplier representatives. CSU web sites were also searched to identify those with a service offer relating to utilisation review.

3. *Review of implementations/case studies*

Implementation sites and case studies were identified through searches of the trade press (HSJ and eHealth Insider) and through semi-structured interviews with suppliers. Information was collected through semi structured interviews with case study sites.

Table 5 : Search strategy (Medline)

1	*UTILIZATION REVIEW/ OR *CONCURRENT REVIEW/
2	"utilisation review".ti
3	"utilization review".ti
4	1 OR 2 OR 3
5	appropriateness.ti,ab
6	exp HEALTH SERVICES MISUSE/
7	HOSPITALIZATION/ OR LENGTH OF STAY/ OR PATIENT ADMISSION/ OR PATIENT DISCHARGE/
8	5 OR 6 OR 7
9	4 AND 8
10	4 [Limit to: Review Articles and (Publication Types Meta Analysis or Systematic Reviews)]
11	interqual.ti,ab
12	"appropriateness evaluation protocol".ti,ab
13	"managed care appropriateness protocol".ti,ab
14	"intensity severity discharge".ti,ab
15	"intensity of service severity of illness discharge screens".ti,ab
16	11 OR 13 OR 14 OR 15
17	9 OR 16
18	17 [Limit to: English Language]
19	18 [Limit to: (Publication Types Journal Article) and English Language]
20	19 [Limit to: Publication Year 1994-2014 and (Publication Types Journal Article) and English Language]

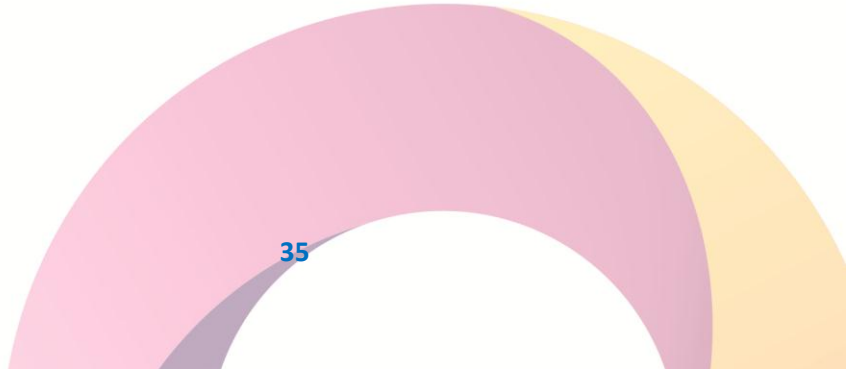
Appendix 2 : Summary overview of evidence - by country

UK

Location	Setting	Sample	Instrument	Findings	Reference
South West England (2 hospitals)	General medicine and geriatrics	Hospital 1 - 634 emergency admissions; Hospital 2 - 619 emergency admissions, both over a 6 month period	ISD-A	In both hospitals, 80% of the admissions were defined as appropriate	Coast J et al (1996) Factors associated with inappropriate emergency hospital admission in the UK, International Journal for Quality in Health Care, 8 (1), 31-39.
DGH in North Staffordshire	Emergency medical admissions	10% random sample of 8210 patients	AEP (without use of over-ride function)	6% of admissions were found to be inappropriate and 45% of subsequent day of care were found to be inappropriate. Over half of patients had a stay in which at least half of the days were judged inappropriate. The most common reason for inappropriateness was waiting for diagnostic tests or treatments which could have been delivered on an outpatient basis. The entire stay was judged to be appropriate in 10.6% of patients. The trust identified 16,500 bed days (a year) could be saved through easier availability of tests, faster reporting, improved discharge planning and increased use of outpatients.	Smith HE et al (1997) Appropriateness of acute medical admissions and length of stay, Journal of the Royal College of Physicians of London, 31 (5), 532.
DGH and community hospitals in Gloucestershire	2 DGHs and 9 CHs	220 DGH and 220 emergency medical admissions from Sep/Oct '98 and	AEP to assess admissions to DGH; CHAEP (Community Hospital	330 of DGH admissions (75%) satisfied at least 1 AEP criterion; a further 69 (16%) were given an override to appropriateness by their responsible clinician. Of the 41 admissions judged inappropriate by AEP, 33 (7.5% of all DGH admissions) justified admission to the CH by the	Donald IP et al (2001) Defining the appropriate use of community hospital beds, British Journal of General Practice, 51, 95-100.

Location	Setting	Sample	Instrument	Findings	Reference
		Jan/Feb '99 (total of 880)	version of AEP developed locally) to define which patients satisfied criteria for community hospital.	CHAEP criteria. 55% of subsequent days of care were judged appropriate. At the CH, 360 (82%) admissions satisfied a CHAEP criterion (of whom 36 also satisfied one of the AEP criteria). 3 admissions met AEP but not CHAEP criteria, indicating that their illness was of a severity more appropriate for the DGH. Only 12 admissions (3%) were given an override to appropriateness by their clinician. The remaining 15% of CH admissions were judged inappropriate.	
Homerton Hospital, Hackney (part of St Barts group)	Adult inpatients	Case-note review of 625 adult inpatients – first 24 hours of care following admission and final 24 hours of care prior to discharge	AEP	31% of admissions were found to be inappropriate, 66% of the last days of stay were found to be inappropriate. Of the admissions found to be inappropriate, 29% (49) could have been cared for at home, 13% (23) were suitable for outpatient care and 58% (98) were deemed appropriate for non-acute beds.In relation to the last day before discharge, 34% were found to be appropriate.	Houghton A et al (1996) Appropriateness of admission and the last 24 hours of hospital care in medical wards in an East London Teaching Group hospital, International Journal for Quality in Health Care, 8(6), 543-553.
2 DGHs	Paediatric admissions	Case records of 47 admissions to 2 DGHs; 13 admissions were also assessed for subsequent days of care	PAEP	Reviewers using PAEP rated 57% of admissions as appropriate, compared to 63% of reviewers using their subjective judgment.	Esmail A (2000) Development of the Paediatric Appropriateness Evaluation Protocol for use in the United Kingdom, Journal of Public Health Medicine, 22 (2), 224-230.
Royal Shrewsbury	Acute	Case note review of 88 patients	AEP	28% of admissions were assessed as inappropriate. The most common reason was because tests or treatment	Campbell J (2001) Inappropriate admissions: thoughts of patients and

Location	Setting	Sample	Instrument	Findings	Reference
Hospital	admissions	admitted over 4 weeks in November 2000		could have been performed in outpatients. The review also included questionnaires with patients and with referring GPs and admitting doctors to assess willingness to consider alternatives to admission.	referring doctors, Journal of the Royal Society of Medicine, 94, 628-631.
Acute hospital in South West region	General medicine and geriatrics	Sample of 677 admissions over a period of 6 months	ISD	19.7% of admissions were assessed as inappropriate; further assessment by a panel of GPs reduced this to between 9.8-15.0%. The most common alternatives to admission were community hospital and urgent outpatient assessment. Potential resource savings (based on average rather than marginal costs) were calculated as relatively low: for community/GP beds, estimated savings were in the range £14007 to £42484 (n=449) and for urgent outpatient assessments, estimated savings were in the range £183617 to £207875 (n=274) but a sensitivity analysis shows that as assumptions on the cost of GP consultation, length of stay and readmissions become less favourable, the savings are significantly reduced or of negative value.	Coast J et al (1995) The hospitals admission study in England: are there alternatives to emergency hospital admission?



Europe

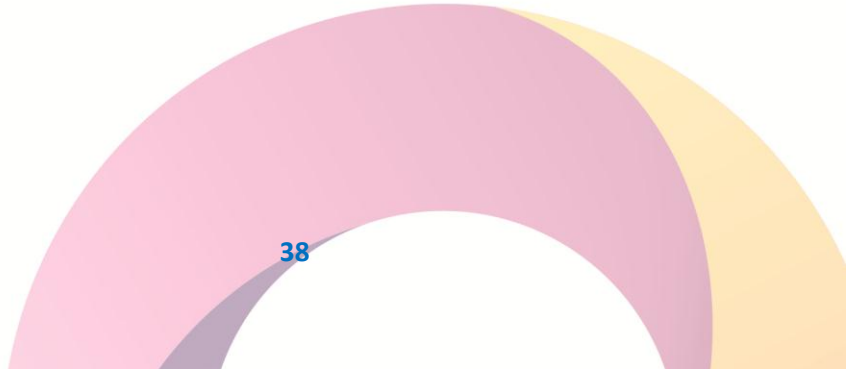
Country	Setting	Sample	Instrument	Findings	Reference
Belgium	23 hospitals	10,921 days of adult acute care (non-intensive) assessed for appropriateness	AEP	24.61% of days were assessed as inappropriate with variation across specialties and hospitals although there was some variation across specialties. The most common reasons for inappropriateness are "waits for an examination" and "lack of extra-hospital structures"	Fontaine P et al (2011) Assessing the causes inducing lengthening of hospital stays by means of the Appropriateness Evaluation Protocol, Health Policy, 99 (1), 66-71.
Denmark	Acute medical admissions in a hospital serving population of 400k	470 patients	AEP	14% were classified as inappropriate. 73 admissions failed to meet any AEP criteria, 131 admissions only met one AEP criterion. Several admissions could have been handled appropriately by a diagnostic unit or through sub-acute referral to an outpatient clinic the following day.	Jepsen HK et al (2013) Every seventh acute medical admission is preventable, Danish Medical Journal, 60 (3), A4595.
Italy	Teaching hospital	373 bed days over 3 days during September 2010 (general medicine)	AEP	44.6% of bed days were rated as inappropriate. Length of stay of over 10 days was significantly associated with higher levels of inappropriateness (odds ratio = 2.04, 95% CI 1.17-3.56). Of the inappropriate days, 12.8% were due to external factors and 87.2% were due to internal factors such as waiting for an investigation or procedure or doctors failing to discharge on time.	Barisonzo R et al (2011) Length of stay as risk factor for inappropriate hospital days: interaction with patient age and comorbidity, Journal of Evaluation in Clinical Practice, 19, 80-85.
Spain	University hospital (426 beds) in southern Spain serving population of 320,000	725 admissions (1350 bed days) in 2005 - a follow up review was	AEP	The level of inappropriateness of admission in the control group was 7.4% (54 patients), whereas in the intervention group it was significantly reduced to 3.2% (23 patients) ($p < 0.001$). The most common cause was waiting for tests or	Soria-Aledo V (2012) Reduction in inappropriate hospital use based on analysis of the causes, BMC Health Services Research, 12, 361.

Country	Setting	Sample	Instrument	Findings	Reference
		conducted in 2007 on a sample of the same size after an improvement cycle		treatment which could be delivered in an outpatient setting. The level of inappropriateness of stays was 24.6% (334 patients) in the control group and 10.4% (137 patients) in the intervention group ($p < 0.001$). The most common causes were waiting for tests or treatment which could be delivered in an outpatient setting, waiting for test results, and "conservative attitude of doctor". The cost of the days considered inappropriate in the study sample, taking into account the mean cost per patient, clinical service and day, was €147,044 in the control group and €66,462 in the intervention group.	
Italy	2 hospital-based rehabilitation wards	371 bed days (adults)	Tool developed by Guilé et al, 2009 for rehabilitation	22.9% of the days of stay were assessed as inappropriate. The most common reason was social and/or family environment issues (34.1%).	Bianco A (2012) Validity and Reliability of a Tool for Determining Appropriateness of Days of Stay: An Observational Study in the orthopedic intensive rehabilitation Facilities in Italy. PLoS ONE 7(11):e50260
Germany	400-bed teaching hospital, Frankfurt	240 medical patients, 262 surgical patients	AEP	33% of all surgical admissions and 28% of subsequent bed days were assessed as inappropriate. 6% of medical admissions and 33% of subsequent bed days were assessed as inappropriate.	Sangha A et al (2002) Metric properties of the appropriateness evaluation protocol and predictors of inappropriate hospital use in Germany: an approach using longitudinal patient data, International Journal for Quality in Health Care, 14 (6), 483-492.

Country	Setting	Sample	Instrument	Findings	Reference
Denmark	Regional teaching hospital	3050 patients in medical admission units (October 1008 to February 2009)	AEP	61.9% of the patients fulfilled the AEP criteria. Patients fulfilling were older ($p < 0.001$), had a higher inhospital mortality ($p < 0.001$), a higher 30-days mortality ($p < 0.001$), a longer length of stay ($p < 0.001$), more readmissions within 30 days ($p < 0.001$) and higher co-morbidity ($p < 0.001$).	Brabrand M (2011) The characteristics and prognosis of patients fulfilling the Appropriateness Evaluation Protocol in a medical admission unit; a prospective observational study, BMC Health Services Research 2011, 11:152.

US, Canada and Australia

Country	Setting	Sample	Instrument	Findings	Reference
Canada	26 hospitals in Manitoba	3904 patients	ISD	After 1 week, 53.2% of patients assessed as needing acute care at admission no longer required acute care. The authors suggest utilisation review should focus on patients with stays longer than 1 week.	DeCoster C(1997) Inappropriate hospital use by patients receiving care for medical conditions: targeting utilization review, CMAJ, 157, 889-896.



Appendix 3 : Overview of evidence - by criteria

Comparison studies

Country	Setting	Sample	Findings	Reference
Canada	Acute cardiology service	75 patients admitted with provisional diagnosis of acute myocardial infarction or unstable angina. Comparison of AEP, MCAP, ISD against an expert panel.	The panel assessed 92% of admissions as appropriate and 67% of subsequent bed days as appropriate. ISD underestimated the appropriateness (80% and 40% respectively); AEP and MCAP overestimated the appropriateness (100% of admissions as appropriate; AEP assessing 85% and MCAP 81% of subsequent days as appropriate). The agreement between the tools and the panel were: ISD κ=0.45; AEP κ = 0.25 and MCAP κ = 0.24). suggesting poor to fair validity. Sensitivity scores : ISD: admissions 1.0; subsequent days of care 0.94 AEP: admissions 1.0; subsequent days of care 0.29 MCAP: admissions 1.0; subsequent days of care 0.33 Specificity scores for all tools for admissions was reported as higher than 0.87. Specificity scores for subsequent days of care were: ISD 0.57 AEP 0.92 MCAP 0.87	Kalant N et al (2000) How valid are utilization review tools in assessing appropriate use of acute care beds? CMAJ, 162 (13), 1809-1813.

Appropriateness Evaluation Protocol

Country	Setting	Sample	Findings	Reference
UK - North Staffordshire Hospital NHS Trust	Emergency admissions	10% of a random sample of 8210 patients, ranging from 15 to 95 years	The study reports a high level of agreement between AEP and a physician panel with scores (on a sample of 40 patients) of 0.83 for admissions and 0.90 for subsequent days of care.	Smith HE et al (1997) Appropriateness of acute medical admissions and length of stay, Journal of the Royal College of Physicians of London, 31 (5), 532.

Country	Setting	Sample	Findings	Reference
UK – Queen Alexandra Hospital, Portsmouth	Acute geriatric admissions:	146 emergency admissions during a 3 week period	Sensitivity of 97%, specificity of 63%, positive predictive value of 95%, and negative predictive value of 75%. Overall agreement between consultants and AEP: 92% (kappa = 0.62).	Tsang P and Severs MP (1995) A study of appropriateness of acute geriatric admissions and an assessment of the Appropriateness Evaluation Protocol, Journal of the Royal College of Physicians of London, 29 (4) 311-314.
UK	Adult inpatients – Homerton Hospital, Hackney (part of St Barts group)	Case-note review of 625 adult inpatients – first 24 hrs of care after admission and last 24 hrs of care before discharge.	40% (157) of those whose admission was assessed as "appropriate" were also assessed as being "appropriately" in hospital in the last 24 hours of care, in comparison with 22% (38) of those whose admission was assessed as "inappropriate" *(p* < 0.0001)."	Houghton A et al (1996) Appropriateness of admission and the last 24 hours of hospital care in medical wards in an East London Teaching Group hospital, International Journal for Quality in Health Care, 8(6), 543-553.
Switzerland	St-Loup hospital (regional)	162 patients totaling 1098 bed days	2 reviewers (1 internal and 1 external) reviewed concurrently and 1 year later, reviewed retrospectively. Both reviewers estimated lower rates of inappropriateness in the concurrent samples than in the retrospective samples.	Santos-Eggimann B (1997) Comparing results of concurrent and retrospective designs in a hospital utilization review, International Journal of Quality in Health Care, 9 (2), 115-120
Spain	University hospital (426 beds) in southern Spain serving population of 320,000	725 admissions (1350 bed days) in 2005 - follow up conducted in 2007 on sample of the same size after improvement cycle.	Inter-rater reliability of κ=0.31 for admissions and κ = 0.77 for subsequent bed days	Soria-Aledo V (2012) Reduction in inappropriate hospital use based on analysis of the causes, BMC Health Services Research, 12, 361.
Germany	400-bed teaching hospital	240 medical patients, 262 surgical patients	Inter-rater agreement for admissions in surgical patients was 74%, κ= 0.44 and in medical patients was 92% and κ=0.31. Inter rater agreement for bed days was 84%, κ=0.58 for surgical patients and	Sangha A et al (2002) Metric properties of the appropriateness evaluation protocol and predictors of

Country	Setting	Sample	Findings	Reference
			76%, κ=0.42 for medical patients.	inappropriate hospital use in Germany: an approach using longitudinal patient data, International Journal for Quality in Health Care, 14 (6), 483-492.
Denmark	Regional teaching hospital	3050 patients in 2 medical admission units (general medicine and cardiology) (October 1008 to February 2009)	Nurses assessed 79.1% of the admissions as appropriate. The Kappa value of nurses agreeing with the AEP criteria was 0.16 ($p < 0.0001$). There was 66.3% agreement between the nurses and AEP, with a sensitivity of 84.8%, a specificity of 30.1%, a positive predictive value of 66.3% and a negative predictive value of 54.9%. Doctors found 76.2% of the admissions appropriate. The Kappa value of doctors agreeing with the AEP criteria was 0.29 ($p < 0.0001$). There was 71.3% agreement between doctors and AEP and a sensitivity of 86.4%, a specificity of 40.9%, a positive predictive value of 71.3% and a negative predictive value of 63.9%.	Brabrand M (2011) The characteristics and prognosis of patients fulfilling the Appropriateness Evaluation Protocol in a medical admission unit; a prospective observational study, BMC Health Services Research 2011, 11:152.
Denmark	Bispebjerg Hospital	470 patients admitted to acute medical department during 3 weeks	Admissions classified as inappropriate were justified, but could have been handled otherwise either as subacute referral to an outpatient clinic or by a short pathway with diagnosis and treatment initiation without admission.	Jepsen HK et al (2013) Every seventh acute medical admission is preventable, Danish Medical Journal, 60 (3), A4595.
UK	2 DGHs and 9 CHs.	220 DGH and 220 emergency medical admissions from Sep/Oct '98 and Jan/Feb '99- total of 880.	AEP used to assess admissions to DGH and CHAEP (Community Hospital version of AEP developed locally) to define which patients satisfied criteria for community hospital. The study included a retrospective review (of 42 acute admissions and 46 community admissions) by a clinical panel to assess the review nurse's use of the criteria. Agreement between the review nurse and review panel for the AEP and DGH was κ = 0.9 (95% confidence interval (CI) = 0.7–1.0); for the CHAEP and CH it was κ = 0.37 (95% CI = 0.1–0.8).	Donald IP et al (2001) Defining the appropriate use of community hospital beds, British Journal of General Practice, 51, 95-100.

Country	Setting	Sample	Findings	Reference
UK	Paediatrics in secondary care	A sample of 50 case notes from total of 418 patients in 3 hospitals in Yorkshire - day of admission assessed (US version of PAEP)	The validity was assessed as 0.29 (95% confidence interval 0.11 to 0.47). Werneke references research which reports kappa score for reliability ranging from 0.46 and 0.89. Validity scores 0.68 for the day of admission and 0.47 (without override) and 0.6 with override for other bed days.	Werneke U and MacFaul R (1996) Evaluation of appropriateness of paediatric admission, Archives of Disease in Childhood, 74, 268-73. Werneke U et al (1997) Validation of the paediatric appropriateness evaluation protocol in British practice, Archives of Disease in Childhood 1997;77:294–298.
UK	Paediatrics	Case records of 47 admissions to 2 DGHs; 13 admissions also assessed for subsequent days of care.	Clinicians using subjective judgment were unable to make a decision in nearly 15% of cases, compared with 6% where the PAEP was used. Inter-rater reliability for reviews on day of admission is reported as κ = 0.85 for clinicians using PAEP compared to κ = 0.35 for clinicians using subjective judgment. Inter rater reliability is lower for reviews on subsequent days of care: κ = 0.54 for trained researchers. Overall, trained researchers had a higher inter-rater reliability (κ = 0.847) than clinicians (κ = 0.543).	Esmail A (2000) Development of the Paediatric Appropriateness Evaluation Protocol for use in the United Kingdom, Journal of Public Health Medicine, 22 (2), 224-230.

Interqual/Intensity Severity Discharge

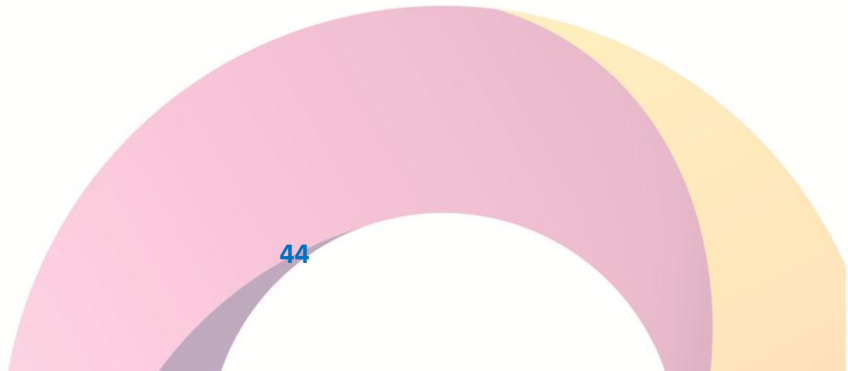
Country	Setting	Sample	Findings	Reference
UK	Emergency admissions to general medicine and geriatrics in 2 hospitals in SW England	Hospital 1 - 634 emergency admissions; Hospital 2 - 619 emergency admissions, both over 6 month period	The validity of the tool was found to be no better than fair to moderate.	Coast J et al (1996) Factors associated with inappropriate emergency hospital admission in the UK, International Journal for Quality in Health Care, 8 (1), 31-39.

Country	Setting	Sample	Findings	Reference
UK	General hospital in England - general medicine and geriatrics	700 admissions during a 6 month period (compared with AEP)	The level of agreement between ISD and AEP was assessed as 85.9%, κ = 0.65 (95% CI 0-46-084) for a sample of 78 admissions. The level of agreement between ISD and a panel of GPs was assessed as 68.1% , κ = 0.32 (95% CI 0-21-0.43). The level of agreement between ISD and a panel of hospital consultants was 51%, κ = 0.18 (95% CI 0.02-0.34). Inter-reviewer reliability was assessed as 85%, κ = 0.64 (95% CI 0.39-0.89).	Inglis AL et al (1995) Appropriateness of hospital utilization: the validity and reliability of the Intensity-Severity-Discharge review system in a United Kingdom acute hospital setting, Medical Care, 33 (9), 952-957.
Canada	26 hospitals in Manitoba	3904 patients	The overall kappa value for the pairs of abstractors (in 3 teams) ranged from 0.52 to 0.96, in the “fair” to “good” agreement range.	DeCoster C(1997) Inappropriate hospital use by patients receiving care for medical conditions: targeting utilization review, CMAJ, 157, 889-896.
US	Urban tertiary teaching hospital	503 patients with chronic heart failure	Interqual criteria were not found to predict the level of care accurately. Only one of the criteria was statistically significant at predicting level of care (blood urea nitrogen).	Wang H et al (2013) The accuracy of Interqual criteria in determining the need for observation versus hospitalization in emergency department patients with chronic heart failure, Critical Pathways in Cardiology 12 (4), 192-196.
Australia	Large acute referral hospital	696 episodes of care (7189 bed days) with a diagnosis of stroke, hip fracture or joint replacement and patients referred for rehabilitation assessment	56% of patient days in the acute hospital met Interqual criteria for acute level of care; 33% of days of stay for rehabilitation referrals met the criteria for acute level of care.	Poulos CJ et al (2011) Determining level of care appropriateness in the patient journey from acute care to rehabilitation, BMC Health Services Research, 11, 291.

Country	Setting	Sample	Findings	Reference
Israel	Acute hospital	3400 bed days	The MPAP tool was found to have higher inter-rater reliability than the AEP (kappa=0.94 and 0.78 respectively).	Mozes B et al (1996) Medical Patients Assessment Protocol: a tool for evaluating the appropriateness of hospital-stay days for acute medical patients; development, reliability and applications, American Journal of Medical Quality 11(1), 18-24.

Oxford Bed Study Instrument

Country	Setting	Sample	Findings	Reference
UK	Large teaching hospital	847 admissions over 3 weeks - criteria used for every alternative patient - 405 eligible patients, 328 reviews completed	The proportion of bed days deemed appropriate decreased from 74% on the first day after admission to 22% on the 8th day after admission. In 88% of cases, the senior nurse and senior registrar were in agreement (Kendall's coefficient of concordance W=0.88 P<0.00001).	Anderson P et al (1988) Use of hospital beds: a cohort study of admissions to a provincial teaching hospital, BMJ 297, 910-912.



References

Advisory Board Company (2014) Bringing it all together: how hospitals can improve care processes with patient flow systems, The Advisory Board Company.

Anderson P et al (1988) Use of hospital beds: a cohort study of admissions to a provincial teaching hospital, BMJ 297, 910-912.

Ash A (1995) The design and analysis of hospital utilization studies, International Journal of Quality in Health Care 7(3), 245-252.

Barisonzo R et al (2011) Length of stay as risk factor for inappropriate hospital days: interaction with patient age and comorbidity, Journal of Evaluation in Clinical Practice, 19, 80-85.

Bianco A (2012) Validity and Reliability of a Tool for Determining Appropriateness of Days of Stay: An Observational Study in the Orthopedic Intensive Rehabilitation Facilities in Italy. PLoS ONE 7(11): e50260

Black N (1995) European collaboration on appropriateness of hospital bed use: a commentary, International Journal of Quality in Health Care 7(3), 185-186.

Brabrand M (2011) The characteristics and prognosis of patients fulfilling the Appropriateness Evaluation Protocol in a medical admission unit; a prospective observational study, BMC Health Services Research 2011, 11:152.

Campbell J (2001) Inappropriate admissions: thoughts of patients and referring doctors, Journal of the Royal Society of Medicine, 94, 628-631.

Coast J (1996) Appropriateness versus efficiency: the economics of utilisation review, Health Policy 36 (1), 69-82.

Coast J et al (1996) Factors associated with inappropriate emergency hospital admission in the UK, International Journal for Quality in Health Care, 8 (1), 31-39.

Coast J et al (1995) The hospital admissions study in England: are there alternatives to emergency admission? Journal of Epidemiology and Community Health, 49, 194-199.

DeCoster C(1997) Inappropriate hospital use by patients receiving care for medical conditions: targeting utilization review, CMAJ, 157, 889-896.

Department of Health (2006) Care and resource utilisation: ensuring appropriateness of care, Department of Health.

Department of Veterans Affairs (2014) VHA Directive 1117: Utilization Management, Department of Veterans Affairs.

Donabedian A (1973) Aspects of medical care administration: specifying requirements for health care, Harvard University Press.

Donald IP et al (2001) Defining the appropriate use of community hospital beds, British Journal of General Practice, 51, 95-100.

Esmail A (2000) Development of the Paediatric Appropriateness Evaluation Protocol for use in the United Kingdom, Journal of Public Health Medicine, 22 (2), 224-230.

Fielden J et al (2014) Shining the spotlight on mental health problems in acute care, HSJ (7 April 2014).

Fontaine P et al (2011) Assessing the causes inducing lengthening of hospital stays by means of the Appropriateness Evaluation Protocol, Health Policy, 99 (1), 66-71.

Foster P and Prairie L (2014) Utilization review essentials, Hospital Medicine Clinics, 3, e255-269.

Gertman PM and Restuccia JD (1981) The Appropriateness Evaluation Protocol: a technique for assessing unnecessary days of hospital care, Medical Care 19 (8), 855-871.

Gray BH and Field MJ (eds) (1989) Controlling costs and changing patient care? The role of utilization management, Institute of Medicine.

Health Foundation (2013a) Improving patient flow: how two trusts focused on flow to improve the quality of care and use available capacity effectively, Health Foundation.

Health Foundation (2013b) Improving the flow of older people: Sheffield Teaching Hospital NHS Trust's experience of the Flow Cost Quality improvement programme, Health Foundation.

Houghton A et al (1996) Appropriateness of admission and the last 24 hours of hospital care in medical wards in an East London Teaching Group hospital, International Journal for Quality in Health Care, 8(6), 543-553.

Imison et al (2012) Older people and emergency bed use: exploring variation, Kings Fund.

Inglis AL et al (1995) Appropriateness of hospital utilization: the validity and reliability of the Intensity-Severity-Discharge review system in a United Kingdom acute hospital setting, Medical Care, 33 (9), 952-957.

Jepsen HK et al (2013) Every seventh acute medical admission is preventable, Danish Medical Journal, 60 (3), A4595.

Kalant N et al (2000) How valid are utilization review tools in assessing appropriate use of acute care beds? CMAJ, 162 (13), 1809-1813.

Lang T et al (1999) A European version of the Appropriateness Evaluation Protocol: goals and presentation, International Journal of Technology Assessment in Health Care, 15 (1), 185-197.

Lorenzo S et al (1999a) An experience of utilization review in Europe: sequel to a BIOMED project, International Journal for Quality in Health Care, 11 (1), 13-19.

Lorenzo S et al (1999b) Reliability study of the European appropriateness evaluation protocol, International Journal for Quality in Health Care, 11 (5), 419-424.

McDonagh MS et al (2000) Measuring appropriate use of acute beds: a systematic review of methods and results, Health Policy, 53, 157-184.

Mitus JA (2008) The birth of InterQual: evidence-based decision support criteria that helped change healthcare, Professional Case Management, 13 (4), 228-233.

Mozes B et al (1996) Medical Patients Assessment Protocol: a tool for evaluating the appropriateness of hospital-stay days for acute medical patients; development, reliability and applications, American Journal of Medical Quality 11(1), 18-24.

NHS England (2013) Commissioning for quality and innovation: 2014/15 model schedules for patient flow improvement through clinical utilisation review, NHS England, Available at: http://www.england.nhs.uk/wp-content/uploads/2013/10/comm-intent.pdf.

O'Neill D and Pearson M (1995) Appropriateness of hospital use in the United Kingdom: a review of activity in the field, International Journal of Quality in Health Care, 7 (3), 239-244.

O'Regan NA et al (2014) The Appropriateness Evaluation Protocol is a poor predictor of in-hospital mortality, Irish Journal of Medical Sciences, 183, 417-421.

Phelps CE (1993) The methodologic foundations of studies of the appropriateness of medical care, New England Journal of Medicine, 329 (17), 1241-1245.

Poulos CJ and Eagar K (2007) Determining appropriateness for rehabilitation or other subacute care: is there a role for utilisation review? Australia and New Zealand Health Policy 2007, 4:3

Poulos CJ et al (2011) Determining level of care appropriateness in the patient journey from acute care to rehabilitation, BMC Health Services Research, 11, 291.

Ramos-Cuadra A et al (1995) The effect of completeness of medical records on the determination of appropriateness of hospital days, International Journal of Quality in Health Care, 7 (3), 267-275.

Restuccia JD (1995) The evolution of hospital utilization review methods in the United States, International Journal of Quality in Health Care, 7 (3), 253-260.

Sangha A et al (2002) Metric properties of the appropriateness evaluation protocol and predictors of inappropriate hospital use in Germany: an approach using longitudinal patient data, International Journal for Quality in Health Care, 14 (6), 483-492.

Santos-Eggimann B (1997) Comparing results of concurrent and retrospective designs in a hospital utilization review, International Journal of Quality in Health Care, 9 (2), 115-120.

Smeets PMJH et al (2000) Assessing the necessity of hospital stay by means of the Appropriateness Evaluation Protocol: how strong is the evidence to proceed? International Journal of Quality in Health Care, 12 (6), 483-493.

Smith HE et al (1997) Appropriateness of acute medical admissions and length of stay, Journal of the Royal College of Physicians of London, 31 (5), 532.

Soria-Aledo V (2012) Reduction in inappropriate hospital use based on analysis of the causes, BMC Health Services Research, 12, 361.

Tsang P and Severs MP (1995) A study of appropriateness of acute geriatric admissions and an assessment of the Appropriateness Evaluation Protocol, Journal of the Royal College of Physicians of London, 29 (4) 311-314.

Vetter N (2003) Inappropriately delayed discharge from hospital: what do we know? BMJ, 326, 927-928.

Wang H et al (2013) The accuracy of Interqual criteria in determining the need for observation versus hospitalization in emergency department patients with chronic heart failure, Critical Pathways in Cardiology 12 (4), 192-196.

Werneke U et al (1997) Validation of the paediatric appropriateness evaluation protocol in British practice, Archives of Disease in Childhood 1997;77:294–298.

Werneke U and MacFaul R (1996) Evaluation of appropriateness of paediatric admission, Archives of Disease in Childhood, 74, 268-73.

Woodhams V et al (2012) Triumph of hope over experience: learning from interventions to reduce avoidable hospital admissions identified through an Academic Health and Social Care Network, BMC Health Services Research, 12, 153.

The Strategy Unit

Working Together to Provide Strategic Transformation in the Changing Environment

The Strategy Unit exists to help others improve health and care. We combine advanced yet practically grounded skills and expertise in analysis, evidence review, strategic financial planning, policy and strategy development, consensus building, programme design and implementation and trusted advisor support for senior leaders. Our core senior team has been working together in these areas for a decade and we have extensive experience and a huge reservoir of shared learning to call upon.

Our service supports commissioners and other health and care organisations in large scale transformation of services designed to improve the outcomes for patients and the public. We have an established reputation for excellence and innovation, working locally and nationally. We work with several partners to deliver services as required, these include GHK/ICF (expert policy analysis and qualitative evaluation); Health Services Management Centre, Birmingham; a number of specialist Associates in the fields of strategic finance, mental health policy, social care, primary care commissioning. With these partners we also contribute to research and national policy development.

How Can the Strategy Unit Help You?

Working closely with a range of other expert services within the CSU as required, we provide a wrap around service to deliver high quality and high value bespoke projects.

We do this by helping you identify and agree:

- What needs to be changed;
- What needs to happen;
- How to make change happen, in the most effective way;
- How to measure and monitor progress to determine if change is successful and sustained.

Clients working with us on a regular basis benefit from regular bulletins and learning events, sharing key findings from our work elsewhere.

Our Promise to You

Each project managed by our team is bespoke and relevant to your business case. Our experience tells us that the best results are achieved through early engagement and our involvement in discussions prior to specification of requirements.

Our working methods are coaching-based wherever possible. We place great emphasis on extensive and managed engagement (with clinicians, patients, stakeholders) working collaboratively with local leadership and providing true innovation in the application of proven analytical methods to get to the heart of the questions that really matter.

We pride ourselves on our integrity and we operate to NHS values and at NHS rates.

We will ensure that the approach we take to working with you is:

- informed by evidence
- focused on patient and public benefit
- makes use of robust statistical techniques and rigorous methodologies

We will work interactively with you to:

- Transfer knowledge
- Present our findings in a professional and bespoke manner tailored to your specific audiences and target markets.

What Else Do We Do?

- Deliver usable, evidence based analytical insight into complex questions in relation to whole systems of care
- Work closely with clients in structuring and defining their constraints
- Work alongside clinical groups to build consensus in understanding the dynamics of demand and utilisation
- Develop large scale models for change
- Produce bespoke, detailed syntheses of the published evidence base on clinical/service topics but also in our ‘methods reviews’, on mechanisms and methodologies of potential application in achieving service change
- Provide expert support through the national reconfiguration assurance process

Strategy Unit

Midlands and Lancashire Commissioning Support Unit

Kingston House

438-50 High Street

West Bromwich

B70 9LD

Tel: 0121 612 1538

Email: strategy.unit@nhs.net